RECOVERING EVENT-CONDITIONED LOCAL SYMMETRY PROJECTORS FOR HYBRID REINFORCEMENT LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Exploiting symmetry allows reinforcement learning to share information across related states and actions. In contact-rich environments, some transformation directions remain compatible with the dynamics while others are broken. Scalar gates cannot represent these surviving subspaces, whereas fixed-basis diagonal gates discard the cross terms of rotated subspaces. We recover event-conditioned local symmetry projectors from paired reset-and-step simulator queries. Hybrid-flow residuals and typed-event information define a Gram operator whose low-defect eigenspace yields a full projector subject to finite compatibility checks. In a direct-wrench MuJoCo orbital-contact task, the projector's translation–rotation cross terms recover the orbit pivot and define a quotient shared by the actor, critics, target networks, and action transport. One projector recovered at a canonical condition per geometry was reused across eight task conditions, reducing recovery calls by 87.5% relative to condition-specific recovery. Out-of-distribution (OOD) evaluation used obstacle centers withheld from policy optimization but independently certified before training. Across 20 paired seeds, sparse recovered full improved OOD regulation AUC over no-op SAC by 0.131. It also outperformed diagonal and spectrum-matched orthogonal controls while closely matching oracle full. With matched MLPs, direct-pivot features did not reproduce the quotient benefit. After charging all recovery calls, recovered full retained a $+0.037$ total-call OOD AUC advantage over no-op.

1 INTRODUCTION

Physical control problems contain transformations that leave the decision problem unchanged or transform its solutions predictably. Equivariant learning encodes such relations by tying functions across group orbits (Cohen & Welling, 2016). In reinforcement learning (RL), known symmetries improve generalization in MDP homomorphic networks, SO(2)-equivariant policies and quadrotor control (van der Pol et al., 2020; Wang et al., 2022a; Yu & Lee, 2023). Equivariant diffusion policies apply these principles to demonstration-based robot learning (Yang et al., 2025; Tie et al., 2025). Standard actor–critic methods must otherwise learn these relations from data.

The useful symmetry in contact-rich control is rarely a single global group. Walls, obstacles and contact transitions can preserve some infinitesimal generators while breaking others, as shown in Fig. 1a. Approximate and partially equivariant methods relax a predefined symmetry (Wang et al., 2022b; Park et al., 2025). For instance, PE-SAC learns where a full group should be used (Chang et al., 2026). This raises a distinct question about which transformation directions remain locally valid. A scalar gate cannot represent a surviving lower-dimensional subspace, and a fixed-basis diagonal gate cannot preserve the cross terms of a rotated projector.

Those cross terms can carry operational geometry. In planar orbital contact, a rank-one projector in $\mathfrak{se}(2)$ couples translation and rotation, and the ratios of its final-column entries identify the orbit pivot. A diagonal approximation retains coordinate-wise weights but loses the cross terms used to recover the pivot. This makes full-projector recovery more than a finer gate, as it can provide the chart used to reduce the control problem.

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.

We recover that geometry inside a supplied candidate Lie algebra. Paired counterfactual simulator queries test flow commutation and event compatibility. Retained directional residuals reconstruct a local Gram, whose low-defect eigenspace defines the projector. Finite reward, event and safety checks determine which recovered projectors enter the table. In the orbital task, each accepted projector defines one quotient shared by the actor, critics, targets and action transport.

Our contributions are: (i) an event-conditioned counterfactual estimator for local symmetry subspaces inside a supplied algebra; (ii) finite-query and directional-reconstruction analysis that links residual fitting to projector and pivot error; and (iii) a controlled MuJoCo study in which one full projector, recovered and certified at a canonical condition per geometry, defines a shared quotient across eight task conditions. Across 20 paired seeds, sparse recovered full outperformed diagonal, spectrum-matched wrong and no-op controls while meeting the prespecified non-inferiority criterion against dense condition-specific recovery and reducing recovery calls by 87.5%. A matched representation control tests whether direct pivot features can replace the quotient, while a call-count-matched control evaluates performance after charging all recovery calls. Hybrid billiard and obstacle-contact experiments verify rank and direction recovery.

2 Related Work

Known group actions support equivariant RL in MDP homomorphic networks, SO(2)-equivariant policies and quadrotor control (van der Pol et al., 2020; Wang et al., 2022a; Yu & Lee, 2023). Equivariant diffusion policies extend these ideas to demonstration-based robot learning (Yang et al., 2025; Tie et al., 2025). Approximate and partial equivariance relax these constraints through symmetry-breaking layers, approximately equivariant dynamics or routed actor–critic models (Wang et al., 2022b; Romero & Lohit, 2022; Kaba & Ravanbakhsh, 2023; Park et al., 2025; Chang et al., 2026). PE-SAC uses learned gates to route between equivariant and unconstrained actor–critic components (Chang et al., 2026). Our focus is selecting and coupling the surviving directions within a supplied candidate Lie algebra.

Automatic symmetry discovery learns transformations or infinitesimal generators from data (Dehmamy et al., 2021; Yang et al., 2023; 2024; Ko et al., 2024). LieSD constructs linear equivariance constraints from a trained predictor's inputs, outputs and gradients, then uses singular value decomposition to recover the Lie algebra (Hu et al., 2025). AtlasD trains local predictors and learns the local symmetry generators under which they are equivariant (Bhat et al., 2025). Related approaches study mixtures of latent symmetries and task-space symmetries (Li et al., 2025; Nitschke & Talebi, 2026). Hybrid contact introduces event-selected flows that are piecewise differentiable, with sensitivities that change across jumps (Goebel et al., 2012; Burden et al., 2016; Kong et al., 2024). Our estimator reconstructs a residual differential directly from paired simulator rollouts within the supplied algebra. Event compatibility selects the probes used to fit its low-defect subspace.

Canonicalization maps inputs to orbit representatives, applies an unconstrained model, and transports outputs back. Learned canonicalization functions use trainable maps compatible with non-equivariant backbone architectures (Kaba et al., 2023). Symmetry-reduced coordinates have also improved deep RL control by using known global actions to identify equivalent states (Zeng & Graham, 2021). Our orbital interface instead obtains its chart from the translation–rotation cross terms of a counterfactually estimated projector. These entries determine the pivot used for canonicalization, while finite reward and event checks govern the use of the recovered chart.

3 Method

3.1 Event-conditioned counterfactual estimator

We consider a deterministic hybrid simulator with state $s \in \mathcal{S}$, mode $m \in \mathcal{M}$, action $a \in \mathcal{A}$, and one-step hybrid flow

$$\mathcal{X} = \mathcal{S} \times \mathcal{M}, \qquad \mathcal{Z} = \mathcal{X} \times \mathcal{A}, \qquad \Phi_{\Delta t} : \mathcal{Z} \to \mathcal{X}, \qquad (s, m, a) \mapsto (s^+, m^+). \tag{1}$$

The method uses reset-and-step simulator access and typed event descriptors, obtaining flow and event responses directly from rollouts rather than analytic vector fields, guard functions or reset

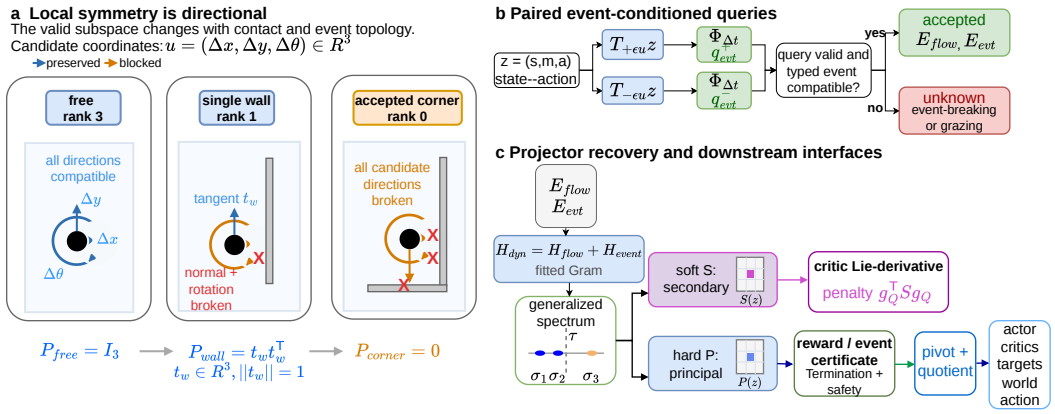

Figure 1: **Event-conditioned projector recovery and use.** **a,** Contact geometry changes the surviving subspace of $\mathfrak{se}(2)$. **b,** Paired queries screen transform and event compatibility before fitting flow residuals. Obstacle and wall recovery additionally include continuous typed-event defects. **c,** Accepted hard projectors define the shared actor–critic quotient, while soft operators supply the secondary wall critic penalty.

Jacobians. A supplied candidate Lie algebra $\mathfrak{g}_0$ with basis $\{e_j\}_{j=1}^d$ acts infinitesimally through

$$\xi_j^S(s) = \frac{\mathrm{d}}{\mathrm{d}\epsilon} T^S_{\epsilon e_j}(s)\Big|_{\epsilon=0}, \qquad \xi_j^A(a) = \frac{\mathrm{d}}{\mathrm{d}\epsilon} T^A_{\epsilon e_j}(a)\Big|_{\epsilon=0}. \tag{2}$$

At each tuple $z = (s, m, a)$, we seek directions that remain compatible with the active event regime. Write $\Phi = \Phi_{\Delta t}$ and let $T_{tu} z$ transform both state and action. We compare transforming before the simulator step with transforming its nominal output:

$$\rho(z, tu) = \Phi(T_{tu} z) \ominus T_{tu} \Phi(z), \qquad \widehat{e}_{\epsilon,u}(z) = \frac{\rho(z, \epsilon u) - \rho(z, -\epsilon u)}{2\epsilon}. \tag{3}$$

Here $\ominus$ denotes the scaled continuous-state endpoint difference in a common tangent chart and discrete event signatures are checked separately. On a smooth event stratum, $\rho(z, tu) = tE(z)u + o(t)$, so the paired residual estimates $E(z)u$.

The orbital residual has six channels: position, orientation scaled by $L = 0.35$ m, linear velocity scaled by $t_v = 0.015$ s, and angular velocity scaled by Lt_v. Each signed pair is retained only when both rollouts have finite residuals and rewards and preserve the nominal contact and termination signatures. Other pairs are discarded before fitting. For each geometry, we query one canonical condition at two radii in five coefficient frames, with 37 directions per frame. At each radius ϵ, directions from all frames are expressed in common metric-whitened coordinates and pooled into U_ϵ. The corresponding retained residual columns form R_ϵ. A full-row-rank design gives

$$\widehat{E}_\epsilon = R_\epsilon U_\epsilon^\dagger, \qquad \widehat{\bar{H}}_\epsilon = \widehat{E}_\epsilon^\top \widehat{E}_\epsilon, \qquad \widehat{\bar{H}} = \frac{1}{|\mathcal{E}|} \sum_{\epsilon \in \mathcal{E}} \mathcal{N}(\widehat{\bar{H}}_\epsilon), \tag{4}$$

where $\mathcal{N}(H) = H/\operatorname{tr}(H)$ for positive trace and $\mathcal{N}(0) = 0$. Insufficient retained directions at either radius produce geometry-level abstention. Thus frames expand the directional design, while radii produce the Grams that are averaged. Figure 1 summarizes the pipeline, while Algorithm 1 gives the complete offline and online procedure.

3.2 Spectral projector and finite compatibility checks

Flow residuals measure commutation with the hybrid step. In obstacle and wall recovery, continuous typed-event residuals additionally measure covariance of contact descriptors, giving

$$H_{\mathrm{dyn}}(z) = H_{\mathrm{flow}}(z) + H_{\mathrm{event}}(z). \tag{5}$$

Orbital recovery uses the event-screened flow Gram in Eq. (4); events determine which directions enter the fit. In physical coefficient coordinates, the spectral step is

$$H_{\mathrm{dyn}}(z)v_i = \sigma_i M_{\mathfrak{g}_0} v_i, \qquad V^\top M_{\mathfrak{g}_0} V = I. \tag{6}$$

The low-defect eigenspace defines the hard projector

$$P(z) = V_{\le\tau} V_{\le\tau}^\top M_{\mathfrak{g}_0}, \tag{7}$$

where $M_{\mathfrak{g}_0}$ is the algebra metric. In whitened coordinates, $\bar{P} = M_{\mathfrak{g}_0}^{1/2} P M_{\mathfrak{g}_0}^{-1/2}$ is an ordinary orthogonal projector. Orbital recovery diagonalizes $\widehat{\bar{H}}$ directly and selects eigenvalues below $\tau(\bar{H}) = \max\{10^{-12}, 0.015\lambda_{\max}(\bar{H})\}$. Each radius and the averaged Gram must select rank one, with at most 10° between the two radius projectors. Termination changes or excessive penetration in any probe reject the geometry. Finally, least-squares fits of paired reward differences yield the mean whitened reward gradient $\widehat{\bar{g}}_r$, screened by

$$\epsilon_{\max} \|\bar{P}\widehat{\bar{g}}_r\|_2 \le \tau_r, \tag{8}$$

where $\epsilon_{\max}$ is the largest probe radius and $\tau_r = 10^{-4}$. This bounds the fitted first-order reward variation along the recovered subspace. Geometries passing these checks enter a fixed projector table, while other geometries abstain.

3.3 Projector-conditioned quotient

The principal online experiment uses soft actor–critic (SAC) (Haarnoja et al., 2018) with one shared quotient interface. In this subsection P denotes the whitened projector. Rotation about $c = (c_x, c_y)$ has unit direction $\bar{v}_c = \frac{(c_y, -c_x, L)^\top}{\sqrt{c_x^2 + c_y^2 + L^2}}$. Its rank-one projector $P_c = \bar{v}_c \bar{v}_c^\top$ retains the pivot in its translation–rotation cross terms:

$$c(P) = \left(-L\frac{P_{12}}{P_{22}},\ L\frac{P_{02}}{P_{22}} \right), \qquad \phi_P(p) = \mathrm{atan2}((p - c(P))_y, (p - c(P))_x)\,. \tag{9}$$

Write the observation as $o = (b, g, v, \omega, m_c) \in \mathbb{R}^8$, with body orientation $b = (\cos\theta, \sin\theta)$, goal displacement g, linear velocity v, angular velocity ω and contact bit m_c. Body position p supplies the chart separately. For planar rotation R and world wrench $a = (f, \tau)$, the network inputs and actor output transport are

$$\begin{aligned} \bar{o}_P &= (R(-\phi_P)b, R(-\phi_P)g, R(-\phi_P)v, \omega, m_c), \\ \bar{a}_P &= (R(-\phi_P)f, \tau), \qquad a_{\mathrm{world}} = (R(\phi_P)f_{\mathrm{can}}, \tau_{\mathrm{can}}). \end{aligned} \tag{10}$$

The actor receives $\bar{o}_P$ and selects a canonical wrench. Online and target critics receive $(\bar{o}_P, \bar{a}_P)$. Current and next replay states use their own ϕ_P. With the exact orbital projector, $\mathcal{C}_{P_c}(s, a) := (\bar{o}_{P_c}, \bar{a}_{P_c})$ satisfies $\mathcal{C}_{P_c}(T_\alpha^S s, T_\alpha^A a) = \mathcal{C}_{P_c}(s, a)$ for rotations about c. This identifies equivalent state–action pairs along the orbit.

Every comparison changes only the supplied operator: full uses $\widehat{P}$, diagonal uses $\mathrm{diag}(\widehat{P})$, wrong uses an orthogonal rank-one projector, and scalar uses an isotropic operator. Missing, rejected and contact-inactive operators leave inputs and actions unchanged. Raw transitions remain in replay, with no auxiliary symmetry loss. Appendix A.6 specifies action feasibility and all operator controls.

3.4 Recovery and quotient stability

The stability analysis follows the computation. Directional residual error perturbs the Gram, spectral separation controls the recovered subspace, and the rotational anchor controls pivot sensitivity.

Proposition 1 (Directional residual-map reconstruction)**.** *For a directional design U, suppose U has full row rank and its residual matrix satisfies $R = EU + N$. With $\widehat{E} = RU^\dagger$, $H = E^\top E$, $\widehat{H} = \widehat{E}^\top \widehat{E}$ and $\nu = \|NU^\dagger\|_2$,*

$$\|\widehat{H} - H\|_2 \le \big(2\|E\|_2 + \nu\big)\nu. \tag{11}$$

Here U is the retained, pooled design at one radius. Trace normalization is locally Lipschitz when the population trace is bounded away from zero; averaging radius Grams bounds error by the corresponding weighted sum. Appendix A.7.3 gives these bounds.

Lemma 1 (Local subspace recovery). *Let $\mathcal{D}_\star$ and $\widehat{\mathcal{D}}$ be the bottom-k eigenspaces of $\bar{H}_\star = M_{\mathfrak{g}_0}^{-1/2} H_\star M_{\mathfrak{g}_0}^{-1/2}$ and $\widehat{\bar{H}} = M_{\mathfrak{g}_0}^{-1/2} \widehat{H} M_{\mathfrak{g}_0}^{-1/2}$, respectively. For $0 < k < d$, population eigengap $\Delta = \lambda_{k+1} - \lambda_k > 0$ and $\|\widehat{\bar{H}} - \bar{H}_\star\|_2 \le \delta < \Delta$, the Davis–Kahan bound (Yu et al., 2015) gives*

$$\|\sin\Theta(\widehat{\mathcal{D}}, \mathcal{D}_\star)\|_2 \le \frac{\delta}{\Delta - \delta}. \tag{12}$$

Angles are measured in whitened coordinates. Applying this fixed-k bound to threshold-selected projectors also requires stable rank. The eigenvalues must remain separated from the threshold. Appendix A.7 states sufficient margins for fixed and relative thresholds.

Corollary 1 (Pivot stability). *For rank-one planar projectors $P, \widehat{P}$ with $P_{22}, \widehat{P}_{22} \ge \alpha > 0$, the quotient pivots from Eq.* (9) *satisfy*

$$\|c(\widehat{P}) - c(P)\|_2 \le \sqrt{2} L \big(\alpha^{-1} + \alpha^{-2}\big) \|\widehat{P} - P\|_F. \tag{13}$$

Conditional finite-query bounds make the residual error explicit through centered-difference bias and simulator tolerance. Proposition 2 also includes neighborhood bias and sampling error for the obstacle and wall estimators. Corollary 2 compares coefficient frames on shared accepted support. Their complete assumptions and proofs appear in Appendices A.7 and B.

The secondary wall study uses a confidence-weighted critic penalty, described with its amortization controls in Appendix C.5.

4 Experiments

We evaluate projector recovery, geometry-level reuse and orbital control, with separate representation and call-count-matched comparisons. Learning curves show mean $\pm$ s.e.m. across policy seeds. Primary orbital differences use a paired two-way bootstrap over policy seeds and policy-held-out geometry centers.

4.1 Hybrid and MuJoCo projector recovery

In a closed billiard with candidate algebra $\mathfrak{g}_0 = \mathbb{R}^2$, recovery yielded rank two in free space, rank one near a wall and rank zero at corners (Fig. 2a), with 99.6% coverage (3,705/3,721) and 100% conditional rank accuracy. In MuJoCo obstacle pushing, contact normals were split into training and held-out sectors. Recovery accepted 110 of 221 held-out contact queries (49.8% coverage), also with 100% conditional rank accuracy (Fig. 2b).

On held-out circular-obstacle contact, a normal-only operator had the correct rank but a 90° principal angle, while a contact-Jacobian nullspace retained an extra rotational mode. Recovery instead achieved a mean angle of 0.083° on accepted queries. Rank and instantaneous nonpenetration therefore do not identify the orbital stabilizer (Appendix C.4).

4.2 From projector recovery to pivot accuracy

The cross terms P_{02} and P_{12} determine the orbital pivot through Eq. (9). In a separate 48-context recovery cohort, $P_{22} \ge 0.486$ and pivot error averaged 5.49 mm, with a 90th percentile of 9.46 mm. The recovered projector therefore yields an accurate quotient chart, whereas $\mathrm{diag}(P)$ removes both cross terms and induces an origin-centered chart.

Separate checks with 25 projective directions or 1% residual noise retained 100% rank accuracy and 90th-percentile angles below 1.1°. Thresholds from 0.0025 to 0.02 selected the same rank (Appendix C.3). We next test this geometry in a shared learning and canonicalization interface.

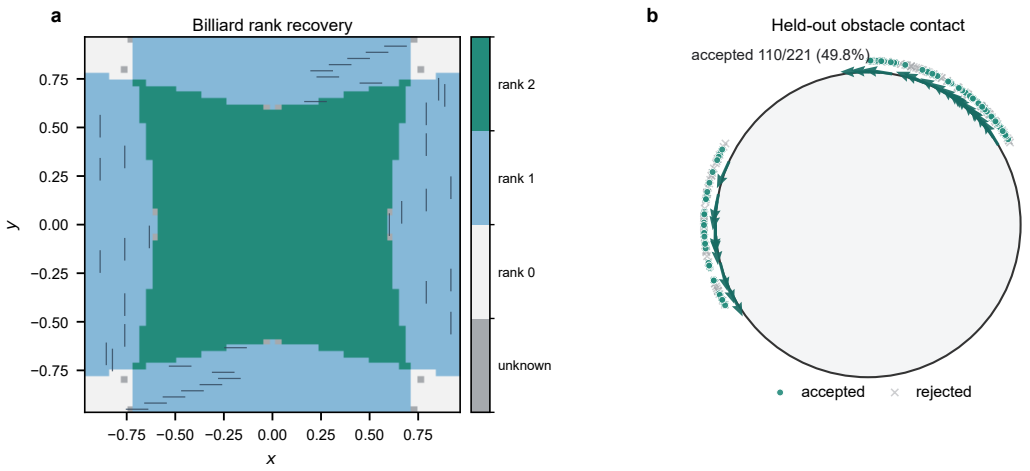

Figure 2: **Local rank and direction recovery. a,** Billiard translation-projector ranks, with short segments showing recovered rank-one directions and gray cells marking rejected queries. **b,** Held-out obstacle-contact queries arranged by contact-normal angle. Colored points mark accepted queries (110/221, 49.8% coverage), and gray crosses mark rejections. Arrows show recovered orbital directions for a subset of accepted queries.

4.3 Principal orbital-contact online control

We constructed a direct-wrench task in MuJoCo (Todorov et al., 2012) in which a non-axisymmetric planar capsule must maintain contact while moving to a goal around a circular obstacle whose center varies by context. A continuous radial preload presses the capsule toward the obstacle. Figure 3 shows the direct force and torque inputs and one frozen-policy rollout. At force-bearing contact, the valid rank-one $\mathfrak{se}(2)$ direction jointly rotates the body, goal and wrench about the obstacle center. The policy observes body orientation, goal displacement, velocity, angular velocity and a contact bit, but not the obstacle center, contact normal or clearance.

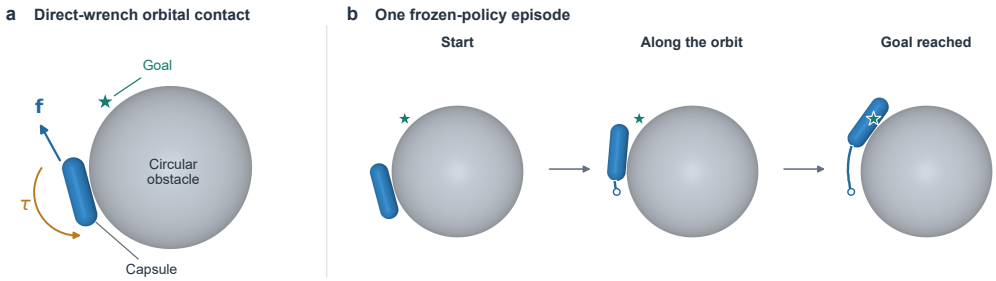

Figure 3: **Orbital-contact task and an example rollout. a,** Top-view MuJoCo scene. Force f and torque τ act directly on the capsule, and the star marks its goal. **b,** Three frames from one successful frozen recovered-full policy episode. The blue line traces the body center, and the open circle marks its initial position.

Geometry-level reuse. All 24 geometry keys, spaced at 15° around the obstacle-center orbit, were certified at one canonical condition before policy training. Training used the 12 even-indexed centers at phases -165° and 15° in both travel directions. Seen-center evaluation used the remaining phases, -75° and 105°, and policy-held-out OOD evaluation used all four phases at the 12 interleaved centers. At a fixed center, phase changes follow the same orbit and direction reversal changes only the generator sign, preserving $P_c = \bar{v}_c \bar{v}_c^\top$. Sparse recovery reused one projector across eight conditions per center, whereas dense recovery certified all 192 conditions.

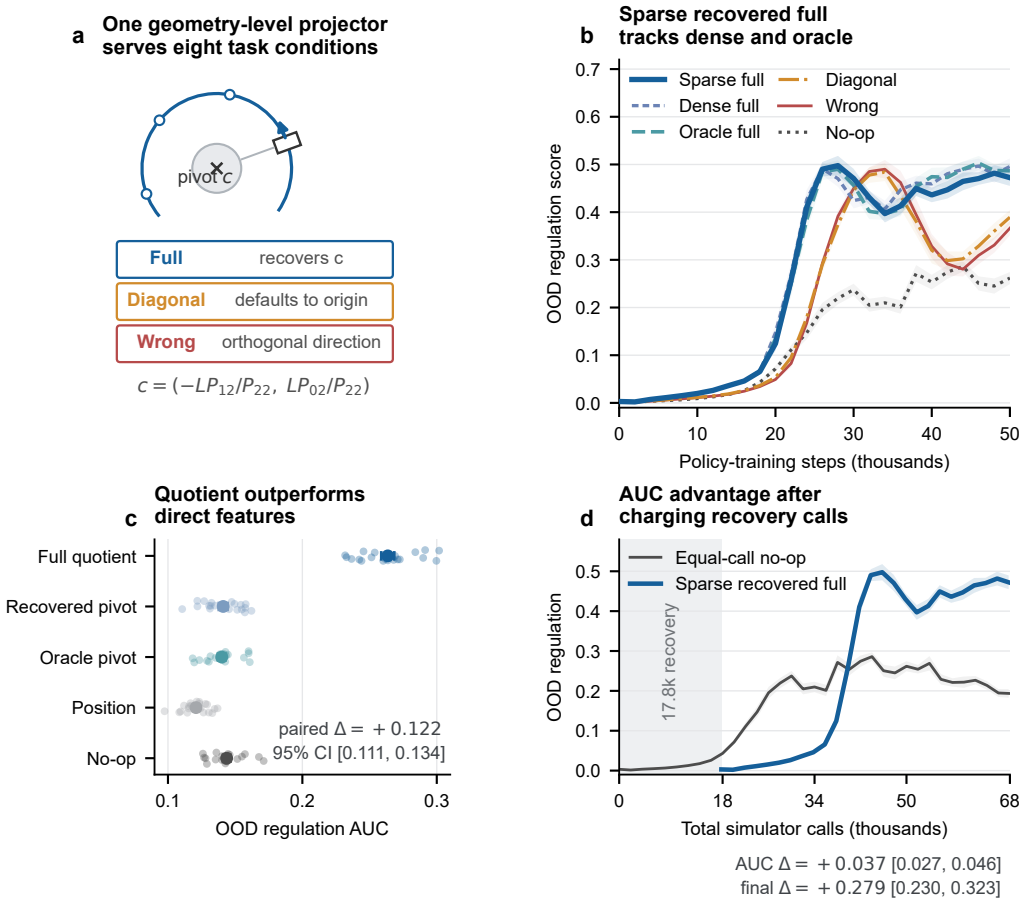

Figure 4: **Geometry-level projector reuse for orbital-contact control. a,** One projector recovered at a canonical condition per geometry is reused across four start phases and two travel directions. **b,** OOD learning curves averaged across eight conditions at each of 12 policy-held-out centers. **c,** Quotient versus direct-feature controls in a separate matched 20-seed cohort. **d,** Recovered full versus call-count-matched no-op, charging all 17,784 recovery calls to each recovered-full policy. Curves in b and d show means over 20 paired seeds, with ± s.e.m. bands. Panel c shows individual seeds (faint points) and means ± s.e.m. (large points and error bars). The annotated 95% CI in c uses a paired seed bootstrap for full quotient minus recovered-pivot features. In d, annotated 95% CIs use a paired two-way bootstrap over seeds and centers for recovered full minus no-op.

Sparse and dense recovery both achieved 100% contact-interior coverage and rank accuracy, rejected every outward guard, and obtained 90th-percentile angles of $0.89°$ (sparse) and $1.08°$ (dense).

Online control. The online comparison used 20 paired seeds, 50k SAC steps, identical networks and optimizers, and one quotient implementation. Sparse full, diagonal and wrong shared the recovered table and differed only in operator. Dense recovery changed table granularity, and sparse oracle full supplied analytic projectors at the same sparse keys. The wrapper used exact geometry identifiers for table lookup, leaving identifiers and centers outside actor and critic inputs. Training used no counterfactual queries. Missing or uncertified keys used the same untransformed interface as no-op. The primary score combines contact occupancy, normalized orbital progress and terminal goal accuracy. Its normalized policy-held-out OOD AUC is the trapezoidal area over the fixed 0–50k grid divided by 50,000.

Sparse recovered full reached 0.268 OOD AUC (Table 1; Fig. 4b). Its paired gains were $+0.131$ over no-op (95% CI $[0.119, 0.142]$), $+0.062$ over diagonal ($[0.049, 0.075]$) and $+0.064$ over wrong ($[0.053, 0.075]$). Return and success AUC also favored full over all three controls (Appendix Table 3).

Using the prespecified 0.02 AUC margin, sparse recovered full was non-inferior to dense recovery, with a paired difference of -0.006 (95% CI $[-0.015, 0.003]$). Its difference from sparse oracle full was -0.002 ($[-0.012, 0.008]$).

A post hoc phase decomposition localized the full-versus-no-op advantage primarily to unseen start phases (Appendix Table 4). On held-out centers, their AUCs were 0.301 versus 0.045 at unseen phases, compared with 0.236 versus 0.230 at training phases.

Representation control. In a separate matched representation-control cohort, all branches shared input dimensionality, parameter count, initialization and SAC training. The full quotient and recovered-pivot feature control used the same recovered geometry and certification support. The quotient gained $+0.122$ OOD AUC over recovered-pivot features (95% paired seed-bootstrap CI $[0.111, 0.134]$; Fig. 4c). Neither recovered- nor oracle-pivot features reproduced the benefit of explicit canonicalization and action transport under this matched MLP control.

Recovery cost. Geometry-level reuse reduced recovery from 192 condition-specific keys (142,272 calls) to 24 keys (17,784 calls), an 87.5% reduction. The sparse cost equals 35.6% of one 50k-step policy run.

The call-count-matched comparison charged all 17,784 recovery calls to each recovered-full policy and gave no-op SAC 17,784 additional ordinary interactions and updates. At the shared 67,784-call budget, recovered full retained a $+0.037$ normalized total-call OOD AUC advantage (95% CI $[0.027, 0.046]$) and a $+0.279$ final OOD score advantage ($[0.230, 0.323]$; Fig. 4d). Appendix C.2 gives the complete task, representation and resource definitions.

Table 1: **MuJoCo orbital-contact control.** Mean $\pm$ s.e.m. over 20 paired seeds; AUC integrates the fixed 0–50k grid. Seen-center evaluation uses unseen start phases; OOD uses all four phases at 12 policy-held-out centers independently certified before training.

Branch	Seen-center AUC	OOD AUC	Final OOD
No-op	0.045 ± 0.003	0.137 ± 0.004	0.262 ± 0.013
Sparse recovered diagonal	0.209 ± 0.004	0.206 ± 0.003	0.389 ± 0.015
Sparse recovered wrong	0.208 ± 0.003	0.204 ± 0.003	0.366 ± 0.015
Sparse oracle full	0.305 ± 0.006	0.270 ± 0.006	0.487 ± 0.016
Dense condition-specific recovered full	0.307 ± 0.006	0.274 ± 0.006	0.496 ± 0.017
Sparse recovered full	0.301 ± 0.007	0.268 ± 0.005	0.472 ± 0.017

5 Discussion and Conclusion

We introduced a simulator-assisted estimator that connects event-conditioned local symmetry projectors to a shared actor–critic quotient. In orbital contact, one projector recovered at a canonical condition per center served eight task conditions, reducing recovery calls by 87.5%. This is within-geometry condition reuse: all evaluation centers were independently certified before policy training. Recovered full retained higher final performance and normalized total-call OOD AUC than no-op after all recovery calls were charged.

In the orbital task, the recovered rank-one projector determines the pivot used by the shared quotient. Deleting its cross terms yields an origin-centred chart, while the matched orthogonal rank-one control changes the generator assignment. These controls test geometric decoding, while the matched feature control tests how the decoded geometry enters learning. The phase-resolved results place the largest advantage over no-op at start phases absent from policy training. Together, these findings support organizing actor–critic learning through recovered orbital coordinates in this benchmark.

AI USE STATEMENT

Generative AI tools were used only for language polishing, literature search, and routine code support. All research ideas, theoretical derivations, methodological contributions, and experimental analyses were developed by the authors. All content assisted by AI was carefully reviewed and verified by the authors, who take full responsibility for the final work.

REPRODUCIBILITY STATEMENT

Appendices A–C provide proofs, configurations, statistical procedures, recovery audits and simulator-call accounting. The anonymous supplement provides source code, predefined splits, policy seeds and required recovery inputs.

REFERENCES

Manu Bhat, Jonghyun Park, Jianke Yang, Nima Dehmamy, Robin Walters, and Rose Yu. AtlasD: Automatic local symmetry discovery. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 4153–4171. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/bhat25a.html.

Samuel A. Burden, S. Shankar Sastry, Daniel E. Koditschek, and Shai Revzen. Event–selected vector field discontinuities yield piecewise–differentiable flows. *SIAM Journal on Applied Dynamical Systems*, 15(2):1227–1267, 2016. doi: 10.1137/15M1016588. URL https://doi.org/10.1137/15M1016588.

Junwoo Chang, Minwoo Park, Joohwan Seo, Roberto Horowitz, Jongmin Lee, and Jongeun Choi. Partially equivariant reinforcement learning in symmetry-breaking environments. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=dRDcVyobhH.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In Maria Florina Balcan and Kilian Q. Weinberger (eds.), *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 2990–2999, New York, New York, USA, 20–22 Jun 2016. PMLR. URL https://proceedings.mlr.press/v48/cohenc16.html.

Nima Dehmamy, Robin Walters, Yanchen Liu, Dashun Wang, and Rose Yu. Automatic symmetry discovery with lie algebra convolutional network. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 2503–2515. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/148148d62be67e0916a833931bd32b26-Paper.pdf.

Rafal Goebel, Ricardo G. Sanfelice, and Andrew R. Teel. *Hybrid Dynamical Systems: Modeling, Stability, and Robustness*. Princeton University Press, 2012. ISBN 9781400842636. doi: 10.23943/princeton/9780691153896.001.0001. URL https://academic.oup.com/princeton-scholarship-online/book/21191.

Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1861–1870. PMLR, 2018. URL https://proceedings.mlr.press/v80/haarnoja18b.html.

Lexiang Hu, Yikang Li, and Zhouchen Lin. Symmetry discovery for different data types. *Neural Networks*, 188:107481, 2025. ISSN 0893-6080. doi: 10.1016/j.neunet.2025.107481. URL https://www.sciencedirect.com/science/article/pii/S0893608025003600.

Sékou-Oumar Kaba and Siamak Ravanbakhsh. Symmetry breaking and equivariant neural networks. *arXiv preprint arXiv:2312.09016*, 2023. URL https://arxiv.org/abs/2312.09016.

Sékou-Oumar Kaba, Arnab Kumar Mondal, Yan Zhang, Yoshua Bengio, and Siamak Ravanbakhsh. Equivariance with learned canonicalization functions. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 15546–15566. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/kaba23a.html.

Gyeonghoon Ko, Hyunsu Kim, and Juho Lee. Learning infinitesimal generators of continuous symmetries from data. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 85973–86003. Curran Associates, Inc., 2024. doi: 10.52202/079017-2730. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/9c26a48b7ebd37bc3e979c8651fc3649-Paper-Conference.pdf.

Nathan J. Kong, J. Joe Payne, James Zhu, and Aaron M. Johnson. Saltation matrices: The essential tool for linearizing hybrid dynamical systems. *Proceedings of the IEEE*, 112(6):585–608, June 2024. doi: 10.1109/JPROC.2024.3440211. URL https://ieeexplore.ieee.org/document/10638633.

Haoran Li, Chenhan Xiao, Muhao Guo, and Yang Weng. Latent mixture of symmetries for sample-efficient dynamic learning. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, pp. 108756–108785. Curran Associates, Inc., 2025. doi: 10.52202/085713-3626. URL https://openreview.net/forum?id=t7NLyOtPEi.

Paul Nitschke and Shahriar Talebi. Hereditary geometric meta-rl: Nonlocal generalization via task symmetries. *arXiv preprint arXiv:2603.00396*, 2026. URL https://arxiv.org/abs/2603.00396.

Jung Yeon Park, Sujay Bhatt, Sihan Zeng, Lawson L.S. Wong, Alec Koppel, Sumitra Ganesh, and Robin Walters. Approximate equivariance in reinforcement learning. In Yingzhen Li, Stephan Mandt, Shipra Agrawal, and Emtiyaz Khan (eds.), *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 4177–4185. PMLR, 03–05 May 2025. URL https://proceedings.mlr.press/v258/park25d.html.

David W. Romero and Suhas Lohit. Learning partial equivariances from data. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 36466–36478. Curran Associates, Inc., 2022. doi: 10.52202/068431-2642. URL https://openreview.net/forum?id=pNHT6oBaPr8.

Chenrui Tie, Yue Chen, Ruihai Wu, Boxuan Dong, Zeyi Li, Chongkai Gao, and Hao Dong. ETSEED: EFFICIENT TRAJECTORY-LEVEL SE(3) EQUIVARIANT DIFFUSION POLICY. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=OheAR2xrtb.

Emanuel Todorov, Tom Erez, and Yuval Tassa. MuJoCo: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5026–5033. IEEE, 2012. doi: 10.1109/IROS.2012.6386109.

Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012. doi: 10.1007/s10208-011-9099-z. URL https://arxiv.org/abs/1004.4389.

Elise van der Pol, Daniel E. Worrall, Herke van Hoof, Frans A. Oliehoek, and Max Welling. MDP homomorphic networks: Group symmetries in reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 4199–4210. Curran Associates, Inc., 2020. URL https://arxiv.org/abs/2006.16908.

Dian Wang, Robin Walters, and Robert Platt. SO(2)-Equivariant Reinforcement Learning. In *International Conference on Learning Representations*, 2022a. URL https://arxiv.org/abs/2203.04439.

Rui Wang, Robin Walters, and Rose Yu. Approximately equivariant networks for imperfectly symmetric dynamics. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 23078–23091. PMLR, 17–23 Jul 2022b. URL https://proceedings.mlr.press/v162/wang22aa.html.

Jianke Yang, Robin Walters, Nima Dehmamy, and Rose Yu. Generative adversarial symmetry discovery. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 39488–39508. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/yang23n.html.

Jianke Yang, Nima Dehmamy, Robin Walters, and Rose Yu. Latent space symmetry discovery. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 56047–56070. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/yang24g.html.

Jingyun Yang, Ziang Cao, Congyue Deng, Rika Antonova, Shuran Song, and Jeannette Bohg. EquiBot: SIM(3)-equivariant diffusion policy for generalizable and data efficient learning. In Pulkit Agrawal, Oliver Kroemer, and Wolfram Burgard (eds.), *Proceedings of The 8th Conference on Robot Learning*, volume 270 of *Proceedings of Machine Learning Research*, pp. 1048–1068. PMLR, 2025. URL https://proceedings.mlr.press/v270/yang25a.html.

Beomyeol Yu and Taeyoung Lee. Equivariant reinforcement learning for quadrotor UAV. In *2023 American Control Conference (ACC)*, pp. 2842–2847. IEEE, 2023. doi: 10.23919/ACC55779.2023.10156379. URL https://ieeexplore.ieee.org/document/10156379/.

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://arxiv.org/abs/1405.0680.

Kevin Zeng and Michael D. Graham. Symmetry reduction for deep reinforcement learning active control of chaotic spatiotemporal dynamics. *Physical Review E*, 104(1):014210, 2021. doi: 10.1103/PhysRevE.104.014210.

A Detailed Method

This appendix gives the full query construction, typed event residuals, spectral operators and shared quotient. It also states the stability results and their assumptions. The secondary wall critic interface appears with its experimental details in Appendix C.5.

A.1 Problem formulation

We consider a deterministic hybrid controlled system with continuous state $s \in \mathcal{S}$, discrete mode $m \in \mathcal{M}$, and action $a \in \mathcal{A}$. A zero-order-hold control is applied for a decision interval Δt. Continuous evolution, guard crossings, and reset maps over this interval induce a hybrid flow map

$$\mathcal{X} = \mathcal{S} \times \mathcal{M}, \qquad \mathcal{Z} = \mathcal{X} \times \mathcal{A}, \qquad \Phi_{\Delta t} : \mathcal{Z} \longrightarrow \mathcal{X}, \qquad (s, m, a) \longmapsto (s^+, m^+). \tag{14}$$

We write

$$\Phi^X_{\Delta t} := \Phi_{\Delta t}, \qquad \Phi^S_{\Delta t} := \mathrm{pr}_{\mathcal{S}} \circ \Phi^X_{\Delta t} \tag{15}$$

for the full state-mode endpoint and its continuous-state component, respectively. The corresponding discrete transition kernel is $P_{\Delta t}(\cdot \mid s, m, a) = \delta_{\Phi^X_{\Delta t}(s,m,a)}$. Reinforcement learning is performed on this induced discrete-time Markov decision process, with discount $\gamma = \exp(-\rho \Delta t)$ for a continuous-time discount rate $\rho > 0$.

We consider a supplied candidate Lie algebra $\mathfrak{g}_0$ of dimension d, which defines the transformation family within which locally valid directions are estimated. Let $\{e_j\}_{j=1}^d$ be a basis of $\mathfrak{g}_0$, and let

$$\xi^S_j(s) = \frac{\mathrm{d}}{\mathrm{d}\epsilon} T^S_{\epsilon e_j}(s)\Big|_{\epsilon=0}, \qquad \xi^A_j(a) = \frac{\mathrm{d}}{\mathrm{d}\epsilon} T^A_{\epsilon e_j}(a)\Big|_{\epsilon=0} \tag{16}$$

denote its infinitesimal actions on states and actions. We consider mode-preserving transformations. Their induced actions on the state-mode and state-mode-action spaces are

$$T^X_g(s, m) = \left(T^S_g s, m\right), \qquad T^Z_g(s, m, a) = \left(T^S_g s, m, T^A_g a\right). \tag{17}$$

We consider state-independent group actions on the action space, $T^A_g(a)$, with generators $\xi^A_j(a)$. For $u = (u_1, \ldots, u_d)^\top \in \mathbb{R}^d$, we write

$$\xi^S_u(s) = \sum_{j=1}^d u_j \xi^S_j(s), \qquad \xi^A_u(a) = \sum_{j=1}^d u_j \xi^A_j(a). \tag{18}$$

The candidate Lie algebra is equipped with a positive-definite metric $M_{\mathfrak{g}_0} \in \mathbb{R}^{d \times d}$. Physical quantities are made dimensionless before this metric is chosen. For example, translational generators are scaled by a reference length L_{ref}. Under passive $M_{\mathfrak{g}_0}$-orthonormal basis changes, matrix representations transform covariantly.

Our objective is to estimate, at each state-action-mode tuple $z = (s, m, a)$, the subspace of candidate coefficients that are locally consistent with the hybrid dynamics and event topology:

$$\mathcal{D}(z) \subseteq \mathfrak{g}_0. \tag{19}$$

Accordingly, $\mathcal{D}(z)$ is treated as a pointwise linear subspace, independently of Lie-subalgebra closure, integrability or the existence of a local group action.

Simulator-assisted setting. The method uses a deterministic reset-and-step simulator to evaluate transformed state-action pairs and report the contact mode and event signature observed during each step. Simulator rollouts provide the flow and event responses used by the estimator.

A.2 Counterfactual hybrid flow-map defects

Let

$$y = \Phi^S_{\Delta t}(s, m, a) \tag{20}$$

be the continuous-state component of the nominal endpoint. For a basis direction e_j, define the transformed inputs

$$z_j^{\pm} = \left(T^S_{\pm\epsilon e_j} s,\, m,\, T^A_{\pm\epsilon e_j} a\right), \qquad y_j^{\pm} = \Phi^S_{\Delta t}(z_j^{\pm}). \tag{21}$$

When the nominal and perturbed rollouts have the same event topology, the centered hybrid flow-map defect is

$$e_j^{\Phi}(z) = \frac{\mathrm{Log}_y(y_j^+) - \mathrm{Log}_y(y_j^-)}{2\epsilon} - \xi_j^S(y) \in T_y\mathcal{S}. \tag{22}$$

Here Log_y maps a nearby state to the tangent space at y. It is implemented componentwise: Euclidean differences are used for translational variables, wrapped differences for S^1 variables, and the Lie logarithm for $SO(3)$ or $SE(3)$ components. In Euclidean state spaces, Eq. (22) reduces to

$$e_j^{\Phi}(z) = \frac{\Phi^S_{\Delta t}(z_j^+) - \Phi^S_{\Delta t}(z_j^-)}{2\epsilon} - \xi_j^S\left(\Phi^S_{\Delta t}(z)\right). \tag{23}$$

Thus e_j^{Φ} estimates $D\Phi_{\Delta t}[\xi_j^S; \xi_j^A] - \xi_j^S \circ \Phi_{\Delta t}$ without differentiating through the simulator. It is the flow-level analogue of a continuous-time Lie-bracket residual.

Collecting the defects gives

$$E_{\Phi}(z) = \begin{bmatrix} e_1^{\Phi}(z) & \cdots & e_d^{\Phi}(z) \end{bmatrix} \in \mathbb{R}^{n_T \times d}, \tag{24}$$

where $n_T = \dim T_y\mathcal{S}$.

A.3 Event-conditioned compatibility

Hybrid flow maps are differentiable within strata that share a stable event topology. Obstacle and wall recovery evaluate all signed basis probes at three perturbation scales and classify each query set as follows:

1. *accepted*: every transformed query is valid, and both signs of every required directional perturbation retain the required event signature at every declared radius;
2. *unknown*: transform invalidity, grazing contact, simultaneous-new-contact ambiguity, a single-scale mismatch, or scale-dependent event topology prevents a reliable local linearization;
3. *event breaking*: a stable, scale-consistent perturbation pattern changes the discrete event sequence.

Unknown and event-breaking query sets are excluded from the obstacle and wall estimates. Figure 1 groups these two rejection classes under "unknown".

Orbital recovery applies directionwise screening within an overdetermined design. At each of two radii, five coefficient frames supply 37 directions each. A direction is retained when both signs have finite flow residuals and rewards, matching nominal event and termination signatures, and penetration within the safety limit. The signature records contact, physical contact, contact-loss termination, termination and truncation. Retained directions from all frames are pooled at each radius. Failure to span the three-dimensional algebra at either radius produces geometry-level abstention; a rejected direction alone does not. Termination changes or penetration above the limit anywhere in the queried set veto the geometry. Scale-rank, scale-angle and projected-reward checks then determine table admission, as specified in Algorithm 1.

For orbital state components (p, θ, v, ω), the endpoint difference in Eq. (3) is

$$\rho(z, tu) = \left(\Delta p,\; L\,\mathrm{wrap}(\Delta\theta),\; t_v\Delta v,\; Lt_v\Delta\omega\right), \qquad L = 0.35\ \mathrm{m}, \quad t_v = 0.015\ \mathrm{s}, \tag{25}$$

where each difference compares the queried endpoint with the transformed nominal endpoint. A unit whitened probe u is converted to physical coefficients $\mathrm{diag}(1, 1, 1/L)u$ before the spatial transformation. The signed residuals in Eq. (3) therefore include the finite transformation of the nominal output. For the same retained pairs, let the row vector q_ϵ contain $[r(T_{\epsilon u}z) - r(T_{-\epsilon u}z)]/(2\epsilon)$. The reward fit is $\widehat{\widehat{g}}_{r,\epsilon} = (q_\epsilon U_\epsilon^{\dagger})^{\top}$, and $\widehat{\widehat{g}}_r$ averages these fits over the two radii.

For accepted obstacle and wall transitions, the simulator additionally returns a typed continuous event descriptor

$$q_{\mathrm{evt}}(z) = \big(t_{\mathrm{evt}}, j_{\mathrm{evt}}, p_{\mathrm{evt}}, n_{\mathrm{evt}}, h_{\mathrm{evt}}\big), \tag{26}$$

where the components denote normalized event time, contact impulse, contact location, contact normal, and any auxiliary feature with a known transformation type, respectively. Only features available and comparable for the given event type are used. No-event samples are routed separately because their event times are censored rather than observed. Each component is assigned a known representation under the candidate transformation. Event time and explicitly invariant auxiliary features transform as scalars, contact locations transform as positions, and impulses and normals transform as vectors.

Let T_g^{evt} denote the resulting typed action on the event descriptor and let

$$\zeta_j^{\mathrm{evt}}(q) = \frac{\mathrm{d}}{\mathrm{d}\epsilon} T^{\mathrm{evt}}_{\epsilon e_j} q \Big|_{\epsilon=0} \tag{27}$$

be its infinitesimal generator. The typed event-equivariance defect is estimated by

$$e_j^{\mathrm{evt}}(z) = \frac{q_{\mathrm{evt}}(z_j^+) - q_{\mathrm{evt}}(z_j^-)}{2\epsilon} - \zeta_j^{\mathrm{evt}}(q_{\mathrm{evt}}(z))\,. \tag{28}$$

Differences in Eq. (28) are computed in the appropriate typed tangent coordinates. Collecting columns gives

$$E_{\mathrm{evt}}(z) = \begin{bmatrix} e_1^{\mathrm{evt}}(z) & \cdots & e_d^{\mathrm{evt}}(z) \end{bmatrix} \in \mathbb{R}^{n_{\mathrm{evt}} \times d}. \tag{29}$$

Subtracting the expected output generator preserves the covariance of typed event quantities. For example, a valid translation may move a contact location, and a valid rotation may rotate a contact impulse, while both yield zero infinitesimal typed defect. As with an ordinary Jacobian Gram, linear combinations of candidate directions remain recoverable within a fixed event-topology stratum.

Let $W_\Phi(z)$ and $W_{\mathrm{evt}}(z)$ be positive-semidefinite residual-space weights. The pointwise dynamics-event Gram matrices are

$$G_{\mathrm{int}}(z) = E_\Phi(z)^\top W_\Phi(z) E_\Phi(z), \tag{30}$$

$$G_{\mathrm{evt}}(z) = G_{\mathrm{int}}(z) + \lambda_{\mathrm{evt}} E_{\mathrm{evt}}(z)^\top W_{\mathrm{evt}}(z) E_{\mathrm{evt}}(z). \tag{31}$$

A covariance-whitened implementation may set $W_\bullet = (\Sigma_\bullet + \lambda_\bullet I)^{-1}$ using covariance estimates from repeated queries or an ensemble. The obstacle and wall pipelines use fixed identity-scaled flow and event weights (both 1), three perturbation scales and event-compatible neighbourhood aggregation. The orbital pipeline fits the six flow channels in Eq. (25) with identity weighting. Event signatures determine which directional pairs enter the fit. The pipeline pools frames before fitting at each radius, then averages the trace-normalized radius Grams. Rewards enter the separate projected-gradient check rather than these Grams.

A.4 Neighborhood dynamics Grams for obstacle and wall recovery

The obstacle and wall pipelines estimate symmetry over event-compatible local neighborhoods. We define a dimensionless state-action metric

$$M_{sa} = \mathrm{diag}\left(L_{\mathrm{ref}}^{-2} I_{\mathrm{pos}}, I_{\mathrm{ang}}, v_{\mathrm{ref}}^{-2} I_{\mathrm{vel}}, a_{\mathrm{ref}}^{-2} I_{\mathrm{act}}\right) \tag{32}$$

and use either k-nearest neighbors or a compact kernel under this metric. For query z, let $\mathcal{N}(z)$ contain neighbors with the same mode and compatible event type, and let normalized kernel weights be $w(z, z')$. The local dynamics Gram is

$$H_{\mathrm{dyn}}(z) = \sum_{z' \in \mathcal{N}(z)} w(z, z') \begin{cases} G_{\mathrm{int}}(z'), & \text{no event,} \\ G_{\mathrm{evt}}(z'), & \text{accepted event.} \end{cases} \tag{33}$$

Unknown and event-breaking samples receive zero regularization confidence and are excluded from the local spectral estimate. The orbital pipeline instead pools retained multiframe directions into one pointwise residual-map fit per radius. The average of its normalized radius Grams proceeds directly to the spectral step in Sec. A.5.

Task compatibility. A dynamics-compatible direction need not preserve the task when rewards or policies break the same transformation. The principal orbital pipeline assesses task compatibility through finite reward, event, termination and safety checks before quotient use, and shares the accepted chart across the actor, critics, targets and action transport.

A.5 Spectral subspace and confidence operators

For obstacle and wall recovery, $H_{\mathrm{dyn}}(z)$ is the neighborhood Gram in Eq. (33). Orbital recovery directly diagonalizes the normalized average $\widehat{\bar{H}}$ from Algorithm 1 in whitened coordinates, where the metric is I_3. Its equivalent Gram in physical coefficient coordinates is $H_{\mathrm{dyn}} = M_{\mathfrak{g}_0}^{1/2} \widehat{\bar{H}} M_{\mathfrak{g}_0}^{1/2}$. In those coordinates, both pipelines have the generalized eigenproblem

$$H_{\mathrm{dyn}}(z) v_i = \sigma_i M_{\mathfrak{g}_0} v_i, \qquad V^\top M_{\mathfrak{g}_0} V = I. \tag{34}$$

The hard projector used for subspace recovery and orbital quotient control is

$$P(z) = V_{\le\tau} V_{\le\tau}^\top M_{\mathfrak{g}_0}, \tag{35}$$

where $V_{\le\tau}$ contains generalized eigenvectors with $\sigma_i \le \tau$. Its $M_{\mathfrak{g}_0}$-orthonormal columns give $P^2 = P$ and $P^\top M_{\mathfrak{g}_0} = M_{\mathfrak{g}_0} P$. For orbital recovery, each radius Gram and their normalized average use $\tau(\bar{H}) = \max\{10^{-12}, 0.015\lambda_{\max}(\bar{H})\}$. The two radius projectors and the averaged projector must each have rank one, with at most 10° between the radius projectors. Per-frame estimates provide design-sensitivity diagnostics. Table admission uses the pooled radius estimates. The secondary wall interface also considers the soft confidence operator

$$S(z) = V \,\mathrm{diag}\left[\mathrm{sigmoid}\left(\frac{\tau - \sigma_i}{T}\right)\right] V^\top M_{\mathfrak{g}_0}. \tag{36}$$

It satisfies $0 \preceq_{M_{\mathfrak{g}_0}} S \preceq_{M_{\mathfrak{g}_0}} I$ and $S^\top M_{\mathfrak{g}_0} = M_{\mathfrak{g}_0} S$, but generally $S^2 \ne S$. Thus S weights directions continuously, while P selects a subspace; away from the threshold, S converges to P as $T \to 0$. The wall amortization baselines learn scalar or diagonal summaries from hard P labels, and direct wall lookup uses P. Operators in the wall critic loss are stopped-gradient quantities. The orbital quotient uses a fixed hard-projector table throughout policy training.

Scalar and diagonal summaries. For a unit wall tangent $t \in \mathbb{R}^2$, the best scalar approximation to $tt^\top$ is $\lambda^\star I_2$ with $\lambda^\star = 1/2$ and error $\|\lambda^\star I_2 - tt^\top\|_F = \sqrt{1/2}$. A fixed-basis diagonal summary retains coordinate weights but removes the cross terms that couple a rotated subspace.

A.6 Projector-conditioned state canonicalization

For physical planar $\mathfrak{se}(2)$ coefficients $u = (u_x, u_y, u_\theta)^\top$, orbital recovery uses whitened coefficients $\bar{u} = (u_x, u_y, Lu_\theta)^\top$ with $L = 0.35$ m. Its deployable policy observation is

$$o = (b, g, v, \omega, m_c) \in \mathbb{R}^8, \qquad b = (\cos\theta, \sin\theta), \tag{37}$$

where g is goal displacement, v is linear velocity, ω is angular velocity and m_c is the contact indicator. Obstacle center, contact normal, force and clearance are excluded. Body position $p \in \mathbb{R}^2$ is carried separately for deterministic coordinate conversion and is not passed to a learned network.

For a symmetric branch operator $A \in \mathbb{R}^{3\times3}$ with a nonzero rotational anchor, define

$$c_A = \left(-L\frac{A_{12}}{A_{22}}, L\frac{A_{02}}{A_{22}}\right), \qquad \phi_A = \mathrm{atan2}((p - c_A)_y, (p - c_A)_x), \qquad p \ne c_A. \tag{38}$$

Let $R_- = R(-\phi_A)$. Every actor, online critic and target critic receives

$$\bar{o}_A = (R_- b, R_- g, R_- v, \omega, m_c), \qquad \bar{a}_A = (R_- f, \tau), \tag{39}$$

for world wrench $a = (f, \tau)$. The actor outputs a canonical wrench and the interface transports its planar force back by $R(\phi_A)$ while leaving the torque unchanged. The map is applied independently to current and next replay states. Raw transitions remain in replay, and there is no symmetry loss,

transformed-transition augmentation or additional coefficient. If A is zero or lacks a reliable rotational anchor, the map uses $\phi_A = 0$ and is exactly the no-op interface.

For a contact context with pivot c, the whitened unit orbital direction and oracle projector are

$$\bar{v}_c = \frac{(c_y, -c_x, L)^\top}{\sqrt{c_x^2 + c_y^2 + L^2}}, \qquad P_c = \bar{v}_c \bar{v}_c^\top. \tag{40}$$

The recovered-full branch supplies the fixed $\widehat{P}_c$. The diagonal and trace-scalar controls supply $\mathrm{diag}(\widehat{P}_c)$ and $\mathrm{tr}(\widehat{P}_c)I_3/3$; both induce the origin-centered chart under the shared implementation. For $e_r = (0,0,1)^\top$, the wrong direction is $w = (I_3 - \widehat{P}_c)e_r / \|(I_3 - \widehat{P}_c)e_r\|_2$ and $A_{\mathrm{wrong}} = ww^\top$. Thus full and wrong match rank, spectrum, trace and Frobenius norm. Missing, rejected and contact-inactive geometry keys return 0.

Finite orbital transformations and action feasibility. For a fixed center c, rotation by α defines the one-parameter action

$$p \mapsto c + R(\alpha)(p - c), \qquad (b, g, v, f) \mapsto (R(\alpha)b, R(\alpha)g, R(\alpha)v, R(\alpha)f). \tag{41}$$

Angular velocity and torque are unchanged, and successive rotations compose by angle addition. Rotating the goal and initial phase along with the body preserves relative goal error, signed progress and radial deviation. The circular obstacle, isotropic planar mechanics and wrench norms are also preserved. Consequently, the task reward terms in Appendix C.2 and the geometric contact and success conditions are invariant under this fixed-center action.

The policy maps Gaussian latent coordinates z to a physical wrench using

$$f = 3\,\frac{z_{xy}}{\sqrt{1 + \|z_{xy}\|_2^2}}, \qquad \tau = 0.16 \tanh z_\tau. \tag{42}$$

Canonical-to-world rotation preserves the force disk and torque interval without component-wise clipping and has unit determinant at fixed state. SAC uses the physical-wrench log density with exact Jacobian corrections for the radial force map and torque squash.

Local probes estimate the fixed-center orbital axis and check numerical flow, event and first-order reward compatibility at the canonical condition. Sparse-reuse rollouts evaluate the resulting approximate chart across the remaining phases and travel directions.

A.7 Theoretical analysis

A.7.1 Subspace recovery

Let $H_\star$ be the declared population Gram target and $\widehat{H}$ its estimate. Their whitened forms are

$$\bar{H}_\star = M_{\mathfrak{g}_0}^{-1/2} H_\star M_{\mathfrak{g}_0}^{-1/2}, \qquad \widehat{\bar{H}} = M_{\mathfrak{g}_0}^{-1/2} \widehat{H} M_{\mathfrak{g}_0}^{-1/2}. \tag{43}$$

For perturbation norm at most δ, Lemma 1 controls the bottom-k subspace at fixed rank $0 < k < d$. Threshold-selected ranks additionally require eigenvalues to remain on their original sides of the cutoff, including when $k = 0$ or $k = d$.

For a fixed cutoff τ, Weyl's inequality gives the sufficient margin $\min_i |\lambda_i(\bar{H}_\star) - \tau| > \delta$. For $\tau(H) = \max\{a, b\lambda_{\max}(H)\}$, its Lipschitz constant is b, hence a sufficient margin is

$$\min_i \big|\lambda_i(\bar{H}_\star) - \tau(\bar{H}_\star)\big| > (1 + b)\delta. \tag{44}$$

Indeed, the eigenvalue and cutoff change by at most δ and $b\delta$, respectively. Orbital recovery uses $a = 10^{-12}$ and $b = 0.015$.

A.7.2 Conditional finite-query and local-sampling error

We next expose sufficient conditions for the perturbation δ in Lemma 1. We distinguish the pointwise population Gram $H_\star^{\mathrm{pt}}(z) := G_\star(z)$ from the neighborhood-averaged population Gram $H_{\star,r}^{\mathrm{loc}}(z) :=$

Algorithm 1 Orbital geometry-level recovery and shared quotient control

Require: simulator $\mathcal{S}$, geometry keys $\mathcal{G}$, canonical contexts, frames $\mathcal{F}$, radii $\mathcal{E}$, projective directions

Initialize an empty geometry-key table $\mathcal{T}$
for each geometry key $g \in \mathcal{G}$ **do**
  Restore the canonical snapshot; check nominal contact and penetration
  if the nominal check fails **then**
    Store abstain for geometry g
    continue
  end if
  Step the nominal zero action; query both signs of all frame–radius directions
  Record finite residuals, rewards, event signatures, termination and penetration
  eligible $\leftarrow$ all probes preserve termination and satisfy the penetration limit
  for each radius $\epsilon \in \mathcal{E}$ **do**
    Retain valid signed pairs and pool all frames into $(U_\epsilon, R_\epsilon, q_\epsilon)$
    if $\mathrm{rank}(U_\epsilon) < 3$ **then**
      eligible $\leftarrow$ false; **break**
    end if
    $\widehat{E}_\epsilon = R_\epsilon U_\epsilon^\dagger$; $\widehat{\bar{H}}_\epsilon = \widehat{E}_\epsilon^\top \widehat{E}_\epsilon$
    $\widehat{\bar{g}}_{r,\epsilon} = (q_\epsilon U_\epsilon^\dagger)^\top$; threshold $\widehat{\bar{H}}_\epsilon$ to obtain $\bar{P}_\epsilon$
  end for
  if not eligible **then**
    Store abstain for geometry g; **continue**
  end if
  Average normalized radius Grams and reward gradients to obtain $(\widehat{\bar{H}}, \widehat{\bar{g}}_r)$
  Threshold $\widehat{\bar{H}}$ to obtain $\bar{P}$; check rank one, radius angle and Eq. (8)
  if these checks pass **then**
    Store $(g, \bar{P})$ in $\mathcal{T}$
  else
    Store abstain for geometry g
  end if
end for
for each SAC rollout or replay update **do**
  Map each task context to its geometry key; missing, inactive or uncertified keys apply no quotient
  Extract the pivot and apply the same operator-conditioned chart to current and next observations
  Share the chart across actor, online critics and target critics; transport the selected canonical-frame action back to the world frame
end for

$\mathbb{E}_{\mu_{z,r}}[G_\star(Z)]$. The result bounds the estimator relative to the pointwise target, with the difference between these two population objects appearing explicitly as neighborhood bias. It is conditional on a fixed event-topology stratum and a common accepted neighborhood, so it does not quantify the probability of acceptance or selection bias outside that stratum.

For an accepted tuple z, let $\psi_z : B_{\epsilon_0}(0) \subset \mathbb{R}^d \to \mathbb{R}^p$ denote the stacked transformed simulator channels in a common typed tangent chart. The stack contains the flow endpoint and the event channels present for the fixed event type, after the dimensionless channel scaling used by the estimator. Let $B_z : \mathbb{R}^d \to \mathbb{R}^p$ be the linear infinitesimal action expected at the transformed output. The population defect differential is

$$E_\star(z) = D\psi_z(0) - B_z. \tag{45}$$

Writing $\widehat{\psi}_z$ and $\widehat{B}_z$ for their numerically evaluated counterparts, the centered estimate has columns

$$\widehat{E}_\epsilon(z)e_j = \frac{\widehat{\psi}_z(\epsilon e_j) - \widehat{\psi}_z(-\epsilon e_j)}{2\epsilon} - \widehat{B}_z e_j. \tag{46}$$

Equations (22) and (28) are the corresponding channelwise implementation.

Assumption A.1 (Shared mode-interior validity)**.** *There is an event-compatible neighborhood $\mathcal{N}_r(z) \subseteq \{z' : \mathrm{d}_{sa}(z', z) \le r\}$ and a radius $\epsilon_0 > 0$. Every perturbation used in the compared coefficient bases remains transform valid, retains one common event signature, and lies in one typed tangent chart for all $z' \in \mathcal{N}_r(z)$ and $0 < \epsilon \le \epsilon_0$. Numerical endpoint and descriptor errors are smaller than the event margin and therefore do not change this classification.*

Assumption A.2 (Smooth channels and bounded numerical error)**.** *For every $z' \in \mathcal{N}_r(z)$ and every unit coefficient vector u used by a compared basis, $t \mapsto \psi_{z'}(tu)$ is three times continuously differentiable and*

$$\sup_{|t| \le \epsilon_0} \left\| \frac{\mathrm{d}^3}{\mathrm{d}t^3} \psi_{z'}(tu) \right\|_2 \le L_3. \tag{47}$$

The numerical channel and output-generator errors satisfy

$$\|\widehat{\psi}_{z'}(v) - \psi_{z'}(v)\|_2 \le \eta_\psi, \qquad \|\widehat{B}_{z'} - B_{z'}\|_2 \le \eta_B \tag{48}$$

throughout the queried ball. The residual weight $W \succeq 0$ is fixed on the neighborhood and $\|E_\star(z')\|_2 \le B_E$.

Assumption A.3 (Local regularity and independent sampling)**.** *Let $G_\star(z') = E_\star(z')^\top W E_\star(z')$. For the dimensionless state–action distance d_{sa} induced by Eq.* (32)*,*

$$\|G_\star(z') - G_\star(z)\|_2 \le L_G \, \mathrm{d}_{sa}(z', z)^\beta, \qquad 0 < \beta \le 1. \tag{49}$$

The local rows $Z_1, \dots, Z_m$ are independent draws from a distribution $\mu_{z,r}$ supported on $\mathcal{N}_r(z)$. With $H^{\mathrm{loc}}_{\star,r} = \mathbb{E}_{\mu_{z,r}}[G_\star(Z)]$ and $X = G_\star(Z) - H^{\mathrm{loc}}_{\star,r}$, assume

$$\|X\|_2 \le R_G, \qquad \left\|\mathbb{E}[X^2]\right\|_2 \le \nu_G^2. \tag{50}$$

Proposition 2 (Conditional finite-query Gram error)**.** *Under Assumptions A.1–A.3, define*

$$\delta_E(\epsilon) = \sqrt{d} \left(\frac{L_3}{6} \epsilon^2 + \frac{\eta_\psi}{\epsilon} + \eta_B \right). \tag{51}$$

Then $\|\widehat{E}_\epsilon(z') - E_\star(z')\|_2 \le \delta_E(\epsilon)$ uniformly on the neighborhood. If $\eta_\psi, \eta_B \le \eta$ and $0 < \epsilon \le \epsilon_0$, this implies

$$\delta_E(\epsilon) \le C_{\mathrm{fd}} \epsilon^2 + C_{\mathrm{sim}} \frac{\eta}{\epsilon}, \qquad C_{\mathrm{fd}} = \frac{\sqrt{d}\, L_3}{6}, \quad C_{\mathrm{sim}} = \sqrt{d}(1 + \epsilon_0). \tag{52}$$

Let

$$\widehat{H}_{\epsilon,r,m}(z) = \frac{1}{m} \sum_{i=1}^{m} \widehat{E}_\epsilon(Z_i)^\top W \widehat{E}_\epsilon(Z_i), \qquad H^{\mathrm{pt}}_\star(z) = G_\star(z), \tag{53}$$

and whiten both matrices by $M_{\mathfrak{g}_0}^{-1/2}$. For any $\delta_{\mathrm{p}} \in (0, 1)$, with probability at least $1 - \delta_{\mathrm{p}}$,

$$\left\| \widehat{\bar{H}}_{\epsilon,r,m}(z) - \bar{H}^{\mathrm{pt}}_\star(z) \right\|_2 \le \delta_H := \|M_{\mathfrak{g}_0}^{-1}\|_2 \Bigg[(2B_E + \delta_E)\|W\|_2 \delta_E + L_G r^\beta + \sqrt{\frac{2\nu_G^2 \log(2d/\delta_{\mathrm{p}})}{m}} + \frac{2R_G \log(2d/\delta_{\mathrm{p}})}{3m} \Bigg]. \tag{54}$$

Consequently, if the population eigengap is $\Delta > \delta_H$, Lemma 1 gives

$$\|\sin \Theta(\widehat{\mathcal{D}}, \mathcal{D}_\star)\|_2 \le \frac{\delta_H}{\Delta - \delta_H}. \tag{55}$$

If additionally $\min_i |\sigma_i(\bar{H}^{\mathrm{pt}}_\star) - \tau| > \delta_H$, the hard threshold selects the same rank as the pointwise population Gram.

The first term in Eq. (54) is the finite-query contribution, the second is neighborhood bias, and the final two terms are the self-adjoint matrix Bernstein fluctuation (Tropp, 2012). For small δ_E, the leading dependence is $\epsilon^2 + \eta/\epsilon + r^\beta + m^{-1/2}$, with the displayed quadratic correction retained in the formal bound. The concentration bound assumes independent, uniformly weighted rows. Fixed or non-IID replay neighborhoods require a separately controlled empirical-deviation term.

Multiscale finite-query Gram error. Let $\epsilon_1, \dots, \epsilon_L \le \epsilon_0$ use the same accepted rows, event signature and channel weights, and let $\alpha_\ell \ge 0$ with $\sum_{\ell=1}^{L} \alpha_\ell = 1$. Define

$$\widehat{H}_{\boldsymbol{\alpha},r,m}(z) = \frac{1}{m} \sum_{i=1}^{m} \sum_{\ell=1}^{L} \alpha_\ell \widehat{E}_{\epsilon_\ell}(Z_i)^\top W \widehat{E}_{\epsilon_\ell}(Z_i). \tag{56}$$

Then Eq. (54) holds with its finite-query term replaced by

$$q_{\boldsymbol{\alpha}} = \sum_{\ell=1}^{L} \alpha_\ell \big(2B_E + \delta_E(\epsilon_\ell)\big) \|W\|_2 \delta_E(\epsilon_\ell), \tag{57}$$

while the neighborhood and sampling terms are unchanged, including for correlated scales. The three-scale obstacle and wall estimators use equal weights.

To see this, apply Eq. (70) at each scale and use convexity of the spectral norm to bound the weighted sum by $q_{\boldsymbol{\alpha}}$. The same rows are then compared with $H^{\mathrm{loc}}_{\star,r}$, so the neighborhood-bias and matrix-Bernstein terms are identical to those in Proposition 2.

Corollary 2 (Shared-acceptance frame comparison)**.** *Work in the $M_{\mathfrak{g}_0}$-whitened coefficient coordinates and let Q be an orthogonal passive basis change. Suppose both estimators share the accepted neighborhood, samples, channel weights, perturbation scales and convex scale weights. Assumptions A.1–A.3 hold uniformly over both bases. For any $\delta_{\mathrm{p}} \in (0,1)$, define $\delta_{H,0}$ and $\delta_{H,Q}$ by Proposition 2, using Eq.* (57) *for multiple scales. Both bounds use the same sampling parameters R_G, ν_G and failure probability δ_{p}. Then, with probability at least $1-\delta_{\mathrm{p}}$,*

$$\bar{H}_{\star,Q} = Q^\top \bar{H}_{\star,0} Q, \qquad \left\| Q \widehat{\bar{H}}_Q Q^\top - \widehat{\bar{H}}_0 \right\|_2 \le \delta_{H,Q} + \delta_{H,0}. \tag{58}$$

If both estimates recover the same k-dimensional population subspace and $\delta_{H,Q}, \delta_{H,0} < \Delta$, their whitened orthogonal projectors satisfy

$$\left\| Q \widehat{P}_Q Q^\top - \widehat{P}_0 \right\|_2 \le \frac{\delta_{H,Q}}{\Delta - \delta_{H,Q}} + \frac{\delta_{H,0}}{\Delta - \delta_{H,0}}. \tag{59}$$

Corollary 2 applies when the compared bases share accepted rows, event classification and the local neighborhood. When a basis change alters these quantities, the estimators target different conditional populations and Eqs. (58)–(59) do not apply. The bound excludes rejected-state selection, event-margin estimation and policy-dependent replay.

A.7.3 Directional residual reconstruction and quotient stability

At each radius, the orbital estimator pools accepted directions across coefficient frames. Let $U = [u_1, \dots, u_K]$ collect these directions and let R contain the corresponding flow residuals. The local linear model is

$$R = EU + N, \qquad \widehat{E} = RU^\dagger, \qquad \widehat{H} = \widehat{E}^\top \widehat{E}, \tag{60}$$

where N contains finite-query and numerical residual error. Full row rank of U implies $UU^\dagger = I$, hence $\widehat{E} - E = NU^\dagger$. Writing $D = NU^\dagger$ gives

$$\widehat{H} - H = E^\top D + D^\top E + D^\top D, \qquad H = E^\top E, \tag{61}$$

and therefore Proposition 1.

Finite-query residual error. For the complete commutation residual in Eq. (3), define $E_\star = D_u \rho(z, 0)$ in common metric-whitened coordinates with the orbital channel scaling. On a common smooth event stratum, assume $\sup_{|t| \le \epsilon_0} \|\mathrm{d}^3 \rho(z, tu)/\mathrm{d}t^3\|_2 \le L_{\rho,3}$ for every retained unit direction u, with $\max \mathcal{E} \le \epsilon_0$. Let η_ρ bound each numerical evaluation error of this complete, scaled residual before differencing. Taylor's formula bounds each column of $N_\epsilon = R_\epsilon - E_\star U_\epsilon$ by b_ϵ. For K_ϵ retained direction pairs,

$$\begin{aligned} b_\epsilon &:= \frac{L_{\rho,3}}{6} \epsilon^2 + \frac{\eta_\rho}{\epsilon}, \\ \nu_\epsilon &:= \|N_\epsilon U_\epsilon^\dagger\|_2 \le \|N_\epsilon\|_F \|U_\epsilon^\dagger\|_2 \le \sqrt{K_\epsilon}\, \|U_\epsilon^\dagger\|_2\, b_\epsilon. \end{aligned} \tag{62}$$

The derivative bound includes the finite transformation of the nominal output. These bounds supply the residual error in Proposition 1 for each radius fit.

Each radius Gram is positive semidefinite by construction, and convex aggregation satisfies

$$\left\| \sum_j \omega_j \widehat{H}_j - \sum_j \omega_j H_j \right\|_F \le \sum_j \omega_j \|\widehat{H}_j - H_j\|_F, \qquad \omega_j \ge 0, \quad \sum_j \omega_j = 1. \tag{63}$$

If $\mathrm{tr}(H) \ge t_0 > 0$ and $\sqrt{d}\,\|\widehat{H} - H\|_F \le t_0/2$, trace normalization is locally Lipschitz:

$$\left\| \frac{\widehat{H}}{\mathrm{tr}(\widehat{H})} - \frac{H}{\mathrm{tr}(H)} \right\|_F \le \left(\frac{2}{t_0} + \frac{2\sqrt{d}\,\|H\|_F}{t_0^2} \right) \|\widehat{H} - H\|_F. \tag{64}$$

Processed spectral target. At a fixed geometry, suppose the radii share the same first-order differential $E_\star$ and $\mathrm{tr}(E_\star^\top E_\star) > 0$. The normalized empirical average and its population target are

$$\widehat{\bar{H}}_{\mathrm{eff}} = \frac{1}{|\mathcal{E}|} \sum_{\epsilon \in \mathcal{E}} \mathcal{N}(\widehat{\bar{H}}_\epsilon), \qquad \bar{H}_{\mathrm{eff},\star} = \mathcal{N}(E_\star^\top E_\star). \tag{65}$$

Under the positivity and small-error conditions above, Eqs. (64) and (63) control the perturbation of this pair. The conversion $\|\Delta H\|_F \le \sqrt{d}\,\|\Delta H\|_2$ connects Proposition 1 to these Frobenius-norm bounds. For orbital recovery, Lemma 1 and the relative-threshold margin in Eq. (44) apply to this processed empirical–population pair. Their perturbation is $\delta = \|\widehat{\bar{H}}_{\mathrm{eff}} - \bar{H}_{\mathrm{eff},\star}\|_2$, and the eigengap is that of $\bar{H}_{\mathrm{eff},\star}$.

For the quotient chart, define $n(P) = (-P_{12}, P_{02})^\top$ so that $c(P) = Ln(P)/P_{22}$. If $P_{22}, \widehat{P}_{22} \ge \alpha$, then

$$\begin{aligned} \|c(\widehat{P}) - c(P)\|_2 &\le L \left(\frac{\|n(\widehat{P}) - n(P)\|_2}{\alpha} + \frac{\|n(P)\|_2 |\widehat{P}_{22} - P_{22}|}{\alpha^2} \right) \\ &\le \sqrt{2} L (\alpha^{-1} + \alpha^{-2}) \|\widehat{P} - P\|_F, \end{aligned} \tag{66}$$

where entries of an orthogonal projector have magnitude at most one. This proves Corollary 1 and connects subspace error to the chart used by the actor and critics.

A.8 Practical scope and limitations

The method identifies linear subspaces within a supplied candidate Lie algebra using counterfactual simulator queries on differentiable event strata. Its bilateral probes characterize subspaces rather than one-sided feasible cones. Orbital quotient use is governed by the finite reward, event, termination and safety checks.

B Proofs and Additional Theoretical Details

B.1 Proof of Lemma 1

Proof. The generalized eigenproblem $Hv = \sigma M_{\mathfrak{g}_0} v$ is converted to an ordinary symmetric eigenproblem by whitening with $M_{\mathfrak{g}_0}^{1/2}$. Let $\bar{H}_\star = M_{\mathfrak{g}_0}^{-1/2} H_\star M_{\mathfrak{g}_0}^{-1/2}$ and $\widehat{\bar{H}} = M_{\mathfrak{g}_0}^{-1/2} \widehat{H} M_{\mathfrak{g}_0}^{-1/2}$. The population and empirical subspaces in the whitened coordinates are spanned by the first k eigenvectors of these two matrices. By assumption, the population gap between the kth and $(k+1)$th eigenvalues is Δ and the perturbation has spectral norm at most δ.

The Davis–Kahan sin-theta theorem gives

$$\|\sin\Theta(\widehat{\bar{\mathcal{D}}}, \bar{\mathcal{D}}_\star)\|_2 \le \frac{\|\widehat{\bar{H}} - \bar{H}_\star\|_2}{\Delta - \|\widehat{\bar{H}} - \bar{H}_\star\|_2} \le \frac{\delta}{\Delta - \delta}. \tag{67}$$

Mapping the whitened eigenvectors back through $M_{\mathfrak{g}_0}^{-1/2}$ gives the same subspace angle in the $M_{\mathfrak{g}_0}$ inner product. This proves Eq. (12). □

B.2 Proof of Proposition 2

Proof. Fix $z' \in \mathcal{N}_r(z)$ and a queried unit direction e_j. Taylor's theorem bounds the centered-difference bias by $L_3\epsilon^2/6$. The numerical channel and output-generator errors give

$$\|[\widehat{E}_\epsilon(z') - E_\star(z')]e_j\|_2 \le \frac{L_3}{6}\epsilon^2 + \frac{\eta_\psi}{\epsilon} + \eta_B. \tag{68}$$

Bounding the spectral norm by the Frobenius norm yields Eq. (51). Equation (52) follows from $\eta_B \le \epsilon_0\eta/\epsilon$ when $\eta_\psi, \eta_B \le \eta$ and $\epsilon \le \epsilon_0$.

Set $E_i = E_\star(Z_i)$, $D_i = \widehat{E}_\epsilon(Z_i) - E_i$ and $\widehat{G}_i = (E_i + D_i)^\top W(E_i + D_i)$. Expanding the pointwise Gram gives

$$\widehat{G}_i - G_\star(Z_i) = E_i^\top W D_i + D_i^\top W E_i + D_i^\top W D_i. \tag{69}$$

Since $\|E_i\|_2 \le B_E$ and $\|D_i\|_2 \le \delta_E$, submultiplicativity implies

$$\|\widehat{G}_i - G_\star(Z_i)\|_2 \le (2B_E + \delta_E)\|W\|_2\delta_E. \tag{70}$$

Write $H^{\mathrm{loc}}_{\star,r} = \mathbb{E}[G_\star(Z)]$ and $X_i = G_\star(Z_i) - H^{\mathrm{loc}}_{\star,r}$. The Hölder condition and support radius give $\|H^{\mathrm{loc}}_{\star,r} - G_\star(z)\|_2 \le L_G r^\beta$. Adding and subtracting the empirical population Grams and their mean yields

$$\|\widehat{H}_{\epsilon,r,m}(z) - G_\star(z)\|_2 \le (2B_E + \delta_E)\|W\|_2\delta_E + \left\|\frac{1}{m}\sum_i X_i\right\|_2 + L_G r^\beta. \tag{71}$$

The self-adjoint matrix Bernstein inequality (Tropp, 2012) and Eq. (50) give, with probability at least $1 - \delta_{\mathrm{p}}$,

$$\left\|\frac{1}{m}\sum_i X_i\right\|_2 \le \sqrt{\frac{2\nu_G^2\log(2d/\delta_{\mathrm{p}})}{m}} + \frac{2R_G\log(2d/\delta_{\mathrm{p}})}{3m}. \tag{72}$$

Finally, whitening multiplies the bound by at most $\|M_{\mathfrak{g}_0}^{-1}\|_2$, since

$$\|M_{\mathfrak{g}_0}^{-1/2}(\widehat{H} - H^{\mathrm{pt}}_\star)M_{\mathfrak{g}_0}^{-1/2}\|_2 \le \|M_{\mathfrak{g}_0}^{-1}\|_2\|\widehat{H} - H^{\mathrm{pt}}_\star\|_2. \tag{73}$$

This proves Eq. (54). Lemma 1 gives the subspace bound, and Weyl's inequality preserves rank when every population eigenvalue is more than δ_H from the cutoff. □

B.3 Proof of Corollary 2

Proof. After coefficient whitening, a passive orthogonal basis change Q replaces the channel map by $\psi_{z,Q}(u) = \psi_{z,0}(Qu)$ and the expected output action by $B_{z,Q} = B_{z,0}Q$. The chain rule therefore gives $E_{\star,Q} = E_{\star,0}Q$ and $\bar{H}_{\star,Q} = Q^\top\bar{H}_{\star,0}Q$. The shared samples also give centered population Grams related by $X_{i,Q} = Q^\top X_{i,0}Q$. Their empirical averages have identical spectral norms, so one Bernstein event controls both sampling errors with probability at least $1 - \delta_{\mathrm{p}}$. Combining this event with the uniform finite-query bounds and mapping the Q-basis estimate back gives

$$\begin{aligned}\|Q\widehat{\bar{H}}_Q Q^\top - \widehat{\bar{H}}_0\|_2 &\le \|Q\widehat{\bar{H}}_Q Q^\top - \bar{H}_{\star,0}\|_2 + \|\widehat{\bar{H}}_0 - \bar{H}_{\star,0}\|_2 \\ &\le \delta_{H,Q} + \delta_{H,0}.\end{aligned} \tag{74}$$

When the threshold selects the same rank, the spectral norm between two equal-rank orthogonal projectors equals the largest sine of their principal angles. Applying Lemma 1 to each estimate and using the triangle inequality through the population projector gives Eq. (59). □

C Additional Experimental Details

C.1 Configuration and reproducibility

Table 2 collects the principal settings. Policy evaluation uses deterministic mean actions, and learning-curve summaries report mean ± s.e.m. across matched seeds. Appendix C.2 defines the orbital data splits, metrics and paired bootstrap procedures.

Table 2: **Configuration of the reported experiments.**

Component	Configuration
Orbital-contact task	Direct-wrench planar capsule in MuJoCo; 220-step horizon; force–torque action; eight-dimensional policy observation; 24 centers, four start phases and two travel directions.
Orbital recovery	Bilateral metric radii $\{0.0004, 0.0008\}$ m; five coefficient frames; 37 projective directions per frame; relative spectral threshold 0.015; one canonical condition per geometry for sparse recovery.
Orbital SAC	50,000 environment steps; seeds 9400–9419; width-128 twin-critic SAC; batch size 256; updates begin after 1,000 steps; Adam learning rate 10^{-4}; one update per environment step.
Obstacle recovery	3,000 replayable snapshots; two train and two held-out contact-normal sectors; perturbation scales $\{5\times10^{-4}, 10^{-3}, 2\times10^{-3}\}$; event-compatible $k = 32$ aggregation; spectral threshold 30; flow and event weights one.
Wall recovery and gates	27,000 queried rows, including 10,403 accepted training-angle rows; scalar and diagonal width-128 MLPs trained for 100 epochs; batch size 512; Adam learning rate 10^{-3}; targets $\mathrm{tr}(P)/3$ and $\mathrm{diag}\, P$.
Wall SAC	150,000 environment steps; ten matched seeds; width-256 actor and twin critics; batch size 256; Adam learning rate 3×10^{-4}; regularization coefficient 0.01; evaluation every 5,000 steps.

C.2 Principal orbital-contact recovery and online control

Task and score. The planar capsule has half-segment 0.055 m, radius 0.025 m and mass 0.25 kg. It moves around a circular obstacle of radius 0.18 m under wrench (f_x, f_y, τ). The commanded force satisfies $\|f\|_2 \le 3$ N, and torque satisfies $|\tau| \le 0.16$ N m. Every simulator substep also applies a radial preload $f_{\mathrm{pre}} = -(F_0 + b_r v_r)e_r$, with $F_0 = 6$ N and $b_r = 1\,\mathrm{N\,s\,m^{-1}}$. Here e_r points outward from the obstacle center and $v_r = v^\top e_r$ is the radial velocity. Appendix A.6 defines the radial force squashing and canonical-to-world action transport. The orbit radius is 0.2043 m, and the simulator step is 0.005 s with frame skip three. Episodes begin from settled force-bearing contact, with a target 60° clockwise or counterclockwise along the orbit. With signed angular progress $\Delta\phi_t$, goal error e_t and radial deviation d_t, the reward is

$$\begin{aligned} r_t = {} & 7\Delta\phi_t - 0.08(e_t/\pi)^2 - 0.03(d_t/0.01)^2 \\ & - 0.008\left(\|f_t\|_2^2/3^2 + \tau_t^2/0.16^2\right) - 0.12\,\mathbf{1}[\text{no contact}] + 2\,\mathbf{1}[\text{success}]. \end{aligned} \tag{75}$$

Success requires force-bearing contact, goal error at most 5° and cumulative signed progress of at least 0.65 rad. For contact occupancy o_c, cumulative progress q, target travel Δ and final error e, the regulation score is

$$o_c\,\frac{1}{2}\left[\mathrm{clip}(q/|\Delta|, 0, 1) + \mathrm{clip}(1 - |e|/|\Delta|, 0, 1)\right]. \tag{76}$$

The primary endpoint integrates this score by the trapezoidal rule over evaluations every 2,000 steps from 0 to 50,000, normalized by the 50,000-step horizon.

Policy and evaluation splits. The 24 obstacle-center angles run from -174.375° to 170.625° in 15° increments. Training uses the 12 even-indexed centers, phases $\{-165^\circ, 15^\circ\}$ and target displacements $\{-60^\circ, +60^\circ\}$, giving 48 conditions. Seen-center evaluation uses phases $\{-75^\circ, 105^\circ\}$ at these centers with both directions, giving another 48 conditions. OOD evaluation uses the 12 odd-indexed

centers, all four phases and both directions, giving 96 conditions. These centers are withheld from policy optimization. The log label `iid` denotes seen-center, unseen-phase evaluation.

Geometry-level recovery. Every center is queried at the canonical $(15^\circ, +60^\circ)$ condition before policy training. Sparse recovery reuses this projector across the other seven phase–direction pairs, whereas dense recovery certifies all eight conditions independently. Policy training uses the resulting fixed tables and issues no counterfactual queries.

Queries start from the settled snapshot with zero commanded wrench and the environment's radial preload. They transform body, goal and wrench while holding the obstacle fixed. Signed pairs require finite residuals and rewards and matching nominal event and termination signatures. Their six scaled flow channels are defined in Eq. (25). At each radius, retained directions from all frames are pooled in common whitened coordinates. Residual-map fitting uses identity channel weighting, followed by trace normalization of each radius Gram and averaging across radii, as specified in Algorithm 1.

The complete 37-direction design has $\kappa(UU^\top) = 1.064$ and $\|U^\dagger\|_2 = 0.289$, invariant under orthogonal frame rotation. After screening, each retained pooled design must have numerical rank three at singular-value tolerance 10^{-10}. Its pseudoinverse controls reconstruction sensitivity in Proposition 1. Each radius Gram and their average must select rank one at threshold $\max\{10^{-12}, 0.015\lambda_{\max}\}$, with at most 10° between radius projectors. Any termination change or penetration above 0.01 m rejects the geometry. Admission also requires $\epsilon_{\max}\|\widehat{P}\widehat{g}_r\|_2 \le 10^{-4}$ for the fitted first-order reward variation. Per-frame projectors are diagnostic outputs.

Both tables accepted all contact-interior queries with 100% conditional rank accuracy and rejected every outward guard. Sparse recovery obtained mean and 90th-percentile principal angles of 0.384° and 0.889°, respectively. The corresponding angles for dense recovery were 0.524° and 1.076°. Their respective costs were 17,784 calls for 24 keys and 142,272 calls for 192 keys.

Online controls and statistical analysis. All branches share seeds, initialization, networks, optimizer and quotient code. Sparse recovered full, diagonal and wrong use the same table and differ only in the supplied operator. Sparse oracle full supplies analytic projectors at the same 24 keys; dense recovered full uses the 192-key table. No-op applies no quotient. Table 1 reports the common 50k-step endpoint, and the call-count-matched comparison below extends no-op training to 67,784 interactions.

For each primary orbital contrast, we form a paired seed-by-center difference matrix. A paired two-way bootstrap independently resamples the seed and policy-held-out-center axes with replacement, then averages the resulting Cartesian product. Intervals are the 2.5th and 97.5th percentiles of 100,000 replicates. For $\Delta_{\mathrm{sparse-dense}} = \mu_{\mathrm{sparse}} - \mu_{\mathrm{dense}}$, the prespecified non-inferiority null is $H_0 : \Delta_{\mathrm{sparse-dense}} \le -0.02$. The margin represents two percentage points on the unit-range regulation score. Non-inferiority requires the lower endpoint of the 95% interval to exceed -0.02.

For full minus no-op, diagonal and wrong, each contrast was positive after averaging over centers for all 20 seeds and after averaging over seeds for all 12 policy-held-out centers.

Alternative performance metrics. Return and success AUC use the same evaluation grid and horizon normalization as regulation AUC (Table 3). Sparse recovered full exceeded all three controls on both metrics. Against no-op, the 95% paired two-way bootstrap intervals were $[2.696, 3.000]$ for return-AUC difference and $[0.093, 0.134]$ for success-AUC difference. Against diagonal, the corresponding intervals were $[0.567, 1.240]$ and $[0.076, 0.133]$. Against wrong, they were $[0.759, 1.323]$ and $[0.094, 0.145]$.

Table 3: **Orbital metric robustness.** Policy-held-out AUC over 0–50k steps, reported as mean ± s.e.m. over 20 paired seeds. Higher values are better.

Branch	Regulation AUC	Return AUC	Success AUC
No-op	0.137 ± 0.004	-5.219 ± 0.068	0.126 ± 0.005
Sparse recovered diagonal	0.206 ± 0.003	-3.285 ± 0.157	0.134 ± 0.011
Sparse recovered wrong	0.204 ± 0.003	-3.411 ± 0.131	0.120 ± 0.009
Sparse recovered full	0.268 ± 0.005	-2.370 ± 0.083	0.240 ± 0.010

Phase-resolved evaluation. This post hoc decomposition groups episodes by whether their initial phase appeared in policy training, retaining pooled OOD AUC as the primary endpoint. Within each branch, seed and center, we average the two evaluation episodes and both directions, then integrate each phase group's 0–50k curve. Equal weighting of the two groups reproduces the original OOD AUC for all six branches. Bootstrap resampling retains phases, directions and episodes within each seed–center cell.

Table 4: **Phase-resolved orbital evaluation.** AUC on the 12 policy-held-out centers, averaging both travel directions. Training phases are -165° and 15°; unseen phases are -75° and 105°. Values are mean ± s.e.m. over 20 paired seeds.

Branch	Training-phase AUC	Unseen-phase AUC
No-op	0.230 ± 0.007	0.045 ± 0.003
Sparse recovered diagonal	0.203 ± 0.004	0.209 ± 0.004
Sparse recovered wrong	0.201 ± 0.003	0.208 ± 0.003
Sparse oracle full	0.236 ± 0.007	0.305 ± 0.006
Dense condition-specific recovered full	0.241 ± 0.007	0.308 ± 0.006
Sparse recovered full	0.236 ± 0.004	0.301 ± 0.007

Full minus no-op was $+0.006$ (95% CI $[-0.012, 0.022]$) at training phases and $+0.256$ ($[0.242, 0.271]$) at unseen phases. Full-minus-diagonal and full-minus-wrong intervals were positive in both groups. No-op achieved 0.045 unseen-phase AUC on both seen and held-out centers, compared with 0.230 at training phases on held-out centers. This phase composition explains its pooled OOD AUC of 0.137 versus the seen-center, unseen-phase score of 0.045.

Representation control. The separate representation-control cohort uses 48 contexts and seeds 7300–7319. Branches share the eight-dimensional observation o_{base}, recovered table and accepted support. Three added coordinates give each branch 11 state-input dimensions and 55,944 trainable parameters, with matched initialization and 50k-step SAC training. Critics additionally receive the action. Let p be body position, $\widehat{c}$ the recovered pivot, c the oracle center, L the reference length and b_{cert} the recovered certificate bit. The inputs are

$$o_{\mathrm{full}} = [\mathcal{C}_{\widehat{P}}(o_{\mathrm{base}}, p), 0, 0, b_{\mathrm{cert}}], \qquad o_{\mathrm{rec}} = [o_{\mathrm{base}}, (p - \widehat{c})/L, b_{\mathrm{cert}}],$$
$$o_{\mathrm{oracle}} = [o_{\mathrm{base}}, (p - c)/L, b_{\mathrm{cert}}], \qquad o_{\mathrm{pos}} = [o_{\mathrm{base}}, p/L, b_{\mathrm{cert}}],$$
$$o_{\mathrm{noop}} = [o_{\mathrm{base}}, 0, 0, b_{\mathrm{cert}}].$$

The full branch canonicalizes observations and critic actions and transports its actor's canonical wrench to the world frame. Feature branches retain world-frame observations and actions. Recovered- and oracle-pivot coordinates are zero when $b_{\mathrm{cert}} = 0$, giving both branches the same certified support.

Table 5 reports the representation-control AUCs. Contrasts use 100,000 paired seed-bootstrap resamples. Full minus recovered pivot was $+0.122$ (95% CI $[0.111, 0.134]$; 20/20 positive seeds). Oracle minus recovered pivot was -0.001 ($[-0.009, 0.007]$; 9/20 positive seeds), while both feature branches remained below the full quotient.

Table 5: **Orbital representation control.** OOD regulation AUC, reported as mean ± s.e.m. over 20 paired seeds in the separate matched cohort.

Representation	OOD AUC
Recovered full quotient	0.264 ± 0.005
Recovered pivot feature	0.141 ± 0.003
Oracle pivot feature	0.140 ± 0.003
Position only	0.121 ± 0.002
Dimension-matched no-op	0.144 ± 0.003

Single-policy call-count-matched control. Each sparse recovered-full policy is charged all 17,784 recovery calls before its 50,000 policy steps. No-op receives 67,784 ordinary interactions, including the same first 50,000 steps used in the primary comparison. Additional interactions use training centers and phases, enter replay and receive one optimizer update each. Evaluation transitions remain outside replay, and evaluation calls are excluded from both budgets.

For total-call AUC, the recovered curve is shifted by 17,784 calls, with its step-zero score held throughout certification. Both curves are trapezoidally integrated over 0–67,784 calls and divided by 67,784. Table 6 distinguishes this shared-budget comparison from the original 0–50k policy-step AUC contrast.

Table 6: **Single-policy call-count-matched control.** Recovered full minus no-op over 20 paired seeds, with 95% paired two-way bootstrap intervals. Final regulation and total-call AUC use 67,784 calls; policy-step AUC compares the original 50k-step runs.

Metric	Difference	95% CI
Final OOD regulation	$+0.279$	$[0.230, 0.323]$
Policy-step OOD AUC (0–50k)	$+0.131$	$[0.119, 0.142]$
Normalized total-call OOD AUC	$+0.037$	$[0.027, 0.046]$

At 67,784 calls, final OOD regulation was 0.472 ± 0.017 for sparse recovered full and 0.194 ± 0.011 for no-op. The total-call AUC contrast was positive for 18/20 seeds and all 12 policy-held-out centers.

C.3 Orbital spectral stability

We audited a separate 48-context orbital recovery cohort. The 37-direction baseline reproduced the original spectra and projectors. A 25-direction design retained 100% rank accuracy, with a 90th-percentile angle of 1.09° versus 1.03° for the baseline. All 48 outward-guard queries were rejected.

Noise tests added independent Gaussian perturbations with standard deviation qs_ρ/ϵ to each retained derivative-residual entry before refitting. Here q is the noise ratio and s_ρ is the context-level RMS of $\epsilon\widehat{e}_{\epsilon,u}$ across retained entries, frames and radii. Each nonzero level used 100 replicates per context; event-validity masks and reward columns remained fixed.

Figure 5 shows both complete scans. Rank accuracy was 100% at the tested thresholds from 0.0025 to 0.02 and at all tested noise levels through 10%. The 90th-percentile angle was 1.02° at 1% noise and 1.73° at 10%.

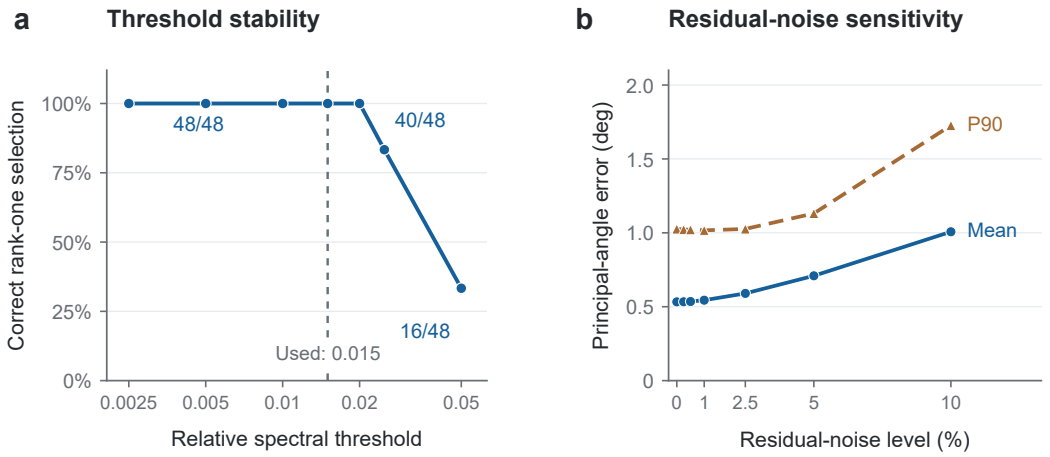

Figure 5: **Orbital spectral-threshold and residual-noise sensitivity. a,** Correct rank-one selection across 48 contexts; the dashed line marks the adopted threshold, 0.015. **b,** Mean and P90 principal-angle error against the analytic orbital projector, pooling 100 replicates per context at each nonzero noise level. P90 denotes a descriptive quantile. Lines connect tested settings.

C.4 Recovery audits and geometry baselines

Finite-query frame dependence. We tested the canonical frame and 16 Haar-distributed orthonormal frames on 221 obstacle and 1,816 wall contact snapshots. Projectors were mapped back to common metric-whitened coordinates. Table 7 separates acceptance variation from rank and direction agreement on common support.

Table 7: **Finite-query frame audit.** Each random frame is compared with the canonical frame. Angle and projector-error quantiles use pairwise common-accepted, same-rank, nontrivial rows. Rank agreement uses all common-accepted rows.

Metric	Circular obstacle	Wall
Canonical/mean coverage	15.38/17.54%	13.49/13.57%
Median Jaccard	0.549	0.782
Rank agreement	99.05%	71.94%
Angle (P90)	0.706°	44.30°
$\|\Delta P\|_{\mathrm{F}}$ (P90)	0.0174	0.9876

There were 419 pairwise common-accepted comparisons for obstacle contact and 3,471 for wall contact. Median angles on the same-rank, nontrivial subset were $0.0105°$ and $0.0212°$, respectively. Obstacle directions agreed closely, whereas wall contact had a pronounced upper error tail. The frame audit uses exponential coordinates and stricter multiscale event checks than the principal direct-coordinate protocol. Its 15.38% obstacle canonical coverage and the principal 49.77% coverage therefore measure different acceptance rules on the same snapshots.

Coverage and privileged geometry. Principal obstacle recovery accepted 2,670 of 3,000 replayable snapshots, including 110 of 221 held-out contact queries (49.77% coverage). On accepted held-out contact queries, rank accuracy was 100%, valid-direction retention was 0.999997 and normal leakage was 1.25×10^{-6}. Of the 111 rejected queries, 61 failed transform validity, 42 were grazing or scale unstable, and eight changed event topology. Table 8 compares the recovered subspace with privileged geometry constructions. Projector error is the mean Frobenius distance to the analytic stabilizer. The angle metric uses the largest principal angle for matching ranks and assigns $90°$ to a rank mismatch.

Table 8: **Held-out circular-obstacle contact.** Recovered errors are conditional on acceptance; geometry baselines use all 221 rows. Angle is the rank-penalized metric defined above.

Method	Coverage	Rank acc.	Projector error	Angle
Analytic stabilizer	100%	100%	0	0°
Contact normal only	100%	100%	1.414214	90°
Contact-Jacobian nullspace	100%	0%	1.000	90°
Counterfactual recovered	49.77%	100%	0.00206	0.083°

The normal-only operator retains a tangent translation rather than the orbital $\mathrm{SE}(2)$ generator. The contact-Jacobian nullspace retains an additional instantaneous rotation. On held-out wall contact, recovery achieved 11.29% coverage, 93.66% conditional rank accuracy and a 5.72° mean rank-penalized angle.

Channel ablations distinguish the two contact regimes. On walls, event-only, flow-only and combined rank accuracies were 94.15%, 22.93% and 93.66%, respectively. On the obstacle, they were 7.27%, 98.18% and 100%. Thus event information carried most wall-recovery accuracy, while flow residuals carried most obstacle-recovery accuracy.

C.5 Wall amortization stress test

Task and observations. The point pusher uses planar velocity actions $a \in [-0.8, 0.8]^2$. Its 15-dimensional observation contains pusher pose and velocity, block pose and velocity, goal position, wall normal and wall offset. Learned gates receive this observation concatenated with the action. Policy training uses wall angles 0°, 15° and 30°; OOD evaluation uses 45°, 60° and 90°. The reward is

$$r_t = 20(d_t - d_{t+1}) - d_{t+1} - 0.02\frac{\|a_t\|_2^2}{0.8^2} - 0.01[\max(F_t^{\text{wall}} - 10, 0)]^2 + 10\,\mathbf{1}[d_{t+1} \le 0.08], \tag{77}$$

where d_t is block–goal distance, and success occurs at $d_{t+1} \le 0.08$ m. Free, near and contact resets use signed block–wall clearance ranges $[0.20, 0.40]$, $[0.01, 0.08]$ and $[-0.002, 0.001]$ m. For wall normal n and tangent $t = (-n_y, n_x, 0)^\top$, the geometry oracle uses $P = I_3$ beyond 0.02 m clearance and $P = tt^\top$ otherwise. Oracle scalar uses $\mathrm{tr}(P)I_3/3$. Oracle-label diagonal fits $\mathrm{diag}\, P$; recovered-label gates fit the corresponding summaries of recovered labels.

Critic regularization. For a differentiable critic $Q_\theta(s, m, a)$, the candidate-basis Lie derivatives are

$$g_Q(z) = \begin{bmatrix} \nabla_s Q_\theta(z)^\top \xi_1^S(s) + \nabla_a Q_\theta(z)^\top \xi_1^A(a) \\ \vdots \\ \nabla_s Q_\theta(z)^\top \xi_d^S(s) + \nabla_a Q_\theta(z)^\top \xi_d^A(a) \end{bmatrix}. \tag{78}$$

Let $S_Q(z)$ be the branch's positive-semidefinite, $M_{\mathfrak{g}_0}$-self-adjoint regularization operator, including its confidence weight. It uses the hard, soft, scalar or diagonal construction in Appendix A.5. The directional penalty is

$$\mathcal{L}_Q^{\text{sym}} = \mathbb{E}_{z\sim\mathcal{D}_{\text{RL}}}\left[g_Q(z)^\top \,\mathrm{sg}[S_Q(z)] M_{\mathfrak{g}_0}^{-1} g_Q(z)\right]. \tag{79}$$

For $\mathbb{E}[u \mid z] = 0$ and $\mathrm{Cov}(u \mid z) = S_Q(z)M_{\mathfrak{g}_0}^{-1}$, the integrand equals $\mathbb{E}[(u^\top g_Q(z))^2 \mid z]$. The regularized critic objective is

$$\mathcal{L}_Q = \mathcal{L}_Q^{\text{SAC}} + \lambda_Q \mathcal{L}_Q^{\text{sym}}, \qquad \lambda_Q = 0.01. \tag{80}$$

The actor uses the standard SAC update. Wall OOD AUC averages success across the three OOD angles and all evaluation checkpoints; Final OOD uses the 150k checkpoint.

Lookup, refresh and augmentation controls. Recovered lookup selects the nearest accepted row within the nearest available wall-angle bucket, using observation, scaled action and clearance features. Periodic refresh relabels 500 replay samples from this offline lookup every 10,000 steps and replaces the cache, requiring no additional simulator queries. Recovered augmentation draws a Gaussian coefficient vector, projects and normalizes it, and applies a 0.02 first-order state–action displacement. It uses one augmented pair per replay sample, clips actions to their feasible interval and fits both critics to the original detached Bellman target. This auxiliary fit replaces the derivative penalty and uses the same coefficient 0.01.

Amortized reconstruction and online performance. Table 9 compares the same descriptor head under oracle and recovered supervision. Oracle labels yielded accurate held-out reconstruction, whereas recovered labels left substantial operator error. Learned online gates therefore used scalar and diagonal outputs.

Table 9: **Descriptor-head full-operator amortization.** Metrics use anisotropic rows. NMSE divides summed squared Frobenius prediction error by summed squared target Frobenius norm. Angles are sample medians within each fit. Reported values average three gate seeds.

Label source	ID NMSE	OOD NMSE	OOD angle
Oracle	0.0173	0.0815	$< 0.0073^\circ$
Recovered	0.2257	0.6333	32.63°

Table 10: **Wall-pushing online control.** OOD success AUC and final success over 45°, 60° and 90° walls, reported as mean $\pm$ s.e.m. over ten matched 150k-step seeds.

Method	OOD AUC	Final OOD
Vanilla SAC	0.510 ± 0.025	0.547 ± 0.051
Full $S = I_3$	0.557 ± 0.029	0.638 ± 0.027
Oracle scalar	0.551 ± 0.018	0.582 ± 0.030
Geometry oracle	0.557 ± 0.030	0.524 ± 0.041
Recovered-label scalar	0.554 ± 0.029	0.607 ± 0.037
Recovered-label diagonal	0.620 ± 0.025	0.631 ± 0.051
Oracle-label diagonal	0.583 ± 0.028	0.596 ± 0.040
Recovered lookup	0.562 ± 0.038	0.544 ± 0.040
Periodic cache refresh	0.568 ± 0.026	0.538 ± 0.028
Recovered augmentation	0.528 ± 0.021	0.556 ± 0.036

In Table 10, recovered-label diagonal exceeded recovered-label scalar by 0.066 OOD AUC (95% CI $[0.003, 0.137]$); their final OOD difference was 0.024 ($[-0.076, 0.142]$). At 11.29% held-out contact coverage, dense gate prediction extends beyond accepted support.

C.6 Simulator-call accounting

Table 11 separates sparse deployment, dense condition-specific recovery and independent recovery audits. Basis-column estimators use one centered pair per direction and scale, giving a nominal $2d$ probe count. The orbital estimator instead uses projective-direction designs; its cost is counted directly. Logged totals include nominal steps when executed. Offline noise injection and cached wall refresh add no simulator calls.

Table 11: **Counterfactual query cost across recovery settings.** Sparse recovery is the principal deployment cost; dense recovery and the spectral audit are separate comparisons.

Setting	Query design	Logged calls
Sparse orbital geometry table	$d = 3$; 24 geometry keys; 5 frames; 37 directions; 2 radii	17,784
Dense orbital condition-specific comparator	$d = 3$; 192 task-condition keys; 5 frames; 37 directions; 2 radii	142,272
Orbital spectral/noise audit	$d = 3$; 48 contexts; three direction designs; clean probes and offline noise	72,144
Billiard $\mathbb{R}^2$ pointwise	$d = 2$; 3,721 states; basis-direction pairs	n/a
Wall $\mathfrak{se}(2)$ offline	$d = 3$; 27,000 rows; 3 scales; basis-direction pairs	511,312
Wall $\mathfrak{se}(2)$ refresh cache	$d = 3$; 7,500 cache entries; lookup maintenance	0
Obstacle $\mathfrak{se}(2)$ contact sectors	$d = 3$; 3,000 snapshots; 3 scales; event-compatible basis pairs	56,520

Measured wall-clock times were 664.6 s for wall offline recovery, 35.5 s for obstacle recovery and 0.15 s for cached wall refresh maintenance.