ON STATISTICAL COMPLEXITY OF LEARNING KINEMATIC TOPOLOGIES AND MANIPULATION MAPS

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning kinematic maps requires separating representational constraints from statistical estimation costs. We develop a framework connecting topology, inverse conditioning, and unknown model structure to representation and learning guarantees. For regular kinematic fibrations, minimum numbers of continuous inverse-selection and manipulation domains persist below a geometric error threshold. Sensitivity constraints yield risk and validation bounds, including an exact Lipschitz–risk frontier for circle inversion. A shared-parameter construction yields growing topological complexity with dimension-independent sample requirements. Chart-dependent budgets and calibration risk quantify the effects of conditioning and parameter sharing. We compute invariants for robot maps and general torus homomorphisms, and formulate protein/RNA kinematics on constrained torsion spaces. A molecular decoder bound transfers angular error to Cartesian error, with observed-coordinate learning distinguished from latent inference. Planar and spatial seven-joint studies, including sixty retained spatial regressors and independent Jacobian checks, isolate representation effects and redundant-solution ambiguity.

1 INTRODUCTION

Low average test error can hide severe kinematic prediction failures caused by coordinate cuts, the absence of a continuous global inverse, or sensitivity near singularities. These causes require different remedies; more data or larger networks alone may not resolve them. Position and velocity learning concern a kinematic map and its differential, but complexity also depends on the task: forward prediction estimates a task pose, inverse prediction selects a configuration, manipulation planning connects a starting configuration to a target task value, and calibration estimates physical parameters. Joint count and inverse multiplicity therefore give an incomplete picture. The regular position map of a planar 2R arm, for example, has two inverse solutions, one continuous inverse selection, and two manipulation-planning domains, yet its inverse sensitivity diverges near full extension. Molecular kinematics raises similar questions: dihedral rotations move downstream atoms, and independent periodic torsions form an ambient hypertorus. Loop closure, steric exclusion, and observation constraints alter the feasible domain or map. Hence, coordinate reconstruction alone does not solve hyper-redundant RNA loop proteomics.

Robot learning has progressed from direct forward and inverse approximations (Geng & Haynes, 1991; Köker et al., 2004) and probabilistic mixture models (Bishop, 1994) to redundancy-aware differential learning and goal babbling (D'Souza et al., 2001; Rolf et al., 2010), online body-schema acquisition (Hersch et al., 2008; Hoffmann et al., 2010), and adaptive Jacobian models (Stalph & Butz, 2012). Spectral-bias analysis (Rahaman et al., 2019) motivates periodic representations using sinusoidal activations, Fourier features, and multiplicative filters (Sitzmann et al., 2020; Tancik et al., 2020; Fathony et al., 2021). Physics-structured dynamics includes Deep Lagrangian Networks (Lutter et al., 2019), contact and constrained formulations (Pfrommer et al., 2021; Daems et al., 2024), manifold-projected dynamics (White et al., 2024), predictive control and humanoid extensions (Schulze et al., 2025, 2026), and Lie-group and port-Hamiltonian models (Duong et al., 2024; Wotte et al., 2024). Graph-based mechanism models (Zhang et al., 2024; Zhu et al., 2024) and differentiable kinematic constraints for task-and-motion planning (Xiao2026,Peng et al., 2026) exploit structures further for learning motion manifolds.

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.

Branch-based IK (DeMers and Kreutz-Delgado, 1992; 1998), kinematic decomposition (Ruiz de Angulo and Torras, 2008), continuous rotation encodings (Zhou et al., 2019), and autoencoding obstructions (Batson et al., 2021) establish the importance of representation learning in kinematics. Learned atlases (Korman, 2018; Schonscheck et al., 2020) already have statistical guarantees under specified denoising models (Liu et al., 2023). IKFlow and GGIK learn solution distributions (Ames et al., 2022; Limoyo et al., 2025). Recent studies address sample efficiency (Lim, 2026), singularities (Rudrasamudram and Malaichamee, 2026), and differentiable numerical IK charts (Cohn et al., 2026).

Algebraic topologists espoused their views on complexity in terms of invariants Pavešić (2019); Zapata & González (2023). Schwarz's sectional category counts local sections (Schwarz, 1966); Farber's topological complexity concerns paths between configurations (Farber, 2003); and Pavešić's map complexity concerns motion from a configuration to a task target (Pavešić, 2018). Work-map and pullback formulations describe related planning problems (Murillo and Wu, 2021; Scott, 2022).

Topology asks how many continuous rules are needed; conditioning asks how strongly a rule amplifies perturbations; statistics asks what must be estimated from data. Our contribution is to establish explicit quantitative connections and separating examples (from manifold reconstruction and computational hardness) — rather than new invariants or charts — using standard intrinsic-dimensional regression principles (Kpotufe, 2011). Experiments isolate periodicity, representation, calibration, conditioning, and redundant-solution ambiguity and distinguish inverse selection from manipulation complexity. Neither count alone determines mean squared error or sample complexity: errors may concentrate near rarely sampled regions, while several modules may share unknown parameters. Our central question is therefore *how map topology, inverse sensitivity, and unknown structure jointly enter a learning guarantee.* In particular, we establish three connections: inverse and manipulation domains retain their exact minimum counts below a geometric tolerance; bounded sensitivity yields error and validation bounds, including an exact circle-inversion law and a finite-sample certificate; and chart sensitivity, sampling probabilities, and loss conversion determine a sufficient sample budget. A shared-parameter construction separates growing topological counts from statistical cost, while structured calibration replaces a nonparametric dimension exponent with a design-matrix. In proteomics, torsional lever arms provide the molecular loss conversion, with additional assumptions for latent angles.

2 Which complexity of Kinematics for Representation and Loss?

A *fiber* is the set of configurations with the same task value. A *section* selects one of them continuously. Let Q be an n-dimensional configuration manifold and Y an r-dimensional task manifold with specified metrics. For full pose, Y may be a region of $\mathrm{SE}(3)$, with translation and rotation weights fixed in advance. Let $W \subset Y$ be an open set of reachable regular values of a smooth map F, so dF_q is surjective at every $q \in F^{-1}(W)$.

$$\text{Write} \quad p = F|_{F^{-1}(W)} : E = F^{-1}(W) \longrightarrow W. \tag{1}$$

We assume explicitly that p is a surjective Hurewicz fibration and W is path-connected. A proper submersion between manifolds without boundary is a sufficient condition by Ehresmann's theorem (Lee, 2012). Constant maximal rank alone does not ensure that a reachable image is a smooth base. Joint limits, obstacles, or singular strata require separate domain checks. For an open cover $\{U_i\}_{i=1}^{M}$ of W, a continuous inverse module is $g_i : U_i \to E$. Its task residual is $e_i(y) = d_W(p(g_i(y)), y)$. Define $B_\varepsilon(p)$ as the least M for which such a cover has $\sup_{U_i} e_i \le \varepsilon$ for every i. Thus $B_0(p) = \mathrm{secat}(p)$, using the unreduced convention $\mathrm{secat}(\mathrm{id}) = 1$. This equality is definitional. It counts domains sufficient to select *one* inverse, not fiber cardinality, network width, or the number of independent parameter vectors. The domains may overlap or be disconnected. Selecting a module may introduce discontinuities. For a global predictor $g : W \to E$ and target distribution μ, we use $R_\mu(g) = \int_W d_W(p(g(y)), y)^2 \, d\mu(y)$. This task risk differs from the supervised joint loss $\mathbb{E} d_Q(\widehat{s}(Y), s(Y))^2$ for a designated section s. In redundant IK, small task error need not imply proximity to that section. In a learning experiment, Independent forward observations take the form $(Q_j, F_\theta(Q_j) + \xi_j)$ in specified output coordinates. Inverse regression instead observes $(I_j, X_j, s_{I_j}(X_j) + \xi_j)$ in specified local charts, with an observed regime label I_j. Noise is added to the *regression response*; reversing noisy forward pairs would introduce a different errors-in-variables problem. A minimax sample count is defined over a class $\mathcal{H}$ of unknown maps, a design law, a noise law, and a loss. It cannot be determined from an individual, fully known F alone.

Table 1: Explicit complexities of the maps used in the paper. κ is a local inverse derivative norm in the stated metrics, not a global chart Lipschitz bound. The first three rows use angular, product-angular, and quotient-sphere metrics, respectively.

Map	Fiber size	B_0	M_0	κ
$z \mapsto z^k$ on S^1, $k \geq 2$	k	2	2	$1/k$
Product squaring on $\mathbb{T}^r$	2^r	$r+1$	$r+1$	$1/2$
Quaternion cover $S^3 \to \mathrm{SO}(3)$	2	4	4	1
Regular planar 2R position map	2	1	2	$\asymp d_{\mathrm{out}}^{-1/2}$

2.1 Manipulation complexity

A manipulation rule takes $(q, y) \in E \times W$ to a path $\Gamma(q, y) \in E^{[0,1]}$ with $\Gamma(q, y)(0) = q$. Its endpoint residual is $d_W(p(\Gamma(q, y)(1)), y)$. Let $M_\varepsilon(p)$ be the minimum size of an open cover of $E \times W$ supporting continuous rules with residual at most ε. For exact rules,

$$M_0(p) = \mathrm{TC}(p) = \mathrm{secat}(\pi_p), \qquad \pi_p(\gamma) = (\gamma(0), p(\gamma(1))). \tag{2}$$

This is Pavešić's manipulation complexity on the regular fibration. It differs from $\mathrm{TC}(E)$, which prescribes both configuration endpoints. For fibrations between manifolds the open-cover and sectional formulations agree; singular maps require more care (Pavešić, 2019). If E is disconnected, rules are defined componentwise and paths stay in the starting component. The fibration over connected W makes every target reachable from every component.

Below, $\mathrm{cat}(W)$ is the Lusternik–Schnirelmann category: the minimum number of open sets covering W whose inclusions can be contracted within W.

Proposition 1 (Classical bounds used in the calculations)**.** *Under the above assumptions, using unreduced category throughout,*

$$B_0(p) \leq \mathrm{cat}(W) \leq M_0(p) \leq \mathrm{TC}(W). \tag{3}$$

If p has a global section, $M_0(p) = \mathrm{TC}(W)$. Over any coefficient field, the cohomological bounds are

$$1 + \mathrm{cl}\ker p^* \leq B_0(p), \qquad 1 + \mathrm{cl}\ker \Delta_p^* \leq M_0(p), \quad \Delta_p(q) = (q, p(q)), \tag{4}$$

where $\mathrm{cl}\, I$ is the greatest length of a nonzero product in the ideal I (zero for the zero ideal).

Here singular cohomology is used on the manifold bases; arbitrary spaces require additional care with the cohomology theory. These are classical estimates (Schwarz, 1966; Pavešić, 2019); Appendix B gives proofs and computations. Learning additionally requires sensitivity, sampling information, and a specified class of unknown maps. Eilenberg–MacLane and Moore spaces provide useful configuration and workspace models. A plane with point obstacles is both types; three-dimensional space with point obstacles has Moore homotopy type. For a point robot with at least two planar obstacles, or any positive number of spatial point obstacles, $B_0 = 1$ but $M_0 = 3$. Appendix G.2 derives these counts and treats singular networks and punctured configuration tori. The two classes overlap and do not exhaust constrained kinematic spaces. Homotopy dimension also need not equal regression dimension.

3 Topological Certificate for Kinematics Learning

Assume that W has a *path-interpolation radius* $\rho > 0$: for $d_W(u, v) < \rho$, a path from u to v can be chosen continuously in (u, v). On a compact Riemannian manifold without boundary, any ρ no larger than its injectivity radius suffices, using the unique short geodesic. For an open regular workspace, such a radius is an additional assumption; paths must remain in W.

Theorem 2 (Uniform-error stability)**.** *For a surjective Hurewicz fibration $p : E \to W$ with path-interpolation radius ρ, and $0 \leq \varepsilon < \rho$, $B_\varepsilon(p) = \mathrm{secat}(p), \qquad M_\varepsilon(p) = \mathrm{TC}(p)$.*

Proof. For an approximate inverse module, interpolate from $p \circ g_i$ to the target and lift this homotopy starting at g_i. Its endpoint is an exact section on the same domain. For an approximate manipulation

rule, lift the short correction path starting at its endpoint and append the lift to the original path. Resulting continuous rule retains its prescribed start, and reaches the exact target. ∴ neither approximate cover can have fewer domains than its exact counterpart. Exact rules give the reverse inequalities. Both equalities follow from classical homotopy lifting (Hatcher, 2002; Schwarz, 1966); their role here is to express the obstruction at a specified tolerance. Related questions about approximate roots are studied by Lau and Marin-Marquez (2026). The correction controls domain count, but need not preserve network architecture, path length, or Lipschitz constants. □

3.1 A CERTIFICATE FROM FINITELY MANY TARGETS

The statement also applies when learning uses a surrogate forward map. If $\widehat{p} : E \to W$ satisfies $\sup_q d_W(\widehat{p}(q), p(q)) \le \eta$, then any inverse or manipulation cover with residual at most ε under $\widehat{p}$ must have at least $B_0(p)$ or $M_0(p)$ domains, respectively, whenever $\eta + \varepsilon < \rho$. The triangle inequality transfers the residual to the true map; $\widehat{p}$ itself need not be a fibration.

Theorem 3 (Average error and independent validation)**.** *Suppose W is compact, $B_0(p) > 1$, $\mathrm{Lip}(p) \le L_p$, and $\mu(B_W(y,t)) \ge at^r$ for every y and $0 < t \le t_0$. Every global K-Lipschitz $g : W \to E$ satisfies, with $H = 1 + L_pK$ and $b = \min\{t_0, \rho/(2H)\}$,*

$$R_\mu(g) \ge \frac{a\rho^2}{4} b^r, \qquad \mu\{e_g \ge \rho/2\} \ge ab^r. \tag{5}$$

Conditioned on a trained predictor, V independent validation targets detect a residual at least $\rho/2$ with probability at least $1 - \delta$ if

$$V \ge (ab^r)^{-1} \log(1/\delta). \tag{6}$$

Appendix C proves this by integrating the residual around a point where it is at least ρ. The same calculation uses manipulation complexity for a valid continuous global path predictor. If $E \times W$ is compact, $M_0(p) > 1$, its endpoint map is K-Lipschitz in the product metric, and the design has lower ball mass at^{n+r}, Eqs. (5)–(6) hold with r replaced by $n + r$ (Appendix B). Thus planning can have an obstruction even when $B_0(p) = 1$. The lower ball-mass assumption prevents the test distribution from avoiding the difficult region. These are validation bounds; neither is a training lower bound. For targets $S = \{y_j\}$, define their fill distance $h_S = \sup_{y \in W} \min_j d_W(y, y_j)$ and measured maximum residual $e_S = \max_j e_g(y_j)$. Every K-Lipschitz predictor obeys

$$\|e_g\|_\infty \le e_S + (1 + L_pK)h_S; \qquad B_0(p) > 1 \implies e_S + (1 + L_pK)h_S \ge \rho. \tag{7}$$

This retains the measured residual instead of assuming exact interpolation. When $e_S < \rho$ and $h_S > 0$, it requires $K \ge [(\rho - e_S)/h_S - 1]/L_p$. Fitting a dense set across an obstruction therefore forces large sensitivity. Low RMS error can hide the resulting seam; the circle calculation makes this precise.

3.2 EXACT LIPSCHITZ–RISK FRONTIER ON THE CIRCLE

Let $p_k : S^1 \to S^1$, $p_k(z) = z^k$, with integer $k \ge 2$, angular geodesic distance, and uniform target measure. This is the orientation map of an ideal periodic transmission. Its sectional category is two: two slit circles admit root selections, while a global section would require $k \deg(g) = 1$.

Theorem 4. *Every continuous g has maximum task residual π. In contrast, two local modules can achieve zero residual.. For $k \ge 2$ and $K \ge 1/k$,*

$$\inf_{\substack{g: S^1 \to S^1 \\ \mathrm{Lip}(g) \le K}} R_\mu(g) = \frac{\pi^2}{3(1 + kK)}. \tag{8}$$

Degree forces the residual to reach π. Integrating the squared Lipschitz cone below that peak gives the lower bound; a piecewise-linear phase attains it. Appendix C gives the construction and full proof. Thus RMS tolerance ε forces $K \ge (\pi^2/(3\varepsilon^2) - 1)/k$ for a global continuous head. For the extremizer, the mass with residual exceeding $\pi/2$ is $1/[2(1 + kK)]$. Detecting that region reliably requires validation size on the order of K at fixed k and confidence. This explicit tradeoff makes a seam visible to both approximation theory and evaluation design.

4 WHAT TOPOLOGY DOES NOT DETERMINE ABOUT LEARNING

4.1 UNBOUNDED DOMAIN COUNT WITH A SHARED STATISTICAL PARAMETER

Proposition 5 (A separation from sample complexity)**.** *For every $r \geq 1$, there is a one-parameter family $p_\alpha : \mathbb{T}^r \to \mathbb{T}^r$ with $B_0(p_\alpha) = M_0(p_\alpha) = r + 1$ for every α, but noisy forward observations suffice to produce an inverse atlas with uniform task error at most $\varepsilon < 1$ using*

$$N = O\big(1 + \sigma^2 \varepsilon^{-2} \log(1/\delta)\big) \tag{9}$$

samples, with constants independent of r. Here task distance is the unnormalized Euclidean product of angular distances; observations have independent Gaussian noise of variance σ^2 in each real output component.

Proof. Use $p_\alpha(z_1, \ldots, z_r) = (e^{i\alpha} z_1^2, z_2^2, \ldots, z_r^2)$. Pullback annihilates the degree-one mod-two cohomology generators, whose r-fold product is nonzero. Equations (3)–(4) and $\mathrm{cat}(\mathbb{T}^r) = \mathrm{TC}(\mathbb{T}^r) = r+1$ give both complexities. Multiplying the first noisy output by $\overline{z_1^2}$ gives a complex Gaussian observation with mean $e^{i\alpha}$. Estimate its phase from the sample mean. Concentration gives Eq. (9). Every section of the estimated map incurs only this phase error in the first task coordinate. Appendix D supplies the details and the analogous manipulation construction. □

The atlas still needs $r + 1$ continuous domains; their functions share one learned parameter. The same statistical estimate supports a manipulation cover with $r + 1$ domains. Sample count is therefore independent of both topological complexities in this family. Storing outputs, constructing covers, and evaluating paths have additional costs, which this statement does not bound. A full Lipschitz regression class with s input coordinates and fixed response noise has the familiar accuracy exponent $s + 2$. Thus forward prediction uses $s = n$, whereas a designated inverse section uses $s = r$. With J observed regimes, a factor J can arise from learning independent functions under balanced sampling. It is not implied by either topological count. Proposition 6 in Appendix E states an explicit uniform upper bound and an integrated-risk lower bound, with their different loss criteria kept separate. We now specialize that upper calculation to a supplied kinematic atlas.

4.2 A SAMPLE BUDGET FOR A SPECIFIED INVERSE ATLAS

The local bound can incorporate the actual map. Refine a supplied inverse atlas into J normalized task charts. In chart i, let the coordinate target $f_i : [0, 1]^r \to C_i \subset \mathbb{R}^m$ be κ_i-Lipschitz, where C_i is closed and convex. Let its feasible decoder $\psi_i : C_i \to E$ be D_i-Lipschitz. The decoded target is an exact section. Assume known chart labels of probability $\pi_i > 0$, uniformly lower-bounded conditional design densities, and response noise scale σ_i. Chart coverage and output feasibility are explicit assumptions, detailed in Appendix E.4. Put $\tau_i = L_p D_i$, the conversion from coordinate error to task error. For $0 < \varepsilon < \min_i \tau_i$, define

$$v_i = \left(1 + \frac{\kappa_i \tau_i}{\varepsilon}\right)^r, \qquad w_i = v_i \left(1 + \frac{m \sigma_i^2 \tau_i^2}{\varepsilon^2}\right). \tag{10}$$

A separate histogram in each chart, projected to C_i before decoding, has uniform task error at most ε on the covered evaluation workspace with probability $1 - \delta$ if

$$N \geq C_r \max_i \frac{w_i}{\pi_i} \log \frac{C_r m \sum_i v_i}{\delta}. \tag{11}$$

Here C_r also depends on the density lower bound. If chart sampling can be chosen, the leading factor has the explicit optimum

$$\min_{\pi_i > 0,\, \sum_i \pi_i = 1} \max_i \frac{w_i}{\pi_i} = \sum_i w_i, \qquad \pi_i^* = \frac{w_i}{\sum_j w_j}. \tag{12}$$

Thus a useful atlas is judged by its chart costs as well as its size. Its underlying continuous sections require $J \geq B_0(p)$, but a minimum-size topological cover need not minimize Eq. (11). The histogram establishes statistical sufficiency and is generally discontinuous; it is not an architecture realizing that cover. Shared parameter models can improve this separate-chart budget, as Proposition 5 shows.

4.3 CONDITIONING AND STRUCTURED FORWARD KINEMATICS

For a differentiable section s, differentiation gives $dF_{s(y)}\, ds_y = \boldsymbol{I}, \qquad \|ds_y\| \geq \frac{1}{\sigma_r(dF_{s(y)})}$, where σ_r is the smallest nonzero singular value. If $n = r$, the local inverse has $ds = (dF)^{-1}$. On a convex task chart where $\sigma_r \geq \gamma$, it is γ^{-1}-Lipschitz, up to chart distortion. For $n > r$, null-space motion can make a selected section much less regular; a lower bound on σ_r alone does not provide that upper bound. Large local Lipschitz constants worsen the sufficient sample budget in Eq. (11), but do not themselves establish a minimax lower bound for a particular mechanism.

A structured forward family gives a complementary rate. For a planar serial arm with known joint offsets and unknown link lengths $\boldsymbol{\ell} \in \mathbb{R}^n$,

$$F_{\boldsymbol{\ell}}(q) = \sum_{j=1}^{n} \ell_j e^{i(q_1+\cdots+q_j)}. \tag{13}$$

Stack real and imaginary features into $\boldsymbol{A} \in \mathbb{R}^{2N\times n}$ and fit ordinary least squares. For full-column-rank $\boldsymbol{A}$, independent Cartesian noise $N(0, \sigma^2)$, and uniform independent test joints, the exact conditional expected forward risk is

$$\mathbb{E}_{\xi, Q_{\mathrm{test}}} |F_{\widehat{\boldsymbol{\ell}}}(Q_{\mathrm{test}}) - F_{\boldsymbol{\ell}}(Q_{\mathrm{test}})|^2 = \sigma^2 \operatorname{tr}((\boldsymbol{A}^\top \boldsymbol{A})^{-1}). \tag{14}$$

Define the design-dependent calibration factor

$$d_{\mathrm{cal}}(\boldsymbol{A}) = \operatorname{tr}\left[\left(\boldsymbol{A}^\top \boldsymbol{A}/N\right)^{-1}\right], \qquad R_{\mathrm{cal}}(\boldsymbol{A}) = \frac{\sigma^2}{N} d_{\mathrm{cal}}(\boldsymbol{A}). \tag{15}$$

Unlike a parameter count, d_{cal} increases under poor excitation. If $\boldsymbol{A}^\top \boldsymbol{A} \succeq cN\boldsymbol{I}$, then $d_{\mathrm{cal}} \leq n/c$. It enters the risk directly, whereas B_0 and M_0 do not. This calculation exploits the n unknown lengths and the known trigonometric structure. Appendix F gives the proof, a confidence bound, and the extension to nonuniform test designs.

4.4 PROTEIN AND RNA KINEMATICS ON TORSION SPACES

With molecular connectivity and local geometry fixed, d independent periodic dihedrals give $\boldsymbol{\theta} \in \mathbb{T}^d$. This is an Eilenberg–MacLane space $K(\mathbb{Z}^d, 1)$: its higher homotopy groups vanish, but its higher homology generally does not. For $d > 1$ it is therefore not a Moore space. Constraints can produce neither type. Appendix G develops the distinction, obstacle models, and the Moore mapping cone of circle inversion. Protein backbone and side-chain models already exploit angular coordinates (Wu et al., 2024; Zhang et al., 2023). For RNA, sugar-pucker compatibility and shared angles between adjacent suites complicate the independent-torsion model (Murray et al., 2003). Appendix I specifies the coordinates and constraints. Write

$$\Theta_\lambda = \{\boldsymbol{\theta} \in \mathbb{T}^d : C_\lambda(\boldsymbol{\theta}) = 0,\ S_\lambda(\boldsymbol{\theta}) > 0\}, \qquad F = \mathcal{O} \circ H_\lambda : \Theta_\lambda \to Y, \tag{16}$$

where H_λ reconstructs anchored Cartesian coordinates, C_λ enforces closure, S_λ enforces strict clearance, and $\mathcal{O}$ selects the observable. Loop closure is a molecular IK problem (Coutsias et al., 2004). A regular stratum with $\operatorname{rank} dC_\lambda = c$ has dimension $s = d - c$; applying our bounds still requires the fibration and chart assumptions. For the ideal phase-observation map $p_{\boldsymbol{A}}(\boldsymbol{\theta}) = \boldsymbol{A}\boldsymbol{\theta} \bmod 2\pi$, let $\boldsymbol{A} \in \mathbb{Z}^{r\times d}$ have full row rank and positive Smith invariant factors $a_1 \mid \cdots \mid a_r$. Then $B_0(p_{\boldsymbol{A}}) = 1 + \#\{j : a_j > 1\}, \qquad M_0(p_{\boldsymbol{A}}) = r + 1$. Appendix H proves this using the induced map on π_1 and cup products. Identity reconstruction has $B_0 = 1$ and $M_0 = d + 1$; a large latent torus alone does not force many inverse modules. Actual Cartesian observation maps need their own calculation. Let b_k bound the weighted Cartesian norm of torsional Jacobian column k and put $\mathcal{L} = (\sum_k b_k^2)^{1/2}$. Proposition 9 gives $\operatorname{Lip}(H_\lambda) \leq \mathcal{L}$ for a fixed-geometry tree decoder. If $\mathcal{O}$ is $L_{\mathcal{O}}$-Lipschitz, use $\tau_i = L_{\mathcal{O}}\mathcal{L}D_i$ in Eq. (11). Its exponent is r for designated inverse sections; observed-coordinate forward regression uses s. Known geometry needs no learned forward map. Unobserved angles additionally require an encoder and identifiability assumptions (See Appendix I and Open X-Embodiment Collaboration et al. (2024); Szikszai et al. (2024); Bose et al. (2024); Abramson et al. (2024); Verma et al. (2025) for topology-aware reconstruction).

Table 2: Measured planar 2R task errors (mean ± sample standard deviation over five seeds). Lengths are 1 and 0.7; near-cut evaluation uses 396 held-out targets. See appendix A for analysis.

N	Output	Task RMSE	95th percentile	Near-cut RMSE
256	Angles	0.611 ± 0.064	1.339 ± 0.238	2.086 ± 0.098
256	Periodic	0.062 ± 0.014	0.121 ± 0.025	0.059 ± 0.029
1024	Angles	0.372 ± 0.030	0.539 ± 0.083	1.658 ± 0.063
1024	Periodic	0.035 ± 0.002	0.068 ± 0.005	0.038 ± 0.013
4096	Angles	0.321 ± 0.026	0.436 ± 0.012	1.476 ± 0.131
4096	Periodic	0.017 ± 0.003	0.032 ± 0.006	0.015 ± 0.003

5 CONTROLLED NUMERICAL STUDIES

We first evaluate the theoretical bounds using simulated benchmarks before extending the analysis to a spatial arm in Section 5.1. Appendix K provides protocol details and manipulation checks, and Appendix I evaluates the molecular decoder bound on spatial chains.

For $k \in \{2, 3, 5\}$ and $K \in \{1, 2, 4, 8, 16, 32, 64\}$, quadrature of Eq. (25) agrees with Eq. (8) to below 2×10^{-15} absolute error. Figure 3 in Appendix C illustrates decreasing mean error alongside a persistent peak. Independent trials match the predicted probability $1 - (1 - [2(1 + kK)]^{-1})^V$ of detecting residuals above $\pi/2$. For $\ell_1 = 1, \ell_2 = 0.7$, we sample uniformly across workspace area on the positive-elbow branch with $q_2 \in [0.2, \pi - 0.2]$. Tanh MLPs with equal hidden widths $(64, 64)$ predict principal angles or sine/cosine pairs (4,482 versus 4,612 parameters). They share data, optimization settings, five random seeds, and an independent 12,000-target test set. For $N \in \{256, 1024, 4096\}$ training/validation pairs, 15% are reserved for early stopping. Periodic outputs consistently reduce task error under this protocol (Table 2, Fig. 1), confirming that the performance degradation near principal-angle cuts stems from representation choice rather than topological multiplicity ($B_0 = 1$). Decoding via $\mathrm{atan2}$ remains well-defined across all periodic predictions. We estimate link lengths across $n \in \{2, 4, 8, 16\}$ and $N \in \{32, 64, 128, 512, 2048\}$, applying Cartesian noise with standard deviation 0.03 over 200 repetitions per setting. Figure 1 (bottom row) compares the exact integrated test error $\|\widehat{\boldsymbol{\ell}} - \boldsymbol{\ell}\|^2$ against $\sigma^2 n/N$. For $n = 16$, the mean ratio d_{cal}/n drops from 1.3119 at $N = 32$ to 1.0037 at $N = 2048$, tracking the empirical deviation from a pure parameter-count scaling (Table 10). Separately, 100 configurations approaching full extension yield a log–log inverse-sensitivity slope of -0.49998 for $q_2 < 0.1$, matching the analytic inverse derivative to within 4.5×10^{-7} relative error via finite differences.

5.1 VALIDATION ON A REDUNDANT SPATIAL SEVEN-JOINT ROBOT

When a spatial arm has more joint coordinates than pose dimensions, inverse kinematics requires selecting from a continuum of valid configurations. We evaluate this setting using a synthetic seven-revolute-joint (7R) manipulator with a six-dimensional end-effector pose to examine two questions: how redundancy affects section-sensitivity calculations, and how supervised regressors respond when presented with multiple valid solutions for a single pose input.

Geometry and designated solutions. We adopt standard Denavit–Hartenberg parameters with $a_j = 0$, offsets $\boldsymbol{d} = (0.34, 0, 0.40, 0, 0.40, 0, 0.126)$ m, and twists $\boldsymbol{\alpha} = (-1, 1, 1, -1, -1, 1, 0)\pi/2$. Around $\boldsymbol{q}_0 = (0.4, -0.65, 0.7, 1.05, -0.5, 0.8, 0.2)$ rad, target poses vary independently by ± 15 mm in translation and ± 0.06 rad in world-frame rotation-vector coordinates. Setting q_7 to either -0.1 or 0.5 rad defines two local inverse selections. Numerical IK produces both labels for every pose, reaching residuals below 10^{-10} in metres and radians. Appendix L contains complete geometric specifications and solver settings. Letting $\boldsymbol{J} = [\boldsymbol{J}_v^\top, \boldsymbol{J}_\omega^\top]^\top$ denote the geometric Jacobian with weighted task tangent $(\Delta\boldsymbol{x}, \lambda\Delta\boldsymbol{\omega})$ ($\lambda = 0.25$ m/rad), we define $\boldsymbol{K} = \mathrm{diag}(1, 1, 1, \lambda, \lambda, \lambda)\boldsymbol{J}$. For either fixed-$q_7$ section, the derivative in this tangent frame satisfies

$$\boldsymbol{D}s = \begin{bmatrix} \boldsymbol{K}^{-1}_{[:,1:6]} \\ \boldsymbol{0}^\top \end{bmatrix}, \qquad \boldsymbol{K}\boldsymbol{D}s = \boldsymbol{I}_6, \qquad \|\boldsymbol{D}s\|_2 \, \sigma_6(\boldsymbol{K}) \geq 1. \tag{17}$$

Across 256 sampled section points, finite-difference evaluations match this formulation to within 4.08×10^{-8} relative error. The sensitivity ratio in Eq. (17) ranges from 1.49 to 2.86, showing that

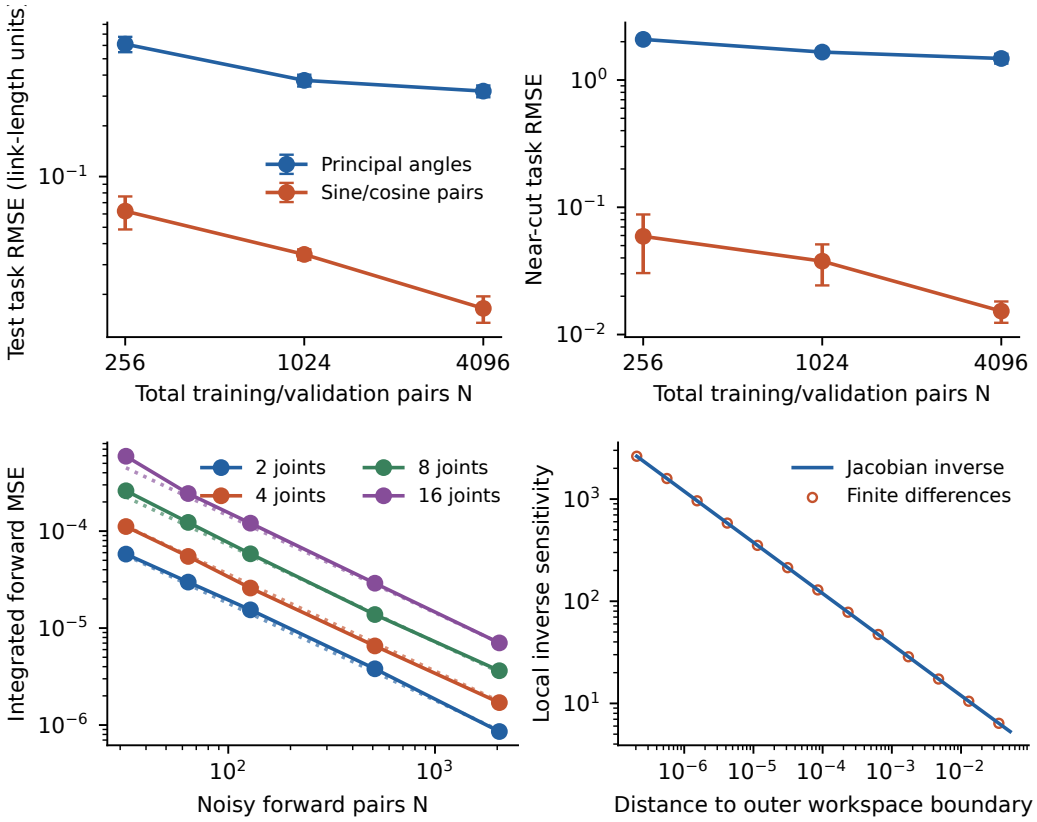

Figure 1: Planar 2R inverse learning on a single domain ($B_0 = 1$). Bars indicate one standard deviation over five seeds. The top-right panel evaluates targets whose principal first-joint angle falls within 0.1 rad of a coordinate cut. Bottom row: Two distinct complexity mechanisms. Left: mean integrated forward error over 200 trials (solid) versus $\sigma^2 n/N$ (dotted). Right: growing inverse sensitivity near a singular boundary.

Table 3: Spatial 7R full-pose prediction with 4,096 unique training/validation poses. Mean $\pm$ sample standard deviation over five seeds. Success requires position error ≤ 5 mm and orientation error $\leq 1°$. Every test pose is evaluated with both designated solutions.

Method	Position (mm)	Orientation (°)	Success (%)
Pooled / angles	77.36 ± 8.25	6.44 ± 0.65	0.00 ± 0.00
Pooled / periodic	75.52 ± 3.56	6.45 ± 0.43	0.00 ± 0.00
Labelled / angles	6.41 ± 1.85	0.77 ± 0.30	48.36 ± 18.08
Labelled / periodic	3.02 ± 0.82	0.36 ± 0.11	91.05 ± 7.50

these specific sections exhibit higher sensitivity than the minimum-norm right inverse. Furthermore, 7,000 Jacobian column checks across 1,000 configurations yield a maximum numerical error of 2.98×10^{-10} in metre/radian coordinates. We evaluate four tanh MLP variants crossing two structural choices: raw angles versus trigonometric (sine/cosine) outputs, and pooled training data versus explicit binary selection labels. Each of $N \in \{256, 1024, 4096\}$ distinct training poses includes both valid joint solutions. We reserve 15% of poses for early stopping (keeping paired labels grouped) and train across five random seeds (60 fits total). All networks use two hidden layers of width 64 (5,063–5,582 parameters), and performance is measured on 8,192 held-out target poses. At $N = 4096$, the labelled periodic model achieves 3.02 mm position RMSE and 0.365° orientation RMSE (Table 3). In contrast, unlabelled pooled periodic prediction retains 75.52 mm position error because conflicting targets blur under the joint-space loss. Directly averaging the two ground-truth joint vectors prior to forward kinematics similarly yields 75.23 mm position RMSE and 6.37° orientation RMSE,

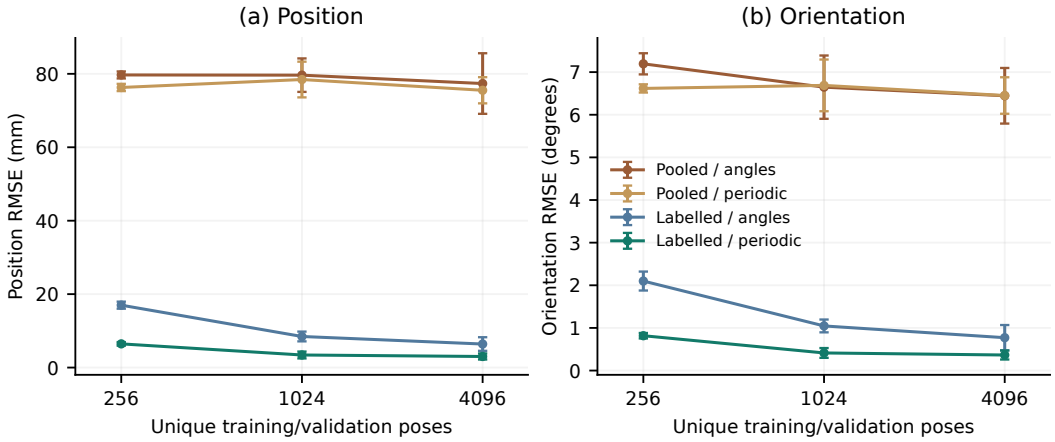

Figure 2: Spatial 7R learning: Explicit selection labels resolve branch ambiguity, whereas periodic representations alone cannot.

demonstrating that linear combinations of valid joint configurations generally fail to produce valid solutions. While branch-aware formulations mitigate this ambiguity (DeMers and Kreutz-Delgado, 1998), this experiment highlights the effect in tandem with our geometric sensitivity bound. For comparison, damped numerical IK initialized at $\boldsymbol{q}_0$ converges on all 8,192 test poses within four iterations (median 3), reaching position and orientation RMSEs below 2×10^{-11} in metres and radians. These numerical experiments confirm the behavior of local redundant pose representations and section-sensitivity bounds.

6 Conclusion

This work connects three sources of complexity in kinematic learning: *topology*, which governs continuous inverse-selection and manipulation domain counts; *conditioning*, which controls inverse sensitivity and enters approximation, risk, and validation guarantees; and *statistical structure*, which governs sample requirements through the unknown model class, shared parameters, dimension, noise, and sampling design. If the restricted smooth kinematic map is a surjective Hurewicz fibration over a path-connected base, continuous path interpolation of radius ρ preserves the exact domain counts for uniform task error below ρ. The sharp circle-inversion frontier, finite-target certificate, shared-parameter separation, and chart-dependent sample budget quantify these connections. Planar and spatial 7-DoF experiments illustrate periodic representations, selection labels, and section sensitivity; torsional decoder bounds connect molecular angular and Cartesian errors. The central contribution is a *transferable analytical framework linking topology, conditioning, and statistical structure to representation choices and data requirements*. It carries over to parallel manipulators through regular loop-closure manifolds and to free-floating manipulators through base-pose and joint configurations, supporting the analysis of learning architectures and sampling on their regular-fibration regions.

Reproducibility Statement

The code includes generators, seeds, hyperparameters, run-level outputs, MLP weights, and plots; Appendix K details the protocols. Additional checks are provided by `analyze_map_complexity.py` and `check_molecular_kinematics.py` in `code/`.

AI Use Statement

ChatGPT/Codex assisted with code implementation for synthetic data generation, proof corrections, and concise editing. The accompanying programs generated the numerical results. The authors retain responsibility for independent verification and all final results.

REFERENCES

Barrett Ames, Jeremy Morgan, and George Konidaris. IKFlow: Generating diverse inverse kinematics solutions. *IEEE Robotics and Automation Letters*, 7(3):7177–7184, 2022. https://arxiv.org/abs/2111.08933.

Joshua Batson, C. Grace Haaf, Yonatan Kahn, and Daniel A. Roberts. Topological obstructions to autoencoding. *Journal of High Energy Physics*, 2021(4):280, 2021. https://arxiv.org/abs/2102.08380.

Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *International Conference on Learning Representations*, 2024. https://arxiv.org/abs/2302.03660.

Thomas Cohn, Seiji Shaw, Harel Biggie, Travis Manderson, Nicholas Roy, and Russ Tedrake. Planning along differentiable charts of constraint manifolds with general-purpose IK solvers. arXiv:2609.10905, 2026.

Evangelos A. Coutsias, Chaok Seok, Matthew P. Jacobson, and Ken A. Dill. A kinematic view of loop closure. *Journal of Computational Chemistry*, 25(4):510–528, 2004. https://doi.org/10.1002/jcc.10416.

David DeMers and Kenneth Kreutz-Delgado. Learning global direct inverse kinematics. In *Advances in Neural Information Processing Systems 4*, pp. 589–595, 1992. Conference held in 1991. https://proceedings.neurips.cc/paper/1991.

David DeMers and Kenneth Kreutz-Delgado. Learning global properties of nonredundant kinematic mappings. *The International Journal of Robotics Research*, 17(5):547–560, 1998. https://doi.org/10.1177/027836499801700506.

Michael Farber. Topological complexity of motion planning. *Discrete & Computational Geometry*, 29:211–221, 2003. doi:10.1007/s00454-002-0760-9.

Christopher R. Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Minimax manifold estimation. *Journal of Machine Learning Research*, 13:1263–1291, 2012. https://jmlr.org/papers/v13/genovese12a.html.

Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002. https://pi.math.cornell.edu/~hatcher/AT/ATpage.html.

Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi Jaakkola. Torsional diffusion for molecular conformer generation. In *Advances in Neural Information Processing Systems*, 2022. https://arxiv.org/abs/2206.01729.

Michael Kapovich and John J. Millson. Universality theorems for configuration spaces of planar linkages. *Topology*, 41(6):1051–1107, 2002. https://doi.org/10.1016/S0040-9383(01)00034-9.

Bobak T. Kiani, Jason Wang, and Melanie Weber. Hardness of learning neural networks under the manifold hypothesis. In *Advances in Neural Information Processing Systems 37*, pp. 5661–5696, 2024. https://arxiv.org/abs/2406.01461.

Arlene K. H. Kim and Harrison H. Zhou. Tight minimax rates for manifold estimation under Hausdorff loss. *Electronic Journal of Statistics*, 9(1):1562–1582, 2015. doi:10.1214/15-EJS1039.

Eric O. Korman. Autoencoding topology. arXiv:1803.00156, 2018.

Samory Kpotufe. k-NN regression adapts to local intrinsic dimension. In *Advances in Neural Information Processing Systems 24*, pp. 729–737, 2011. https://arxiv.org/abs/1110.4300.

Joshua Lau and Vicente Marin-Marquez. Continuous approximate roots of polynomial equations via shape theory. *Topology and its Applications*, 390:109894, 2026. https://arxiv.org/abs/2510.01613.

John M. Lee. *Introduction to Smooth Manifolds*. Springer, second edition, 2012.

Dong-Won Lim. How many training samples are needed for the inverse kinematics solutions by artificial neural networks. arXiv:2605.23583, 2026.

Oliver Limoyo, Filip Marić, Matthew Giamou, Petra Alexson, Ivan Petrović, and Jonathan Kelly. Generative graphical inverse kinematics. *IEEE Transactions on Robotics*, 41:1002–1018, 2025. https://arxiv.org/abs/2209.08812.

Hao Liu, Alex Havrilla, Rongjie Lai, and Wenjing Liao. Deep nonparametric estimation of intrinsic data structures by chart autoencoders: Generalization error and robustness. arXiv:2303.09863, 2023.

Runze Ma, Zhongyue Zhang, Zichen Wang, Chenqing Hua, Jiahua Rao, Zhuomin Zhou, and Shuangjia Zheng. RiboFlow: Conditional de novo RNA co-design via synergistic flow matching. In *Advances in Neural Information Processing Systems*, 2025.

Aniceto Murillo and Jie Wu. Topological complexity of the work map. *Journal of Topology and Analysis*, 13(1):219–238, 2021. doi:10.1142/S179352532050003X.

Laura J. W. Murray, W. Bryan Arendall III, David C. Richardson, and Jane S. Richardson. RNA backbone is rotameric. *Proceedings of the National Academy of Sciences*, 100(24):13904–13909, 2003. https://doi.org/10.1073/pnas.1835769100.

Petar Pavešić. Complexity of the forward kinematic map. *Mechanism and Machine Theory*, 117:230–243, 2017. doi:10.1016/j.mechmachtheory.2017.07.015.

Petar Pavešić. A topologist's view of kinematic maps and manipulation complexity. *Contemporary Mathematics*, 702:61–83, 2018. https://arxiv.org/abs/1707.03899.

Petar Pavešić. Topological complexity of a map. *Homology, Homotopy and Applications*, 21(2):107–130, 2019. https://arxiv.org/abs/1809.09021.

Vishnu Rudrasamudram and Hariharasudan Malaichamee. Singularity avoidance in inverse kinematics: A unified treatment of classical and learning-based methods. arXiv:2604.13405, 2026.

Vicente Ruiz de Angulo and Carme Torras. Learning inverse kinematics: Reduced sampling through decomposition into virtual robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 38(6):1571–1577, 2008. doi:10.1109/TSMCB.2008.928232.

Stefan C. Schonscheck, Jie Chen, and Rongjie Lai. Chart auto-encoders for manifold structured data. arXiv:1912.10094, version 2, 2020.

Albert S. Schwarz. The genus of a fiber space. *American Mathematical Society Translations, Series 2*, 55:49–140, 1966.

Jamie Scott. On the topological complexity of maps. *Topology and its Applications*, 314:108094, 2022. https://arxiv.org/abs/2011.10646.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2009.

Kevin E. Wu, Kevin K. Yang, Rianne van den Berg, Sarah Alamdari, James Y. Zou, Alex X. Lu, and Ava P. Amini. Protein structure generation via folding diffusion. *Nature Communications*, 15:1059, 2024. https://doi.org/10.1038/s41467-024-45051-2.

Yangtian Zhang, Zuobai Zhang, Bozitao Zhong, Sanchit Misra, and Jian Tang. DiffPack: A torsional diffusion model for autoregressive protein side-chain packing. In *Advances in Neural Information Processing Systems*, 2023. https://arxiv.org/abs/2306.01794.

Zichen Zhong, Haoliang Sun, Yukun Zhao, Yongshun Gong, and Yilong Yin. Riemannian MeanFlow for one-step generation on manifolds. arXiv:2603.10718, 2026.

Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. On the continuity of rotation representations in neural networks. In *Proceedings of CVPR*, pp. 5745–5753, 2019. https://arxiv.org/abs/1812.07035.

Z. Geng and L. Haynes. Neural network solution for the forward kinematics problem of a Stewart platform. In *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, 1991. https://doi.org/10.1109/ROBOT.1991.132029.

C. M. Bishop. Mixture density networks. Technical Report NCRG/94/004, Neural Computing Research Group, Aston University, Birmingham, U.K., 1994.

A. D'Souza, S. Vijayakumar, and S. Schaal. Learning inverse kinematics. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 298–303, 2001. https://doi.org/10.1109/IROS.2001.973374.

R. Köker et al. A study of neural network based inverse kinematics solution for a three-joint robot. *Robotics and Autonomous Systems*, 49:227–234, 2004. https://doi.org/10.1016/j.robot.2004.09.010.

M. Hersch, E. L. Sauser, and A. Billard. Online learning of the body schema. *International Journal of Humanoid Robotics*, 5(2):161–181, 2008. https://doi.org/10.1142/S0219843608001376.

M. Rolf, J. J. Steil, and M. Gienger. Goal babbling permits direct learning of inverse kinematics. *IEEE Transactions on Autonomous Mental Development*, 2(3):216–229, 2010. https://doi.org/10.1109/TAMD.2010.2062511.

M. Hoffmann, H. Marques, A. Arieta, H. Sumioka, M. Lungarella, and R. Pfeifer. Body schema in robotics: A review. *IEEE Transactions on Autonomous Mental Development*, 2(4):304–324, 2010. https://doi.org/10.1109/TAMD.2010.2086454.

P. O. Stalph and M. V. Butz. Learning local linear Jacobians for flexible and adaptive robot arm control. *Genetic Programming and Evolvable Machines*, 13:137–157, 2012. https://doi.org/10.1007/s10710-011-9147-0.

M. Lutter, C. Ritter, and J. Peters. Deep Lagrangian networks: Using physics as model prior for deep learning. In *International Conference on Learning Representations (ICLR)*, 2019.

N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. A. Hamprecht, Y. Bengio, and A. Courville. On the spectral bias of neural networks. In *Proc. 36th International Conference on Machine Learning (ICML)*, vol. 97, pp. 5301–5310, 2019.

V. Sitzmann, J. N. P. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 7462–7473, 2020.

M. Tancik, P. P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. T. Barron, and R. Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 7537–7547, 2020.

R. Fathony, A. K. Sahu, D. Willmott, and J. Z. Kolter. Multiplicative filter networks. In *International Conference on Learning Representations (ICLR)*, 2021.

S. Pfrommer, M. Halm, and M. Posa. ContactNets: Learning discontinuous contact dynamics with smooth, implicit representations. In *Proc. Conference on Robot Learning (CoRL)*, vol. 155, pp. 2279–2291, 2021.

R. Daems, J. Taets, F. Wyffels, and G. Crevecoeur. KeyCLD: Learning constrained Lagrangian dynamics in keypoint coordinates from images. *Neurocomputing*, 573:127175, 2024.

T. Duong, A. Altawaitan, J. Stanley, and N. A. Atanasov. Port-Hamiltonian neural ODE networks on Lie groups for robot dynamics learning and control. *IEEE Transactions on Robotics*, 40:3695–3715, 2024. https://doi.org/10.1109/TRO.2024.3428433.

Y. P. Wotte, F. Califano, and S. Stramigioli. Optimal potential shaping on SE(3) via neural ordinary differential equations on Lie groups. *The International Journal of Robotics Research*, 43(14):2221–2244, 2024. https://doi.org/10.1177/02783649241256044.

H. Zhu, W. Xu, J. Huang, and J. Li. DisGNet: A distance graph neural network for forward kinematics learning of Gough-Stewart platform. *arXiv preprint arXiv:2402.09077*, 2024. https://arxiv.org/abs/2402.09077.

Y. Peng et al. RobKiNet: Robotic kinematics informed neural network for optimal robot configuration prediction. *Robotics and Autonomous Systems*, 105541, 2026. https://doi.org/10.1016/j.robot.2026.105541.

Z. Zhang, L. Yang, C. Sun, W. Shang, and J. Pan. CafkNet: GNN-empowered forward kinematic modeling for cable-driven parallel robots. *arXiv preprint arXiv:2402.18420*, 2024. https://arxiv.org/abs/2402.18420.

A. White, A. Büttner, M. Gelbrecht, V. Duruisseaux, N. Kilbertus, F. Hellmann, and N. Boers. Projected neural differential equations for learning constrained dynamics. *arXiv preprint arXiv:2410.23667*, 2024. https://arxiv.org/abs/2410.23667.

L. Schulze, J. Peters, and O. Arenz. Context-aware deep Lagrangian networks for model predictive control. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6939–6946, 2025. https://doi.org/10.1109/IROS60139.2025.11246292.

L. Schulze, J. D. Negri, V. Barasuol, V. S. Medeiros, M. Becker, J. Peters, and O. Arenz. Floating-base deep Lagrangian networks. In *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, 2026.

Cesar A. Ipanaque Zapata and Jesús González. Higher topological complexity of a map. *Turkish Journal of Mathematics*, 47(6):1616–1633, 2023. https://arxiv.org/abs/2212.03441.

Junjin Xiao, Dongyang Li, Yandan Yang, Shuang Zeng, Tong Lin, Xinyuan Chang, Feng Xiong, Mu Xu, Xing Wei, Zhiheng Ma, Qing Zhang, and Wei-Shi Zheng. Learning action manifold with multi-view latent priors for robotic manipulation. *arXiv preprint arXiv:2605.11832*, 2026. https://arxiv.org/abs/2605.11832.

Open X-Embodiment Collaboration et al. Open X-Embodiment: Robotic learning datasets and RT-X models. In *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, 2024. https://arxiv.org/abs/2310.08864.

Marcell Szikszai, Marcin Magnus, Siddhant Sanghi, Sachin Kadyan, Nazim Bouatta, and Elena Rivas. RNA3DB: A structurally-dissimilar dataset split for training and benchmarking deep learning models for RNA structure prediction. *Journal of Molecular Biology*, 436(12):168552, 2024. https://doi.org/10.1016/j.jmb.2024.168552.

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Ron Levie, Michael M. Bäuerle, S. M. Ali Hassani, David Koh, Andrew W. Senior, Stephen Cavanagh, Thomas Keck, and John M. Jumper. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, 2024. https://doi.org/10.1038/s41586-024-07487-w.

Joey Bose, Guillaume Huguet, Stanislaw Jastrzebski, Prakash Patel, Nikolay Malkin, Yoshua Bengio, and Guy Wolf. SE(3)-stochastic flow matching for protein backbone generation. In *International Conference on Learning Representations (ICLR)*, 2024. https://openreview.net/forum?id=kJFIH23hXb.

Yogesh Verma, Markus Heinonen, and Vikas Garg. Let physics guide your protein flows: Topology-aware unfolding and generation. *arXiv preprint arXiv:2509.25379*, 2025. https://arxiv.org/abs/2509.25379.

A A ROBOT EXAMPLE SEPARATING BRANCHES, CHARTS, AND SINGULARITIES

Consider the obstacle-free planar 2R arm with $\ell_1 > \ell_2 > 0$, periodic joints, and position map $F(u_1, u_2) = u_1(\ell_1 + \ell_2 u_2)$ on $\mathbb{T}^2$. Its regular workspace is the annulus $\ell_1 - \ell_2 < |z| < \ell_1 + \ell_2$. Put

$$c(z) = \frac{|z|^2 - \ell_1^2 - \ell_2^2}{2\ell_1\ell_2}, \quad u_2(z) = c(z) + i\sqrt{1 - c(z)^2}, \quad u_1(z) = \frac{z}{\ell_1 + \ell_2 u_2(z)}. \tag{18}$$

These expressions are continuous, have unit modulus, and satisfy $F(u_1(z), u_2(z)) = z$. Therefore $B_0(p) = 1$. Since the annulus is homotopy equivalent to S^1, Eq. (3) gives $M_0(p) = 2$. These are distinct from its two IK solutions and from $\mathrm{TC}(\mathbb{T}^2) = 3$ for the unrestricted joint torus. The other elbow branch changes the square-root sign. Principal real-valued angles introduce cuts absent from the S^1-valued section. Radial and short-angle interpolation on the annulus gives $\rho = 2(\ell_1 - \ell_2)$, so $M_\varepsilon(p) = 2$ for $\varepsilon < \rho$, while $B_\varepsilon(p) = 1$. The same global section becomes ill-conditioned near the annulus boundary. In ordinary local joint coordinates, $\det \boldsymbol{J} = \ell_1\ell_2 \sin q_2$, $\|\boldsymbol{J}^{-1}\| \asymp |q_2|^{-1} \asymp d_{\mathrm{out}}^{-1/2}$ $(q_2 \to 0)$, where $d_{\mathrm{out}} = \ell_1 + \ell_2 - |F(q)|$. On patches at least distance d from this boundary and a fixed distance from the inner boundary, controlled coordinates give $\kappa_i \le C d^{-1/2}$. With balanced labels and fixed positive noise, Eq. (11) then gives the sufficient envelope

$$N = \widetilde{O}\big(\boldsymbol{J}(d)\, d^{-1}\varepsilon^{-4}\big)\,. \tag{19}$$

The coordinate-patch count $\boldsymbol{J}(d)$ and chart assumptions remain part of the calculation; $B_0 = 1$ does not set them. This is a nonparametric upper budget, not a lower bound for the known analytic arm. Equation (18) already solves that inverse exactly. The example shows how one map can have small topological counts and large local statistical constants.

B MAP COMPLEXITY: DEFINITIONS, BOUNDS, AND COMPUTATIONS

B.1 THE PLANNING PROBLEM FIXES THE INVARIANT

The path space has the compact-open topology. For manifolds, a continuous map into this path space is equivalent to a jointly continuous family of paths. The prescribed-start condition is exact: $\Gamma(q, y)(0) = q$. Approximation is permitted only in its task endpoint. All paths remain in E; imposing speed, torque, length, or collision constraints beyond membership in E would define a different admissible class.

The composite $\pi_p = (\mathrm{id}_E \times p) \circ (\mathrm{ev}_0, \mathrm{ev}_1)$ is a Hurewicz fibration. It is surjective because lifting a path in the connected base from $p(q)$ to any y supplies a feasible path starting at q. This reasoning also handles disconnected E: each component maps onto W, and a lifted path stays in its starting component. Thus the sectional category of π_p counts the declared continuous rules. For singular maps, exact sections, homotopy sections, and different classes of domains need not give the same answer; we make no identification outside the regular fibration assumptions.

There are several uses of "topological complexity of a map." The pullback invariant of Scott (2022), equivalent to the work-map formulation of Murillo and Wu (2021), counts domains in $E \times E$ on which one chooses a task-space path from $p(q_0)$ to $p(q_1)$. Our $M_0(p)$ instead takes $(q, y) \in E \times W$ and returns a path in E with its start fixed. For a fibration, it can equivalently count task paths from $p(q)$ to y, because such a family lifts continuously with the prescribed start. This is the mixed planning formulation. The input spaces and endpoint requirements explain the difference between the notions; changing a normalization by one does not remove that difference.

B.2 PROOF OF PROPOSITION 1

On a categorical open set $U \subset W$, a homotopy from a constant map to its inclusion lifts from a constant point of the corresponding fiber. Its endpoint is a section, proving $B_0(p) \le \mathrm{cat}(W)$. Conversely, fix $q_0 \in E$ and restrict a manipulation cover to $\{q_0\} \times W$. Projecting each path by p contracts the corresponding target domain to $p(q_0)$. Hence $\mathrm{cat}(W) \le M_0(p)$. To prove $M_0(p) \le \mathrm{TC}(W)$, pull a motion-planning cover of $W \times W$ back along $(q, y) \mapsto (p(q), y)$ and lift its task paths starting at q.

If p admits a section s, restricting any manipulation cover along $(y_0, y_1) \mapsto (s(y_0), y_1)$ and projecting the resulting paths gives a motion-planning cover of $W \times W$. This proves $\mathrm{TC}(W) \le M_0(p)$ and hence equality.

For the cohomological bounds, a class in the kernel of the pullback of a fibration restricts to zero on every domain admitting a section. If m such domains cover the base, the product of m kernel classes is zero, by the relative cup product. Therefore a nonzero product of length ℓ forces at least $\ell + 1$ domains. Apply this first to p. For π_p, constant paths give a homotopy equivalence $c : E \to E^{[0,1]}$ and $\pi_p \circ c = \Delta_p$. Hence $\ker \pi_p^* = \ker \Delta_p^*$, proving the second inequality. Manifold bases and their open covers suffice for this singular-cohomology argument; no assertion about arbitrary non-ANR bases is needed.

B.3 Computing Table 1

For $p_k(z) = z^k$, a global section would give $k \deg s = 1$, which is impossible for $k \ge 2$. Two overlapping slit circles give local roots, so $B_0 = 2$. Since $\mathrm{cat}(S^1) = \mathrm{TC}(S^1) = 2$, Eq. (3) gives $M_0 = 2$ as well. In local angular coordinates the inverse derivative is $1/k$, while the forward Lipschitz constant is k. The integers B_0 and M_0 therefore stay fixed when k changes, although the sharp risk frontier changes through $L_p K = kK$.

For product squaring on $\mathbb{T}^r$, the r degree-one mod-two classes pull back to zero, and their product is nonzero. Thus $B_0 \ge r + 1$. Since $\mathrm{cat}(\mathbb{T}^r) = \mathrm{TC}(\mathbb{T}^r) = r + 1$, both B_0 and M_0 equal $r + 1$. Every local inverse derivative is $I/2$ in product angular coordinates. The fiber has 2^r elements, so fiber cardinality, topological domain count, and local sensitivity have visibly different dependence on r.

For the quaternion cover $p : S^3 \to \mathbb{RP}^3 \cong \mathrm{SO}(3)$, write $H^*(\mathbb{RP}^3; \mathbb{Z}_2) = \mathbb{Z}_2[a]/(a^4)$. The class a pulls back to zero while $a^3 \neq 0$, proving $B_0 \ge 4$. The category is four; the group structure gives $\mathrm{TC}(\mathrm{SO}(3)) = \mathrm{cat}(\mathrm{SO}(3)) = 4$. Hence $B_0 = M_0 = 4$. With the unit-sphere metric and the quotient metric $d([u], [v]) = \min\{d_{S^3}(u, v), d_{S^3}(u, -v)\}$, this is a local isometry and $\kappa = 1$. If task rotation distance is instead the usual physical rotation angle, it is twice this quotient distance and the inverse derivative norm is $1/2$. These counts concern quaternion lifting, not every robot whose workspace contains orientations.

For the regular 2R position map, the two elbow sheets each map diffeomorphically to the annulus. Equation (18) gives $B_0 = 1$. The annulus retracts onto S^1, so $\mathrm{cat}(W) = \mathrm{TC}(W) = 2$ and $M_0 = 2$. One can construct the two rules explicitly. In polar task coordinates interpolate radii linearly, and interpolate angles using either the branch $(-\pi, \pi)$ or $(0, 2\pi)$ for the target-minus-start angle. Their domains exclude antipodal and coincident directions, respectively, and cover every pair. Lift the interpolated task path through the elbow sheet containing the start. Its initial configuration is exactly q, and its endpoint reaches the target. For disconnected E, the positive- and negative-elbow rules with the same angular branch combine into one continuous rule on a disconnected domain. For the error threshold, write $a = \ell_1 - \ell_2$. Two annulus points at Euclidean distance less than $2a$ cannot have antipodal directions, since antipodal points are separated by the sum of their radii, which exceeds $2a$. Linear interpolation of radii and short-arc interpolation of angles therefore define a continuous path interpolation for $\rho = 2a$. All radii stay in the annulus. Theorem 2 gives the claimed finite-error manipulation count; the exact global section gives $B_\varepsilon = 1$ for every nonnegative tolerance.

B.4 Approximate manipulation and its risk bound

Let $\Gamma : V \to E^{[0,1]}$ be an approximate rule and write $a(q, y) = \Gamma(q, y)(1)$. Set $H(q, y, t) = h(p(a(q, y)), y, t)$, using the specified interpolation on W. Lift H starting at a. Concatenate Γ on $[0, 1/2]$ with the lifted correction on $[1/2, 1]$, with linear time rescaling on each half. The two pieces agree at the midpoint, depend continuously on (q, y), retain the initial point q, and end in $p^{-1}(y)$. The domain is unchanged, proving $M_\varepsilon = M_0$ for $\varepsilon < \rho$.

For a global rule, let $Z = E \times W$ with metric $d_Z((q, y), (q', y')) = (d_E(q, q')^2 + d_W(y, y')^2)^{1/2}$, and suppose its endpoint map a is K-Lipschitz. Then $e(q, y) = d_W(p(a(q, y)), y)$ is $(1 + L_p K)$-Lipschitz. If Z is compact and $M_0 > 1$, the preceding correction argument forces $\max_Z e \ge \rho$. Under the lower mass condition $\nu(B_Z(z, t)) \ge a t^d$ with $d = n + r$, integration over a radius-b ball

about a maximizer gives

$$R_\nu(\Gamma) \geq \frac{a\rho^2}{4} b^{n+r}, \qquad \nu\{e \geq \rho/2\} \geq a b^{n+r}, \qquad b = \min\{t_0, \rho/[2(1 + L_p K)]\}. \tag{20}$$

Consequently $V \geq (ab^{n+r})^{-1} \log(1/\delta)$ independent input pairs suffice to detect such a residual with probability $1 - \delta$. Only the endpoint sensitivity enters this bound, although path validity is needed to invoke M_0. Predicting unconstrained endpoints as a function of (q, y) does not satisfy that requirement.

C Proof details and scope of the representation results

C.1 Lifting approximate sections

Let $C_\rho = \{(u, v) : d_W(u, v) < \rho\}$ and let $h : C_\rho \times [0, 1] \to W$ be continuous with $h(u, v, 0) = u$ and $h(u, v, 1) = v$. For a module $g : U \to E$ with residual bounded by $\varepsilon < \rho$, define $H(y, t) = h(p(g(y)), y, t)$. Then $H(y, 0) = p(g(y))$. The homotopy lifting property supplies $\widetilde{H} : U \times [0, 1] \to E$ with $p\widetilde{H} = H$ and $\widetilde{H}(y, 0) = g(y)$. The map $s(y) = \widetilde{H}(y, 1)$ is a continuous section. This argument requires the fibration over the whole interpolation region, not just the existence of an inverse derivative at each training point.

For a compact smooth Riemannian manifold without boundary, completeness and positive injectivity radius provide the specified interpolation for pairs closer than that radius. In a punctured or truncated workspace, a short ambient line or geodesic can leave the valid domain. The main theorem is applied only when interpolation is certified within the base. Its conclusion is about existence of continuous modules. It does not assert that the lifted module lies in a fixed finite neural architecture, has controlled weights, or can be computed efficiently.

C.2 General risk bound

Write $e(y) = d_W(p(g(y)), y)$. For $y, y' \in W$, the reverse triangle inequality gives

$$|e(y) - e(y')| \leq d_W(p(g(y)), p(g(y'))) + d_W(y, y') \tag{21}$$

$$\leq (L_p K + 1) d_W(y, y'). \tag{22}$$

If $\sup e < \rho$, the lifting argument would yield a global section. Therefore $\sup e \geq \rho$, and compactness attains the supremum at a point y_0. On $B_W(y_0, b)$, $e \geq \rho - Hb \geq \rho/2$. The lower mass condition and integration yield Eq. (5). The probability that V independent targets miss a set of mass at least ab^r is at most $\exp(-Vab^r)$, proving Eq. (6). A trained random predictor can be conditioned on its training data, so the validation argument remains valid if the validation targets are independent and K is a valid bound for that fitted predictor. Reusing training data in this calculation would be unjustified.

C.3 Fill distance and a uniform certificate

For any y, choose a nearest sampled target y_j. The residual Lipschitz bound gives $e_g(y) \leq e_g(y_j) + (1 + L_p K) d_W(y, y_j)$. Taking the supremum proves the first inequality in Eq. (7); combining it with $\sup e_g \geq \rho$ proves the second. The resulting lower bound on K is informative when its right-hand side is positive. The same argument holds for valid planners on $E \times W$, replacing B_0 by M_0 and using the endpoint Lipschitz constant.

A random test set can also give a fill-distance certificate. Let $\mathcal{N}(W, t)$ be a covering number and suppose $\mu(B(y, t)) \geq at^r$ for $t \leq t_0$. For $h/2 \leq t_0$, a deterministic $h/2$-net and a union bound give

$$\mathbb{P}\{h_S > h\} \leq \mathcal{N}(W, h/2) \exp\{-Va(h/2)^r\}. \tag{23}$$

If each net ball receives one test point, every target lies within h of some test point. Thus $V \geq [a(h/2)^r]^{-1} \log(\mathcal{N}(W, h/2)/\delta)$ ensures $\|e_g\|_\infty \leq e_S + (1 + L_p K)h$ with probability at least $1 - \delta$. This statement needs a certified K and a quantitative covering bound. Sample count alone does not provide either. Unlike a spacing argument that assumes negligible training residual, Eq. (7) includes the observed residual explicitly.

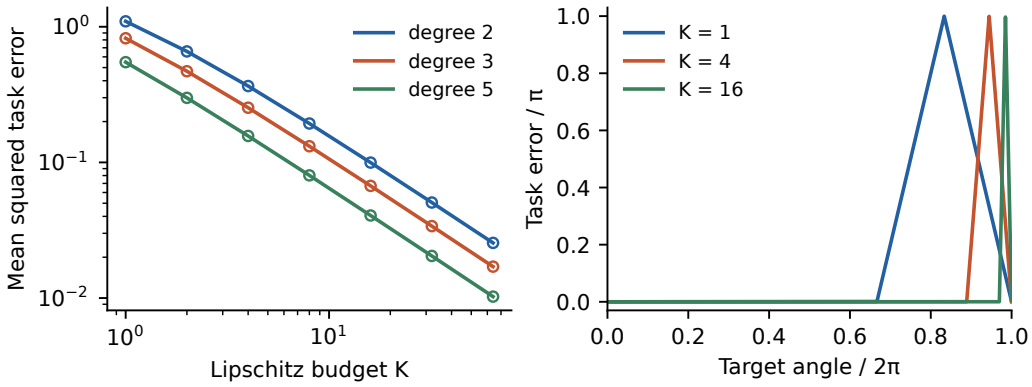

Figure 3: Exact periodic inverse tradeoff. Left: the sharp theoretical mean squared error (lines) and quadrature (markers). Right: narrower residual peaks preserve maximum error π. Distances are angular and squared errors use radians squared.

C.4 Circle lower bound, sharpness, and sampling

For any continuous $g : S^1 \to S^1$, let $h(z) = p_k(g(z))z^{-1}$. Its degree is $k \deg(g) - 1 \neq 0$, so h is surjective and takes the value -1. Thus e_g attains π. If $\mathrm{Lip}(g) \leq K$, let $H = 1 + kK$ and center the angular coordinate t at a maximizer. The two arcs of length π/H are disjoint except at their center, and

$$R_\mu(g) \geq \frac{1}{2\pi} 2 \int_0^{\pi/H} (\pi - Ht)^2 dt = \frac{\pi^2}{3H}. \tag{24}$$

For the attaining construction, put $w = 2\pi/(1 + kK)$ and define $g(e^{i\theta}) = e^{i\phi(\theta)}$ by

$$\phi(\theta) = \begin{cases} \theta/k, & 0 \leq \theta \leq 2\pi - w, \\ (2\pi - w)(2\pi - \theta)/(kw), & 2\pi - w \leq \theta \leq 2\pi. \end{cases} \tag{25}$$

For the upper construction, $K \geq 1/k$ ensures $w \leq \pi$ and bounds both slopes of ϕ in Eq. (25) by K. Integrating the derivative along any shortest input arc bounds output geodesic distance, establishing the global Lipschitz condition including the identified endpoint. Outside the seam, $k\phi(\theta) - \theta = 0$. At $\theta = 2\pi - w + u$, $0 \leq u \leq w$, it equals $-2\pi u/w$. Hence

$$e_g(\theta) = \min\{2\pi u/w,\, 2\pi(1 - u/w)\}, \qquad R_\mu(g) = \frac{w\pi}{6}. \tag{26}$$

The set where $e_g > \pi/2$ occupies the central half of the seam and has measure $w/(4\pi) = 1/(2H)$. For V uniform independent targets, the exact detection probability is $1 - (1 - 1/(2H))^V$. A deterministic angular validation grid with spacing Δ also obeys

$$\max_{\text{grid}} e_g \geq \pi - H\Delta/2, \tag{27}$$

since a grid point is at distance at most $\Delta/2$ from a peak. Thus an unresolved grid maximum must not be substituted for the continuous supremum.

The value π depends on the metric normalization. Under unit-circle chordal distance the worst residual is two; the corresponding sharp integrated expression has a different constant. All circle numbers in this paper use angular distance.

C.5 What module count means for an implementation

A continuous exact local module is a section on its declared open domain. The cover can reuse parameters, and its index may be chosen discontinuously. Sectional category does not count network layers or independent experts. A model returning a distribution, an unordered solution set, or an invalid ambient vector is a different representation class. Normalizing a vector with a possible zero, or taking an angle with a cut, does not establish a globally continuous map into the required manifold. A continuous six-dimensional orientation encoding changes the quaternion-lift problem; it does not contradict the genus of $S^3 \to \mathrm{SO}(3)$.

D PROOF OF THE STATISTICAL SEPARATION

For the product squaring cover $p_0 : \mathbb{T}^r \to \mathbb{T}^r$, let $a_1, \dots, a_r$ be the standard degree-one mod-two cohomology generators. Then $p_0^* a_j = 2a_j = 0$, while $a_1 \smile \cdots \smile a_r \neq 0$. Schwarz's kernel cup-length lower bound gives $\operatorname{secat}(p_0) \geq r + 1$. Conversely a categorical cover of $\mathbb{T}^r$ has $r + 1$ sets. Each inclusion on such a set is null-homotopic, so a constant map into a fiber can be lifted along a homotopy to obtain a section. Hence $\operatorname{secat}(p_0) \leq r + 1$. Postcomposing with translation of the first target circle is a homeomorphism and does not change this number.

For observed configurations $z_j \in \mathbb{T}^r$, let the first complex response be $Y_{j1} = e^{i\alpha} z_{j1}^2 + \xi_{j1}$, with real and imaginary noise coordinates independently $N(0, \sigma^2)$. Other response coordinates can have the same noise law and are not needed because their mapping is known. Put

$$Z = \frac{1}{N} \sum_{j=1}^{N} Y_{j1} \overline{z_{j1}^2}, \qquad \widehat{\alpha} = \arg Z. \tag{28}$$

Conditional on all configurations, $Z - e^{i\alpha}$ is a two-dimensional isotropic Gaussian with covariance $\sigma^2 \boldsymbol{I}/N$. Thus

$$\mathbb{P}(|Z - e^{i\alpha}| > t) = \exp(-Nt^2/(2\sigma^2)). \tag{29}$$

For $t \leq 1/2$, elementary planar geometry gives $d_{S^1}(e^{i\widehat{\alpha}}, e^{i\alpha}) \leq 2t$. Taking $t = \varepsilon/2$ yields the asserted sufficient sample count (with $N \geq 1$). If $\sigma = 0$, one pair recovers the phase exactly. No distributional coverage assumption on the input configurations is needed for this sufficient statistic.

Let $\widehat{s}_i$ be any sectional cover for $p_{\widehat{\alpha}}$. At every point of its domain,

$$p_\alpha(\widehat{s}_i(y)) = (e^{i(\alpha - \widehat{\alpha})} y_1, y_2, \dots, y_r). \tag{30}$$

The product angular distance to y is exactly the phase error, independent of r. There is no need to learn each local section independently. At fixed confidence, a two-point Gaussian testing argument for phases separated by a constant multiple of ε also requires $\Omega(\sigma^2 \varepsilon^{-2})$ samples when $\sigma > 0$ and ε is small. The result is a counterexample to a universal factor $B_0(p)$ in sample count, not a guarantee that all robot geometries admit such efficient sharing.

The same estimate supports manipulation. Choose an exact manipulation cover for $p_{\widehat{\alpha}}$ with $r + 1$ domains, whose existence follows from $\operatorname{cat}(\mathbb{T}^r) = \operatorname{TC}(\mathbb{T}^r) = r + 1$. Each rule starts at the specified configuration and ends at some a satisfying $p_{\widehat{\alpha}}(a) = y$. Under the true map its endpoint has residual exactly $d_{S^1}(e^{i\alpha}, e^{i\widehat{\alpha}})$. Paths remain in the common configuration torus. Hence the sample count in Eq. (9) is sufficient for an approximate manipulation cover as well. The construction assumes the known family is available for synthesizing these covers; it does not assert efficient atlas discovery from unstructured observations.

E LOCAL STATISTICAL BOUNDS AND INVERSE CONDITIONING

E.1 LOCAL REGRESSION AND THE DIMENSION THAT MATTERS

Let J denote the number of *observed statistical regimes*, not sectional category. Given $I = i$, suppose the input $X \in [0, 1]^s$ has density bounded above and below, and the response is $f_i(X) + \xi \in \mathbb{R}^m$, where $\|f_i\|_\infty \leq 1$, $\operatorname{Lip}(f_i) \leq L$, and coordinates of ξ are conditionally independent, mean-zero σ-sub-Gaussian. Assume $\mathbb{P}(I = i) \geq p_{\min} > 0$.

Proposition 6 (A local statistical envelope)**.** *For $0 < \varepsilon < 1$, a histogram estimator achieves maximum-over-regimes uniform Euclidean error ε with probability at least $1 - \delta$ when*

$$N \geq \frac{C_s}{p_{\min}} (1 + L/\varepsilon)^s (1 + m\sigma^2/\varepsilon^2) \log \frac{C_s m J (1 + L/\varepsilon)^s}{\delta}. \tag{31}$$

For J independent scalar Lipschitz targets, uniform regime probabilities and Gaussian noise, the minimax expected integrated *squared risk obeys, in the small-error regime with fixed positive L, σ,*

$$\inf_{\widehat{f}} \sup_{(f_1, \dots, f_J)} \mathbb{E} \frac{1}{J} \sum_i \|\widehat{f}_i - f_i\|_{L^2}^2 \geq c_s L^{2s/(s+2)} (\sigma^2 J/N)^{2/(s+2)}. \tag{32}$$

Table 4: Different learning tasks induce different statistical experiments. The nonparametric exponents assume fixed nonzero response noise and a full Lipschitz target class.

Task	Input dimension	Accuracy scaling / additional requirement
Forward regression $q \mapsto F(q)$	n	$\widetilde{O}(\varepsilon^{-(n+2)})$
Designated section $y \mapsto s(y)$	r	$\widetilde{O}(\varepsilon^{-(r+2)})$; section labels and conditioning
Inverse with free fiber coordinate (y, z)	$r + \dim z$	Depends on the chosen conditional parameterization
Structured planar calibration	n joint inputs	$O(n\sigma^2\varepsilon^{-2})$ under a well-conditioned design

These are standard nonparametric mechanisms (Kpotufe, 2011; Tsybakov, 2009); The proofs below give an explicit bin-count and Gaussian-noise argument and an Assouad construction. The upper and lower bounds use different risk criteria, stated explicitly; their accuracy exponents agree up to logarithms. The upper estimator is not continuous and is used only to establish statistical sufficiency. Continuous architecture constraints require a separate approximation analysis.

For fixed noise and balanced independent regimes, Eq. (31) has order $\widetilde{O}(J\varepsilon^{-(s+2)})$. This factor J comes from independent unknown functions and sampling allocation. It can be equated with $B_0(p)$ only after exhibiting a realizable kinematic family with those independent degrees of freedom and suitable labels. An arbitrary smooth robot with a fixed geometry does not supply that construction.

E.2 Uniform histogram upper bound

Take a regular grid on $[0, 1]^s$ with side $h = 1/M_h$ where

$$M_h = \max\{1, \lceil 4L\sqrt{s}/\varepsilon \rceil\}. \tag{33}$$

There are M_h^s cells. In a regime-cell pair, average the response vectors. The deterministic difference between any $f_i(x)$ in the cell and the average of its sampled noiseless values is at most $L\sqrt{s}h \le \varepsilon/4$. If $L\sqrt{s} \le \varepsilon/4$, the one-cell choice has the same property.

Let the conditional density lower bound be $c_0 > 0$. Every regime-cell has probability at least $p_{\min}c_0h^s$. A multiplicative Chernoff bound and a union bound over JM_h^s pairs imply that all counts are at least $Np_{\min}c_0h^s/2$ with probability $1 - \delta/2$, provided

$$Np_{\min}c_0h^s \ge C \log(2JM_h^s/\delta). \tag{34}$$

Conditional on inputs and labels, the mean of a coordinate's noise is $\sigma/\sqrt{N_{ib}}$-sub-Gaussian. A second union bound over all coordinates, regimes, and cells bounds every coordinate mean by

$$\sigma\sqrt{2\log(4mJM_h^s/\delta)/N_{ib}}. \tag{35}$$

Multiplying by $\sqrt{m}$ bounds Euclidean noise error. The count requirement in Eq. (31), with a sufficiently large C_s absorbing c_0 and dimension factors, makes the noise term at most $\varepsilon/2$. Combining the two events proves the statement. The occupancy term remains even at zero noise. In particular, the noisy exponent $s + 2$ is not a universal noiseless rate.

The bound is an explicit estimator argument. A bare Hoeffding uniform-convergence bound on squared losses at risk tolerance ε^2 would instead introduce ε^{-4} and would not by itself establish the claimed noisy regression rate. Faster least-squares arguments require their own variance assumptions.

E.3 Independent-regime lower bound

For clarity let all regime inputs be uniform on the cube and outputs be scalar with $N(0, \sigma^2)$ noise. In each regime place ch^{-s} disjoint bumps of radius proportional to h, amplitude $c'Lh$, and uniformly bounded Lipschitz constant. Independently choose the sign of each bump. For sufficiently small h, all functions lie in the prescribed bounded Lipschitz class.

Two parameter vectors differing on one bump have integrated squared difference of order L^2h^{s+2} within that regime. A single observation falls in that regime with probability $1/J$, so the Gaussian

KL divergence of their N-sample distributions is at most

$$C\frac{NL^2h^{s+2}}{J\sigma^2}. \tag{36}$$

Choose $h = c(\sigma^2 J/(NL^2))^{1/(s+2)}$ to bound this divergence by a fixed small constant. Assouad's lemma lower-bounds the probability of estimating each sign incorrectly. Summing the squared separations over Jh^{-s} signs, including the $1/J$ risk weight, yields cL^2h^2 and therefore Eq. (32). This constructs an abstract regression family with independent targets; it does not establish that every bump combination is realizable as the inverse of a fixed rigid serial-chain model.

E.4 Proof of the chart-dependent budget

The atlas assumptions separate a topological cover from its statistical implementation. Let U_i be open section domains covering the workspace under study, with continuous sections s_i. Each is represented by a normalized input chart on the portion used for regression, $\chi_i : [0,1]^r \to W$, and a coordinate function $f_i : [0,1]^r \to C_i$ such that

$$p(\psi_i(f_i(x))) = \chi_i(x), \qquad \psi_i(f_i(x)) = s_i(\chi_i(x)). \tag{37}$$

The chart images cover all evaluation targets, and the corresponding section domains cover W. It is enough, for example, to use finitely many compact coordinate patches with continuous extensions to surrounding open section domains. These conditions are a supplied representation assumption; on a general noncompact workspace such a finite representation need not exist. If only a compact subset is covered, the guarantee and topological lower count apply to that declared restricted problem, not to an unobserved larger workspace.

All coordinate changes, target branches, and a rule for choosing a valid chart at inference are supplied. Training labels I have probabilities π_i; conditional on $I = i$, the normalized input density is at least a common $c_0 > 0$. Labels can represent a sampling mechanism on overlapping domains and need not form a disjoint geometric partition. Outputs are $f_i(X) + \xi$, with independent observations and conditionally mean-zero, coordinatewise σ_i-sub-Gaussian response noise. The functions have bounded range, $\mathrm{Lip}(f_i) \le \kappa_i$, and the known decoder ψ_i maps the entire closed convex set C_i into E with Lipschitz constant at most D_i. This last condition prevents noisy predictions from leaving the feasible domain.

Set $\eta_i = \varepsilon/(L_p D_i)$ and use a histogram of side $1/M_i$ with

$$M_i = \max\{1, \lceil 4\sqrt{r}\,\kappa_i/\eta_i \rceil\}. \tag{38}$$

The number of cells is at most $C_r v_i$. Every chart-cell has probability at least $\pi_i c_0 M_i^{-r}$. A union bound over all $\sum_i M_i^r$ cells gives positive counts and the lower bound $N\pi_i c_0 M_i^{-r}/2$ once the occupancy requirement is met. Conditional noise concentration, followed by a union bound over cells and m coordinates, bounds Euclidean noise error by $\eta_i/2$ when each such count is at least a constant times $m\sigma_i^2\eta_i^{-2}\log(C_r m\sum_i v_i/\delta)$. The bias is at most $\eta_i/4$ by construction. These requirements are jointly implied by Eq. (11).

Project each cell average to C_i. Orthogonal projection onto a closed convex set cannot increase distance to the true $f_i(x) \in C_i$, so the coordinate guarantee survives. Decoding gives joint error at most $D_i\eta_i$ and then task error at most $L_p D_i \eta_i = \varepsilon$. The uniform guarantee permits any supplied valid selector on overlaps, even a discontinuous one. It does not make a histogram continuous inside its section domain.

Finally, for every probability vector π,

$$\max_i w_i/\pi_i \;\ge\; \sum_i \pi_i(w_i/\pi_i) = \sum_i w_i. \tag{39}$$

Choosing $\pi_i = w_i/\sum_j w_j$ attains equality. This optimizes the displayed sufficient budget over label sampling, with the conditional within-chart designs held fixed. It does not prove minimax optimal experimental design or include the cost of learning unknown chart boundaries, sensitivities, labels, or shared structure.

For the 2R calculation in Eq. (19), work away from the inner singular boundary and truncate the outer boundary by $d > 0$. Use patches with controlled normalized coordinate distortion, uniformly

bounded D_i and L_p, and $\kappa_i \le Cd^{-1/2}$. For fixed noise, $r = 2$, and small d, ε, Eq. (10) gives $w_i = O(d^{-1}\varepsilon^{-4})$. Balanced labels give $\max_i w_i/\pi_i = O(J(d)d^{-1}\varepsilon^{-4})$. The displayed $\widetilde{O}$ hides the explicit logarithm in Eq. (11). Neither the coordinate-patch count nor the density constants are determined by $B_0 = 1$. If the analytic kinematic parameters are known, this nonparametric envelope is unnecessary; the exact section can simply be evaluated.

E.5 Coordinate and loss conversion

If an output-chart decoder ψ_i is D_i-Lipschitz, coordinate error η gives joint-space error at most $D_i\eta$. If F is L_F-Lipschitz on the decoded region, it gives task error at most $L_F D_i \eta$. These are sufficient upper transfers; their converses need local injectivity and conditioning. In redundant systems, points far apart along a fiber can have identical task values. Also, a smooth manifold's small-scale covering exponent equals its dimension under controlled charts, but its global doubling constant need not equal that dimension exactly. Volume, metric scale, density, and chart distortion belong in the constants.

E.6 Right inverses of a Jacobian

Let $\boldsymbol{J} : T_qQ \to T_yY$ be surjective and $\boldsymbol{A} : T_yY \to T_qQ$ satisfy $\boldsymbol{JA} = \boldsymbol{I}$. For a unit left singular vector $\boldsymbol{u}_r$ associated to the smallest singular value σ_r, every solution of $\boldsymbol{Jv} = \boldsymbol{u}_r$ has norm at least $1/\sigma_r$. Hence $\|\boldsymbol{A}\| \ge 1/\sigma_r$. For a section, $\boldsymbol{A} = ds_y$. If $n = r$, equality holds pointwise because $\boldsymbol{A} = \boldsymbol{J}^{-1}$. If $n > r$, $\boldsymbol{A} = \boldsymbol{J}^\dagger + \boldsymbol{N}$ with $\boldsymbol{JN} = 0$, and the null-space term is not controlled by σ_r. Moreover, prescribing $ds = \boldsymbol{J}^\dagger$ pointwise need not integrate to a section. We make no such integrability assumption.

F Structured calibration and the planar example

F.1 Exact conditional risk for link-length estimation

Let $\theta_j = \sum_{a \le j} q_a$. Under independent uniform joint angles, the triangular map $\boldsymbol{q} \mapsto \boldsymbol{\theta}$ is an automorphism of the torus, so $\theta_1, \dots, \theta_n$ are themselves independent uniform angles. Equivalently one can compute the cross terms directly. For any real coefficient error $\boldsymbol{v}$,

$$\mathbb{E}_Q \left| \sum_j v_j e^{i\theta_j} \right|^2 = \sum_j v_j^2. \tag{40}$$

Conditional on a full-column-rank design $\boldsymbol{A}$, least squares has error $\widehat{\boldsymbol{\ell}} - \boldsymbol{\ell} = (\boldsymbol{A}^\top \boldsymbol{A})^{-1}\boldsymbol{A}^\top \boldsymbol{\xi}$, a Gaussian vector with covariance $\sigma^2(\boldsymbol{A}^\top \boldsymbol{A})^{-1}$. Taking the trace proves Eq. (14). On the event $\boldsymbol{A}^\top \boldsymbol{A} \succeq cN\boldsymbol{I}$, Gaussian concentration also gives, with conditional probability $1 - \delta$,

$$\|\widehat{\boldsymbol{\ell}} - \boldsymbol{\ell}\|^2 \le \frac{\sigma^2}{cN}\left[n + 2\sqrt{n\log(1/\delta)} + 2\log(1/\delta)\right]. \tag{41}$$

The design condition must be checked or established for the sampling scheme. It is not guaranteed by configuration dimension or by topology. In particular, repeated unexciting configurations can make calibration unidentifiable. The link lengths are the only unknown parameters in this experiment; unknown axes, offsets, compliance, and sensor distortions define larger model classes. If the lengths and all other geometry are known, no forward-model learning is required.

For a general known feature map $\boldsymbol{\Phi}(q) \in \mathbb{R}^{m \times d_\theta}$ and linear kinematic family $F_{\boldsymbol{\theta}}(q) = \boldsymbol{\Phi}(q)\boldsymbol{\theta}$, let $\boldsymbol{G}_\nu = \mathbb{E}_{q\sim\nu}\boldsymbol{\Phi}(q)^\top \boldsymbol{\Phi}(q)$ for the test design. Stacking the training features gives $\boldsymbol{A}$. Whenever $\boldsymbol{A}$ has full column rank, the same covariance calculation gives

$$\mathbb{E}_{\boldsymbol{\xi}} \|F_{\widehat{\boldsymbol{\theta}}} - F_{\boldsymbol{\theta}}\|^2_{L^2(\nu)} = \sigma^2 \operatorname{tr}\big(\boldsymbol{G}_\nu(\boldsymbol{A}^\top \boldsymbol{A})^{-1}\big) = \frac{\sigma^2}{N} d_{\mathrm{cal}}(\boldsymbol{A}; \nu), \tag{42}$$

$$d_{\mathrm{cal}}(\boldsymbol{A}; \nu) = \operatorname{tr}\big(\boldsymbol{G}_\nu(\boldsymbol{A}^\top \boldsymbol{A}/N)^{-1}\big). \tag{43}$$

This factor measures the interaction of the unknown feature coefficients, training excitation, and test loss. It is unchanged under invertible reparameterizations of $\boldsymbol{\theta}$. It is not the usual ridge effective

dimension and need not be bounded by d_θ. For the planar arm under uniform test joints, $\boldsymbol{G}_\nu = \boldsymbol{I}_n$, yielding Eq. (15). Since every feature column has squared norm N, $\mathrm{tr}(\boldsymbol{A}^\top \boldsymbol{A}/N) = n$; the arithmetic-harmonic mean inequality then gives $d_{\mathrm{cal}} \geq n$, with equality for an orthogonal design. An unidentifiable training design requires a separate analysis; we do not replace the inverse by a pseudoinverse and silently discard test-relevant bias.

F.2 A SECTION FOR THE 2R ARM

On the open annulus, $-1 < c(z) < 1$, so $u_2(z)$ is a continuous point of the open upper semicircle. The squared modulus of $\ell_1 + \ell_2 u_2$ equals $\ell_1^2 + \ell_2^2 + 2\ell_1\ell_2 c(z) = |z|^2$. Since the denominator is nonzero, u_1 has unit modulus and the pair is an exact section. Fixing either elbow sign makes the forward restriction a diffeomorphism onto the annulus. The complete regular restriction is the disjoint union of two such sheets. This demonstrates why two preimages do not imply sectional category two.

In real angle coordinates the Jacobian is

$$\boldsymbol{J} = \begin{pmatrix} -\ell_1 \sin q_1 - \ell_2 \sin(q_1 + q_2) & -\ell_2 \sin(q_1 + q_2) \\ \ell_1 \cos q_1 + \ell_2 \cos(q_1 + q_2) & \ell_2 \cos(q_1 + q_2) \end{pmatrix}. \tag{44}$$

Its determinant is $\ell_1\ell_2 \sin q_2$. As $q_2 \to 0$, $\sigma_{\max}(\boldsymbol{J}) \to \sqrt{(\ell_1 + \ell_2)^2 + \ell_2^2} > 0$ and

$$\sigma_{\min}(\boldsymbol{J}) = \frac{|\det J|}{\sigma_{\max}(\boldsymbol{J})} \asymp |q_2|, \qquad d_{\mathrm{out}} = \frac{\ell_1\ell_2}{2(\ell_1 + \ell_2)} q_2^2 + O(q_2^4). \tag{45}$$

This proves the sensitivity exponent. Local chart constants diverge as this boundary is approached even though the sectional category on the whole regular annulus stays one.

G CLASSICAL TOPOLOGICAL FACTS AND DISTINCTIONS

Configurations and planning. For ℓ genuinely periodic independent joints and $n - \ell$ nonperiodic unconstrained joints, $Q = \mathbb{T}^\ell \times \mathbb{R}^{n-\ell} \simeq \mathbb{T}^\ell$. Its fundamental group is $\mathbb{Z}^\ell$ and $H_j(Q; \mathbb{Z}) = \mathbb{Z}^{\binom{\ell}{j}}$ for $0 \leq j \leq \ell$. Cup length gives category at least $\ell + 1$ and the product bound gives at most $\ell + 1$. A connected topological group has $\mathrm{TC} = \mathrm{cat}$ under the unreduced convention, hence $\mathrm{TC}(Q) = \ell + 1$. Cutting a periodic joint to a genuine interval removes that factor. Collision constraints or closed-chain constraints can change these conclusions. These are standard facts, not new contributions.

G.1 EILENBERG–MACLANE SPACES AND MOORE SPACES

An Eilenberg–MacLane space $K(G, j)$ has $\pi_j \cong G$ and all other positive-degree homotopy groups zero; for $j > 1$, G is abelian. A connected Moore space $M(G, j)$ instead has reduced integral homology $\widetilde{H}_j \cong G$ and zero reduced homology in other degrees, with simple connectivity required when $j > 1$ (Hatcher, 2002, Sections 1.B, 2.2, and 4.2). The two definitions control different invariants. In particular,

$$\mathbb{T}^d = K(\mathbb{Z}^d, 1), \qquad H^*(\mathbb{T}^d; \mathbb{Z}) = \Lambda_{\mathbb{Z}}(u_1, \ldots, u_d), \quad |u_i| = 1. \tag{46}$$

The universal cover is $\mathbb{R}^d$, so higher homotopy vanishes. Nevertheless $u_1 \cdots u_d \neq 0$: asphericity does not remove higher cohomology or the cup-length obstruction. For $d > 1$, the nonzero groups H_1 and H_2 already rule out a Moore description. The circle is both $K(\mathbb{Z}, 1)$ and $M(\mathbb{Z}, 1)$.

There is also a direct reason not to replace a torus by a space that retains only its first homology. The wedge $\bigvee_{i=1}^d S^1$ is a Moore space $M(\mathbb{Z}^d, 1)$, but its fundamental group is the free group F_d, not $\mathbb{Z}^d$ when $d \geq 2$. It is itself $K(F_d, 1)$. Its positive-degree cup products vanish. For $d \geq 2$ its category is two and its topological complexity is three: two distinct degree-one classes give a nonzero product of two zero-divisors, and the graph dimension gives the matching upper bound. By comparison, the torus has category and topological complexity $d + 1$. Matching a list of torsional loops or H_1 therefore does not preserve the planning calculation.

A Moore space associated with a kinematic map. Moore spaces do have a concrete role in the circle example. Attach a disk to the target circle along the degree-k map p_k. Its mapping cone is

$$C(p_k) = S^1 \cup_{p_k} D^2 = M(\mathbb{Z}/k\mathbb{Z}, 1), \qquad k \geq 2. \tag{47}$$

The cellular boundary $C_2 \to C_1$ is multiplication by k, so $\widetilde{H}_1 = \mathbb{Z}/k\mathbb{Z}$ and all other reduced integral homology is zero. This records the cokernel of the induced map on first homology. A section of p_k would require an integer right inverse to multiplication by k, which does not exist. The Moore space is thus an auxiliary record of the degree obstruction, not the configuration or target space of the mechanism. For $k = 2$ it is $\mathbb{RP}^2$, whose universal cover is S^2; it is not $K(\mathbb{Z}/2\mathbb{Z}, 1)$. That Eilenberg–MacLane space is $\mathbb{RP}^\infty$. Likewise, the physical orientation space $\mathrm{SO}(3) \cong \mathbb{RP}^3$ is neither a Moore space nor an Eilenberg–MacLane space: $H_1 = \mathbb{Z}/2\mathbb{Z}$, $H_3 = \mathbb{Z}$, and $\pi_3 = \mathbb{Z}$.

Pose and quaternion lifts. As manifolds $\mathrm{SE}(3) \cong \mathbb{R}^3 \times \mathrm{SO}(3)$ and $\mathrm{SO}(3) \cong \mathbb{RP}^3$. Therefore $\mathrm{TC}(\mathrm{SE}(3)) = \mathrm{TC}(\mathrm{SO}(3)) = 4$. For the antipodal cover $S^3 \to \mathbb{RP}^3$, the degree-one class $a \in H^1(\mathbb{RP}^3; \mathbb{Z}_2)$ pulls back to zero and $a^3 \neq 0$, giving sectional category at least four; the dimensional upper bound is four. This counts local continuous quaternion sign selections on the entire orientation space. It is not a universal inverse-robot branch count. Matrices or the continuous rotation encodings of Zhou et al. (2019) pose different representation problems.

G.2 Configuration and workspace models with obstacles

The two types arise in actual free spaces, not only as auxiliary constructions. Let P_k be a set of $k \geq 1$ distinct points. The standard complement calculation gives

$$W_{n,k} = \mathbb{R}^n \setminus P_k \simeq \bigvee_{j=1}^{k} S^{n-1}, \qquad n \geq 2. \tag{48}$$

The same homotopy type holds for the exterior of k pairwise disjoint closed round balls: within disjoint neighborhoods of the balls, radial coordinates identify this complement with the complement of their centers. These are models for a point robot, or for a translational robot whose inflated obstacles satisfy this geometry. Arbitrary obstacle shapes or unions require a separate calculation.

In two dimensions, $W_{2,k}$ is both $K(F_k, 1)$ and $M(\mathbb{Z}^k, 1)$, where F_k is the free group on k generators. The universal cover of its graph model is a tree. In three dimensions, $W_{3,k}$ is $M(\mathbb{Z}^k, 2)$ and is not an Eilenberg–MacLane space: any sphere summand is a retract and contributes a nonzero $\pi_3(S^2)$. Table 5 collects the consequences. The entries describe homotopy types; an open free workspace need not be homeomorphic to its wedge model.

Table 5: Free-space models and exact planning counts, using unreduced category. For the identity map or a redundant projection $p : W \times \mathbb{T}^s \to W$, $B_0 = 1$ and $M_0 = \mathrm{TC}(W)$ in every row. P_k denotes k removed points; F_j denotes a free group.

Space W	Homotopy type and class	$\mathrm{cat}(W)$	$\mathrm{TC}(W)$
Regular 2R annulus	S^1: $K(\mathbb{Z}, 1)$ and $M(\mathbb{Z}, 1)$	2	2
$\mathbb{R}^2 \setminus P_1$	S^1: $K(\mathbb{Z}, 1)$ and $M(\mathbb{Z}, 1)$	2	2
$\mathbb{R}^2 \setminus P_k$, $k \geq 2$	$\bigvee_k S^1$: $K(F_k, 1)$ and $M(\mathbb{Z}^k, 1)$	2	3
$\mathbb{R}^3 \setminus P_k$, $k \geq 1$	$\bigvee_k S^2$: $M(\mathbb{Z}^k, 2)$, not $K(G, j)$	2	3
$\mathbb{T}^2 \setminus P_k$, $k \geq 1$	$\bigvee_{k+1} S^1$: $K(F_{k+1}, 1)$ and $M(\mathbb{Z}^{k+1}, 1)$	2	3

Proposition 7 (Inverse and manipulation counts in obstacle models)**.** *For $k \geq 1$, $\mathrm{cat}(W_{2,k}) = \mathrm{cat}(W_{3,k}) = 2$ and*

$$\mathrm{TC}(W_{2,k}) = \begin{cases} 2, & k = 1, \\ 3, & k \geq 2, \end{cases} \qquad \mathrm{TC}(W_{3,k}) = 3. \tag{49}$$

For $p = \mathrm{id}_W$ or the projection $p : W \times \mathbb{T}^s \to W$, these are the manipulation counts $M_0(p)$, while $B_0(p) = 1$. For any surjective Hurewicz fibration over $W_{3,k}$ satisfying the standing assumptions, $B_0(p) \leq 2$ and $2 \leq M_0(p) \leq 3$; a global section forces $M_0(p) = 3$.

Proof. A noncontractible wedge of spheres is a suspension and has category two. Hence $\mathrm{TC} \leq 2\,\mathrm{cat} - 1 = 3$. The circle has $\mathrm{TC} = 2$ because it is a connected topological group. On a wedge of at least two circles, distinct rational degree-one classes a, b give a nonzero product of zero-divisors $(a \otimes 1 - 1 \otimes a)(b \otimes 1 - 1 \otimes b)$, proving $\mathrm{TC} \geq 3$. On a wedge of two-spheres, a degree-two class a supported on one summand has $a^2 = 0$ and

$$(a \otimes 1 - 1 \otimes a)^2 = -2a \otimes a \neq 0 \quad \text{over } \mathbb{Q}, \tag{50}$$

so the same lower bound applies (Farber, 2003). Identity and projection maps have global sections. Proposition 1 gives the map-specific conclusions. □

These examples separate obstacle count from continuous domain count. Adding isolated obstacles increases a Betti number but need not increase M_0. Nor does a multiply connected free space force multiple inverse modules: selecting the desired point is a global inverse of id_W, while choosing a path to it is a different problem. For a general robot map, one must retain p and its induced cohomology map; the topology of Q or W alone does not determine $B_0(p)$.

Singular spaces and singular kinematic maps. A branched network $X = \bigvee_k S^1$ can itself be a singular one-dimensional configuration space and workspace. The identity map is a fibration, so the same definitions give $B_0 = 1$ and $M_0 = 3$ for $k \geq 2$. This purely topological calculation does not require a smooth vertex neighborhood. Statistical analysis on the network still needs a metric and a sampling law, with appropriate treatment of its vertices.

For a mechanism with singular fibers, specify a collision-free configuration domain, a reachable task domain, and the regular restriction in Eq. (1) before using the fibration bounds. The planar 2R map provides one example: deleting its singular boundary values leaves the annulus in Table 5, and the actual kinematic map has $B_0 = 1$, $M_0 = 2$. Retaining rank changes requires an additional argument; a topological description of the image does not establish the homotopy-lifting property. Planning separately on regular strata also does not settle how paths pass between them.

Why the two classes are not exhaustive. Removing finitely many points from $\mathbb{T}^n$, $n \geq 3$, preserves $\pi_1 = \mathbb{Z}^n$ but changes the higher topology. The k small linking spheres obey one homology relation, their oriented sum being zero. In particular,

$$H_{n-1}(\mathbb{T}^n \setminus P_k; \mathbb{Z}) \cong \mathbb{Z}^{n+k-1}. \tag{51}$$

There is no additional independent linking homology class for a single puncture. The punctured torus is nevertheless not aspherical: its universal cover is $\mathbb{R}^n$ minus the infinite discrete preimage of the punctures. This cover is $(n-2)$-connected and has nonzero $(n-1)$-homology, hence nonzero π_{n-1} by the Hurewicz theorem. Since H_1 and H_{n-1} are both nonzero, the space is not Moore either. Thus a periodic configuration space with even one forbidden point can lie in neither class. In dimension two the behavior differs: a genus-one surface with k punctures retracts onto $k+1$ circles, as recorded in the table.

More generally, the universality theorem of Kapovich and Millson (2002, Corollary C) realizes any connected smooth compact manifold as each of finitely many diffeomorphic components of a planar-linkage moduli space. This rules out a general Eilenberg–MacLane/Moore classification of kinematic configuration spaces. The useful claim is conditional: once a particular free space is shown to have one of these homotopy types, its invariants can make the planning calculation explicit.

What transfers to learning. The resulting $B_0(p)$ constrains the number of continuous inverse domains and $M_0(p)$ constrains feasible path rules. The stability theorem transfers these counts to $\varepsilon > 0$ only when the stated metric interpolation radius exists; a homotopy classification alone supplies no such radius. The sampling budget still uses chart geometry and the dimension of the regression input. For instance, $W_{3,k}$ is three-dimensional even though its wedge model is two-dimensional. A homotopy equivalence need not preserve distances, design density, or loss, and cannot replace $r = 3$ by $r = 2$ in Eq. (11). Similarly, the known identity map needs no training data. Any statistical comparison must specify which kinematic quantities are unknown. The constrained molecular spaces in Eq. (16) require the same checks; their ambient torus does not fix their free-space topology.

H EXACT COMPLEXITIES OF TORUS OBSERVATION MAPS

The Eilenberg–MacLane description is useful because maps between tori induce integer homomorphisms between their fundamental groups. A particularly transparent kinematic baseline is the homomorphism $p_{\boldsymbol{A}}(\boldsymbol{\theta}) = \boldsymbol{A}\boldsymbol{\theta} \bmod 2\pi$. It describes integer combinations of periodic phases. It need not describe a Cartesian molecular observable.

Proposition 8 (Smith factors determine inverse domain count)**.** *Let $\boldsymbol{A} \in \mathbb{Z}^{r\times d}$ have rank $r \geq 1$, with $d \geq r$. Write its positive Smith invariant factors as $a_1 \mid \cdots \mid a_r$ and let $t = \#\{j : a_j > 1\}$. Then $p_{\boldsymbol{A}} : \mathbb{T}^d \to \mathbb{T}^r$ is a surjective smooth fibration and*

$$B_0(p_{\boldsymbol{A}}) = t + 1, \qquad M_0(p_{\boldsymbol{A}}) = r + 1. \tag{52}$$

Each fiber has $\prod_j a_j$ connected components, each a $(d-r)$-torus. A global section exists exactly when all $a_j = 1$, equivalently when $\boldsymbol{A}$ has an integer right inverse. In standard product-angular metrics, both domain counts persist for error $\varepsilon < \pi$.

Proof. Full row rank makes the derivative surjective. Compactness gives properness, and the image is an open and closed subgroup of the connected target, hence the whole target. Smith normal form gives unimodular integer matrices $\boldsymbol{U}, \boldsymbol{V}$ with $\boldsymbol{U}\boldsymbol{A}\boldsymbol{V} = (\operatorname{diag}(a_1, \ldots, a_r)\ 0)$. They induce torus diffeomorphisms, so they preserve the exact domain counts and identify the fiber as stated.

In Smith coordinates the $r - t$ unit factors have global sections and the $d - r$ unused coordinates can be fixed. The remaining covering over $\mathbb{T}^t$ has sectional category at most $\operatorname{cat}(\mathbb{T}^t) = t+1$. Pulling that cover back to $\mathbb{T}^r$ proves the upper bound. If $t > 0$, choose a prime ℓ dividing the smallest nonunit factor. Divisibility of the Smith factors makes ℓ divide all t nonunit factors. Their t degree-one target classes lie in $\ker p^*_{\boldsymbol{A}}$ over $\mathbb{F}_{\boldsymbol{\ell}}$, and their product is nonzero in the torus exterior algebra. Equation (4) gives $B_0 \geq t + 1$. For $t = 0$, an integer right inverse supplies a global section. Conversely, any continuous section induces an integer right inverse on π_1. Finally, $\operatorname{cat}(\mathbb{T}^r) = \operatorname{TC}(\mathbb{T}^r) = r + 1$, so Eq. (3) fixes M_0. The target flat torus has injectivity radius π; Theorem 2 proves the last assertion in the original metrics. □

This is an application of classical sectional-category and torus cohomology tools, not a new topological invariant. It sharpens several distinctions in a directly calculable family. Projection $(\theta_1, \ldots, \theta_d) \mapsto (\theta_1, \ldots, \theta_r)$ has $B_0 = 1$ and $M_0 = r + 1$, irrespective of the number of unobserved torsions. Product squaring has $B_0 = M_0 = r + 1$. In contrast, $\boldsymbol{A} = \operatorname{diag}(2, 3)$ has Smith factors $(1, 6)$, hence six inverse points but only $B_0 = 2$, with $M_0 = 3$. Counting the nonunit entries of an arbitrary diagonal matrix would be wrong; the divisibility-ordered Smith factors are essential.

Conditioning is additional information. The linear lift has derivative $\boldsymbol{A}$, and a locally chosen minimum-norm right inverse has derivative $\boldsymbol{A}^\dagger$ with norm $1/\sigma_r(\boldsymbol{A})$. The integer change of coordinates in the proof is generally not an isometry. For example,

$$\boldsymbol{A}_h = \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix}, \quad h \in \mathbb{Z}, \qquad B_0(p_{\boldsymbol{A}_h}) = 1, \quad M_0(p_{\boldsymbol{A}_h}) = 3, \quad \|\boldsymbol{A}_h^{-1}\| \asymp |h| \quad (|h| \to \infty). \tag{53}$$

Identical topological counts can therefore coexist with unbounded sensitivity. Smith factors cannot replace the metric quantities in Eq. (11).

I MOLECULAR KINEMATICS, LATENT TORSIONS, AND ERROR PROPAGATION

I.1 WHAT THE AMBIENT TORUS REPRESENTS

Fix a molecular graph, atom identities, connectivity, and local geometry λ. Choose a rooted tree of rigid fragments joined by rotatable bonds. After fixing a root frame, the internal-coordinate decoder $H_\lambda : \mathbb{T}^d \to \mathbb{R}^{3a}$ gives the positions of a labeled atoms. Its rotations are periodic, so it is globally defined on the torus even when principal real-valued angle coordinates have seams. Bond lengths and bond angles are held fixed in this model. If they are learned or allowed to vary, their continuous coordinates must be added; their domain is not automatically another circle factor.

For proteins, ϕ and ψ describe backbone rotations, ω describes peptide-bond torsion, and side chains supply residue-dependent χ coordinates. Fixing peptide planarity or a trans state removes a variable from that model; retaining cis/trans flexibility changes it. Termini and the chosen atom set affect the count. Ring-containing residues, disulfide closure, and excluded volume also prevent a universal formula that treats every named torsion as freely rotatable. For RNA, backbone $(\alpha, \beta, \gamma, \delta, \epsilon, \zeta)$ and glycosidic χ coordinates describe local conformation, but the ribose ring couples its coordinates. A model with varying sugar pucker must supply compatible ring geometry or solve its closure constraints. In the seven-angle RNA *suite* representation, the terminal δ angles belong to neighboring sugars and are shared between adjacent suites (Murray et al., 2003). Thus neither $\mathbb{T}^{7L}$ for L nucleotides nor a product of independent suite models can be identified with the feasible conformation space without additional assumptions. Variable-length molecules form a family of spaces, not one fixed torus.

For example, an anchored acyclic backbone containing only the N–C_α–C atoms of L residues has $d = 2(L-1)$ free torsions if every ω is fixed, or $d = 3(L-1)$ if those torsions also vary. The omitted terminal angles are not defined by this atom set. These counts describe the ideal tree before steric or loop constraints. They give $\mathrm{TC}(\mathbb{T}^d) = d + 1$ for full configuration-to-configuration planning, while an embedded full-coordinate decoder still has $B_0 = 1$. Side-chain variables can be added only with their residue-specific masks and ring constraints.

One can cut closure bonds to define the tree decoder and restore their geometry through $C_\lambda(\boldsymbol{\theta}) = 0$. Strict steric inequalities are represented by $S_\lambda(\boldsymbol{\theta}) > 0$ in Eq. (16). On a region where dC_λ has constant full row rank c, the feasible set is a smooth $(d-c)$-manifold before boundary contact. Hard contacts, rank changes, and different chemical states require separate strata. Its topology need not be that of the ambient torus. The tree decoder remains a useful smooth extension for error bounds even when interpolation through the ambient torus would violate molecular constraints. Such an interpolation is not a feasible molecular motion.

I.2 Choose the molecular task before its complexity

There are several distinct maps. Forward kinematics takes torsions to atomic coordinates or selected observables. If all geometry is specified, this map is evaluated analytically and requires no training samples. A learning problem arises when geometry, calibration parameters, a correction model, or the relevant distribution is unknown. Full-coordinate reconstruction into torus-valued angles has a global inverse whenever the anchored decoder is an embedding. In that ideal unconstrained case $B_0(H_\lambda) = 1$ and $M_0(H_\lambda) = d + 1$, since its image is homeomorphic to $\mathbb{T}^d$. A discontinuity in principal angle labels is not a discontinuity of this torus-valued inverse.

Loop closure instead prescribes an endpoint frame or anchor geometry and seeks torsions consistent with it. Its connection to robot IK is explicit in Coutsias et al. (2004). For a regular six-dimensional endpoint-pose task with d free torsions, the local fiber dimension is $d - 6$, when the differential has rank six. Closure can have multiple branches and singular values even in a small loop. Partial atom positions, distances, and experimental observables give still other maps $\mathcal{O} \circ H_\lambda$, often with nontrivial fibers. For each, B_0 concerns selecting a feasible conformation; M_0 additionally requires a feasible path from a prescribed initial conformation. Neither invariant counts low-energy states or establishes folding kinetics. A fixed closed-loop conformation space is a fiber of a closure map, not its task base.

When rigid pose is irrelevant, we fix an anchored frame or work in the quotient by $\mathrm{SE}(3)$. All-atom RMSD after rigid alignment is no larger than the anchored Cartesian error used below. Atom permutations or unresolved symmetries are separate quotients and can introduce further ambiguity. A single known gauge does not resolve a noninjective observable.

I.3 Torsional lever arms and the Cartesian loss

Equip $\mathbb{T}^d$ with the product-angular metric. For positive weights w_a summing to one, use $\|x\|_w^2 = \sum_a w_a \|x_a\|^2$ for Cartesian error; equal weights give squared anchored RMSD. Let $\mathcal{D}_k$ be the atoms downstream of torsion k, u_k its unit axis, and o_k a point on that bond axis. Choose constants ℓ_{ak} that bound $\|x_a(\boldsymbol{\theta}) - o_k(\boldsymbol{\theta})\|$ for all $\boldsymbol{\theta}$ and $a \in \mathcal{D}_k$. The intervening contour length is one such bound for a fixed-geometry tree. Put

$$b_k^2 = \sum_{a \in \mathcal{D}_k} w_a \ell_{ak}^2, \qquad \mathcal{L}^2 = \sum_{k=1}^{d} b_k^2. \tag{54}$$

Proposition 9 (A task-loss conversion for torsional decoders)**.** *For the fixed-geometry tree decoder,*

$$\frac{\partial x_a}{\partial \theta_k} = \begin{cases} u_k \times (x_a - o_k), & a \in \mathcal{D}_k, \\ 0, & a \notin \mathcal{D}_k. \end{cases} \tag{55}$$

For all torsions $\boldsymbol{\theta}, \boldsymbol{\theta}'$, writing $\Delta_k = d_{S^1}(\theta_k, \boldsymbol{\theta}'_k)$,

$$\|H_\lambda(\boldsymbol{\theta}') - H_\lambda(\boldsymbol{\theta})\|_w \le \sum_k b_k \Delta_k \le \mathcal{L}\, d_{\mathbb{T}^d}(\boldsymbol{\theta}, \boldsymbol{\theta}'). \tag{56}$$

Consequently, if $\mathcal{O}$ is $L_{\mathcal{O}}$-Lipschitz in this Cartesian metric, $F = \mathcal{O} \circ H_\lambda$ is $L_{\mathcal{O}}\mathcal{L}$-Lipschitz. This also bounds its restriction to any feasible subset with the restricted angular metric, or with the larger intrinsic path metric.

Proof. An infinitesimal rotation about a unit axis has velocity $u_k \times (x_a - o_k)$ for a downstream atom and leaves upstream atoms fixed. Its weighted column norm is bounded by b_k. Rotate the angles one at a time along their shortest circular arcs, integrate each column bound, and apply the triangle inequality. Cauchy–Schwarz yields the product-metric bound. Composing with $\mathcal{O}$ gives the final statement. The paths used in this proof need only be valid in the ambient tree model. □

The bound quantifies why equal angular errors need not have equal structural consequences. Earlier torsions can move more atoms through longer lever arms. A tighter local description is the pullback quadratic form

$$G_{\boldsymbol{\theta}} = \boldsymbol{J}_H(\boldsymbol{\theta})^\top \operatorname{diag}(w_1 I_3, \dots, w_a I_3) \boldsymbol{J}_H(\boldsymbol{\theta}), \quad \|H_\lambda(\boldsymbol{\theta} + v) - H_\lambda(\boldsymbol{\theta})\|_w^2 = v^\top G_{\boldsymbol{\theta}} v + O(\|v\|^3). \tag{57}$$

Here v is a small local angular displacement. This expression follows by Taylor expansion of the smooth decoder; $\mathcal{L}$ is a conservative global bound, not an estimate of typical structural error. On a feasible stratum, use a matrix $\boldsymbol{Z}_{\boldsymbol{\theta}}$ with orthonormal columns spanning $\ker dC_\lambda$. The relevant task Jacobian is $d\mathcal{O}\, \boldsymbol{J}_H \boldsymbol{Z}_{\boldsymbol{\theta}}$. Its smallest nonzero singular value controls the section lower bound in Eq. (**??**). Using the unrestricted ambient Jacobian would miss the effect of closure. In a redundant task, null-space motion still prevents an inverse Lipschitz upper bound from following from that singular value alone.

Statistical consequences. For a supplied molecular inverse atlas with feasible decoders ψ_i, retain all assumptions of Appendix E.4 and use $\tau_i = L_{\mathcal{O}}\mathcal{L}D_i$. Then its leading separate-chart cost is

$$\sum_i \left(1 + \frac{\kappa_i L_{\mathcal{O}} \mathcal{L} D_i}{\varepsilon}\right)^r \left(1 + \frac{m\sigma_i^2 L_{\mathcal{O}}^2 \mathcal{L}^2 D_i^2}{\varepsilon^2}\right) \tag{58}$$

under the allocation in Eq. (12), with the same logarithmic factor. It is a sufficient budget for the stipulated noisy chart-regression problem. Molecular closure does not automatically provide the convex coordinate ranges or globally feasible decoders required there. A finite compact evaluation region and verified charts are part of the model.

For an unrestricted Lipschitz *forward* regression class on supplied s-dimensional feasible charts, Proposition 6 instead has accuracy exponent $s + 2$ under fixed positive response noise. On the ambient torus $s = d$; regular equality constraints can reduce it to $d - c$. Dimensional constants, chart distortion, sampling coverage, and the number of charts remain in the budget. Energy concentration or correlated torsions do not by themselves justify replacing s by a smaller dimension. Conversely, known internal-coordinate structure or shared unknown parameters can avoid the full nonparametric class. The calibration separation in Section 4 illustrates this possibility without asserting that arbitrary molecular geometry is linear in its unknown parameters.

I.4 When the dihedral angles are genuinely latent

Angles computed from complete, nondegenerate labeled atomic coordinates are available regression coordinates, even if a generative model calls them latent. With partial observations $X = \mathcal{O}(H_\lambda(\Theta)) + \xi$, they can instead be unobserved and nonidentifiable. If two torsion vectors at angular distance $\Delta > 0$ produce the same noiseless observation, every deterministic encoder has error at least $\Delta/2$ on

one of them. This follows immediately from the triangle inequality. More data drawn from the same ambiguous observation model cannot remove that obstruction.

If both the decoder and latent law are unconstrained and unknown, the pairs (H, ν) and $(H \circ h^{-1}, h_{\#}\nu)$ produce identical coordinate distributions for any torus diffeomorphism h. Physical decoder structure, sufficiently informative observations, interventions, or an explicit equivalence convention are therefore needed to identify the angles. A learned Euclidean latent vector is not automatically a dihedral torus.

A useful conditional guarantee is still possible. Suppose an available encoder e has $d_{\mathbb{T}^d}(e(X), \Theta) \le \eta$ on the evaluation set, and $\widehat{F}$ has uniform error at most ε against F on the union of the true and encoded angle sets. Then

$$d_Y(\widehat{F}(e(X)), F(\Theta)) \le \varepsilon + L_{\mathcal{O}}\mathcal{L}\eta. \tag{59}$$

This is an error-transfer statement conditional on an encoder guarantee. It neither estimates that encoder's sample complexity nor licenses applying the supervised bounds to noisy inferred inputs. If hard feasibility is required, the encoder or its subsequent decoder must also land in Θ_λ; small angular error alone does not certify closure or clearance.

Relation to molecular generative learning. Torsional diffusion conditions on fixed local geometry and excludes independently rotating ring torsions from its free variables (Jing et al., 2022). DiffPack learns side-chain torsion distributions conditioned on a backbone (Zhang et al., 2023), whereas FoldingDiff generates backbone angle sequences (Wu et al., 2024). RiboFlow combines RNA frames, torsion prediction, and sequence information (Ma et al., 2025). Riemannian flow matching (Chen and Lipman, 2024) and the 2026 Riemannian MeanFlow preprint (Zhong et al., 2026) evaluate distributions of protein torsion pairs and seven-dimensional RNA suite angles. These are local angle-distribution benchmarks; success on them does not establish accurate whole-chain closure, Cartesian reconstruction, or physical folding dynamics.

A distribution over feasible conformations is a different output from one continuous inverse selection. Our results give no universal sample count for a diffusion or flow model. They also do not claim that randomness removes every obstruction: if a conditional generator $G(y, z)$ is continuous in y and satisfies $F(G(y, z)) = y$ for every target and every fixed seed z, fixing any seed produces a continuous section. Thus $B_0(F) > 1$ rules out that particular global exact generator. Distribution-valued outputs, weaker accuracy criteria, or multiple domains change the question. Learning the sequence-conditioned conformational law or an energy landscape remains distinct from learning deterministic internal-to-Cartesian kinematics.

I.5 A REPRODUCIBLE CHECK OF THE DECODER CALCULATION

The script `code/check_molecular_kinematics.py` tests the geometric calculation on ideal spatial chains with $d \in \{8, 32, 128\}$ torsions, $a = d + 3$ atoms, unit bond lengths, fixed bond angle 1.9 radians, and an anchored initial triple. All atoms have weight $1/a$. For this particular chain, using contour lengths in Eq. (54) gives

$$\mathcal{L}^2 = \frac{1}{d+3} \sum_{j=1}^{d} \sum_{k=1}^{j} k^2 = \frac{d(d+1)^2(d+2)}{12(d+3)}. \tag{60}$$

This worst-case upper constant grows as $d^{3/2}$; it is not a lower bound or a fitted growth law for real proteins.

Using seed 25092026, we draw 100 uniform torus configurations per dimension and independent unit angular directions. We check the analytical Jacobian-vector product against a central difference with step 10^{-6}, test finite angular perturbations of norms 10^{-4}, 10^{-2}, and 10^{-1}, and verify a 2π coordinate shift. The generated table reports the measured derivative error and the maximum ratio of Cartesian perturbation to the rightmost bound in Eq. (56). The script also computes Smith factors from exact integer minors for the torus examples. These are arithmetic and numerical consistency checks of proved statements, not protein/RNA training results.

A biological evaluation should specify atom correspondence, torsion masks, fixed versus varying bond geometry, closure constraints, and sequence or structural splits before training. It should report angular and aligned Cartesian error, closure residual, steric violations, and behavior near singular task

Table 6: Ideal spatial-chain checks, 100 configurations per dimension. $\mathcal{L}$ is the proved global upper constant; the local norm is measured from the weighted Jacobian. The final columns are maxima over the stated checks. These are not molecular learning benchmarks.

d	$\mathcal{L}$	Mean local norm	Derivative error	Error/bound
8	7.006	2.643	1.62e-09	0.3726
32	53.113	9.766	4.02e-09	0.1114
128	419.701	35.354	1.00e-08	0.0262

Jacobians, alongside held-out likelihood or distributional metrics when the output is an ensemble. Such a study is needed to test whether a chart or structure-sharing strategy improves molecular learning in practice; the present experiments do not supply that evidence.

J Unknown representations are a separate problem

Our learning bounds assume the regression input and its metric are available. An observation embedding $\phi : Q \to \mathbb{R}^D$ must be injective, with quantitative distortion control, before intrinsic dimension can be transferred to the observed regression problem. A general camera observation can be noninjective, and the resulting ambiguity is not covered by a manifold-embedding assumption.

Explicit manifold recovery is a different estimation objective. Genovese et al. (2012) obtain a Hausdorff rate of order $N^{-2/(d+2)}$ up to logarithms for a specific perpendicular bounded-noise model. Kim and Zhou (2015) study tight rates for a different on-manifold sampling model; these results cannot be substituted without matching the observation assumptions. We do not assign a universal reconstruction penalty to kinematics learning.

Even a small Hausdorff error does not by itself provide an algorithmically available correspondence, a bi-Lipschitz chart, tangent accuracy, or an inverse section. If an *available* map $T : \mathcal{M} \to \widehat{\mathcal{M}}$ satisfies $\sup_x \|T(x) - x\| \le \eta$, a target extension f is L_f-Lipschitz, and $\widehat{f}$ has uniform error at most ε against that extension on $\widehat{\mathcal{M}}$, then the triangle inequality gives

$$\sup_{x \in \mathcal{M}} \|\widehat{f}(T(x)) - f(x)\| \le \varepsilon + L_f \eta. \tag{61}$$

This is a conditional upper transfer, not a minimax lower reduction. A constant target, for example, can be predicted without reconstructing its input manifold. Likewise, a forward Lipschitz constant alone cannot bound inverse error near a singularity. The computational hardness results of Kiani et al. (2024) address another resource and should not be conflated with information-theoretic sample counts.

K Experimental specification and complete results

The script `code/run_experiments.py` produces the original numerical studies. The spatial scripts listed in Appendix L add the redundant-arm study. All results are on simulated data. Randomness is generated with NumPy's generator and specified seeds; there are no external datasets. Original planar lengths are dimensionless link-length units; spatial 7R lengths are in metres. CPU timings are recorded for provenance and are not used for a solver-speed claim.

K.1 Circle and independent validation

The quadrature uses SciPy's adaptive integration with explicit split points at the seam boundary and peak, absolute tolerance 10^{-12}. A 65,536-point uniform grid records an auxiliary sampled maximum, which is not equated to the exact supremum. For validation detection, use degree two, $K \in \{4, 16, 64\}$, $V \in \{32, 128, 512\}$, and 2,000 independent trials per setting, with seed 807. Detect an angular residual strictly greater than $\pi/2$. All settings, theoretical values, and empirical frequencies appear in the supplied CSVs.

Table 7: Detection of a residual above $\pi/2$ in the degree-two circle construction. Observed frequencies use 2,000 independent validation trials per setting.

K	Validation size V	Exact probability	Observed frequency
4	32	0.839	0.846
4	128	0.999	0.999
4	512	1.000	1.000
16	32	0.386	0.388
16	128	0.858	0.851
16	512	1.000	0.999
64	32	0.117	0.112
64	128	0.392	0.386
64	512	0.863	0.857

Table 8: Additional measured 2R metrics over five seeds. Maximum means the largest residual among the fixed 12,000 test targets, not a certified workspace supremum. Success is the fraction with task error below 0.01.

N	Output	Sampled max	Joint RMSE (rad)	Success fraction
256	Angles	3.292 ± 0.060	0.645 ± 0.098	0.003 ± 0.002
256	Periodic	0.242 ± 0.041	0.114 ± 0.049	0.047 ± 0.021
1024	Angles	3.323 ± 0.043	0.387 ± 0.015	0.018 ± 0.005
1024	Periodic	0.132 ± 0.016	0.053 ± 0.002	0.144 ± 0.036
4096	Angles	3.330 ± 0.033	0.312 ± 0.024	0.031 ± 0.012
4096	Periodic	0.080 ± 0.035	0.024 ± 0.002	0.428 ± 0.102

K.2 Planar inverse-learning protocol

Let $\beta_0 = 0.2$ and $r(\beta) = \sqrt{1 + 0.7^2 + 1.4\cos\beta}$. Sample $\theta \sim \mathrm{Unif}[-\pi, \pi)$ and $r^2 \sim \mathrm{Unif}[r(\pi - \beta_0)^2, r(\beta_0)^2]$ independently; this is uniform workspace area. Use $x = (r\cos\theta, r\sin\theta)$ and label the positive-elbow solution from Eq. (18). Inputs are divided by 1.7. The angle targets are $(q_1, q_2)/\pi$, with principal $q_1 \in [-\pi, \pi)$; periodic targets are $(\cos q_1, \cos q_2, \sin q_1, \sin q_2)$. These losses differ by design, as this is a representation ablation. They are compared with the same physical task-space test metric.

Both models use scikit-learn's `MLPRegressor`, tanh hidden layers of width 64 and 64, Adam with initial learning rate 0.002, L^2 regularization parameter 10^{-5}, batches of at most 128, and at most 1,000 epochs. Internal validation fraction is 0.15; early stopping uses validation score, tolerance 10^{-6}, and patience 60. Fit seeds are $0, \ldots, 4$. For each sample size, training data use generator seeds $100 + \mathrm{seed}$; the same targets serve both methods. Across sample sizes these generator streams need not produce nested training sets. The independent fixed test set uses seed 2901 and contains 12,000 points, of which 396 satisfy the near-cut criterion $||q_1| - \pi| < 0.1$. Twelve of the 30 fits reach the epoch limit. The smallest periodic pair norm across all test predictions is 0.4860.

Predicted angles decode by multiplying by π, and periodic outputs decode pairwise with $\mathrm{atan2}$. The test task residual is the Euclidean distance between the analytic forward image of the prediction and the target. We record RMSE, mean, 95th percentile, sampled maximum, near-cut RMSE, angular joint RMSE, success fraction below 0.01, epoch count, parameter count, elapsed fitting time, and the minimum periodic pair norm on the test set. The raw result files retain convergence warnings; reaching the epoch budget is not treated as proof of optimizer convergence. Dispersion in all MLP tables is sample standard deviation across five fits, not a confidence interval for population risk. The analytic IK solution has numerical residual below 10^{-12} by construction, so these comparisons make no superiority claim over it.

K.3 Forward calibration and conditioning

For n joints, use true lengths $\mathrm{linspace}(0.5, 1, n)$. Draw independent joint angles uniformly on $[-\pi, \pi)$, build the real/imaginary feature matrix of Eq. (13), and add independent Gaussian noise of standard deviation 0.03 to each Cartesian coordinate. Fit unconstrained ordinary least squares.

Table 9: Forward calibration over 200 independent repetitions per setting. Conditional expectation is $\sigma^2 \operatorname{tr}((A^\top A)^{-1})$, averaged over training designs; normalized MSE is $N\,\mathrm{MSE}/(n\sigma^2)$, with $\sigma = 0.03$.

Joints n	Pairs N	Observed mean MSE	Conditional expectation	Normalized MSE
2	32	$5.79e-05$	$5.71e-05$	1.030
2	64	$2.98e-05$	$2.83e-05$	1.060
2	128	$1.54e-05$	$1.41e-05$	1.097
2	512	$3.8e-06$	$3.52e-06$	1.082
2	2048	$8.57e-07$	$8.79e-07$	0.975
4	32	0.000111	0.000119	0.991
4	64	$5.48e-05$	$5.76e-05$	0.975
4	128	$2.59e-05$	$2.85e-05$	0.920
4	512	$6.54e-06$	$7.05e-06$	0.931
4	2048	$1.7e-06$	$1.76e-06$	0.969
8	32	0.000261	0.000253	1.159
8	64	0.000123	0.000119	1.093
8	128	$5.84e-05$	$5.78e-05$	1.038
8	512	$1.38e-05$	$1.42e-05$	0.981
8	2048	$3.63e-06$	$3.52e-06$	1.033
16	32	0.000592	0.00059	1.316
16	64	0.000244	0.000256	1.085
16	128	0.000121	0.00012	1.074
16	512	$2.9e-05$	$2.85e-05$	1.032
16	2048	$7.03e-06$	$7.06e-06$	1.000

Table 10: Calibration complexity from the same 4,000 recorded trials. The ratio d_{cal}/n is one for an orthogonal design and larger under imperfect excitation. Entries are mean $\pm$ sample standard deviation over 200 designs; no models are retrained.

Unknown lengths n	$N = 32$	$N = 2048$
2	1.0143 ± 0.0202	1.0003 ± 0.0004
4	1.0534 ± 0.0327	1.0008 ± 0.0005
8	1.1255 ± 0.0367	1.0017 ± 0.0005
16	1.3119 ± 0.0504	1.0037 ± 0.0005

There are 200 repetitions for each (n, N), with generator seed sequences $(871, n, N, \mathrm{seed})$. The exact integrated test MSE is $\|\widehat{\boldsymbol{\ell}} - \boldsymbol{\ell}\|^2$, and the conditional noise expectation is Eq. (14). These are two different quantities and are both recorded. The algorithm knows the planar serial structure and joint offsets; only link lengths are learned.

For inverse conditioning, set $q_1 = 0$ and use 100 logarithmically spaced q_2 values in $[0.001, 0.5]$. Compute singular values by SVD. Compute the derivative of the explicit inverse independently with central differences in each Cartesian input direction and step $h = \min\{10^{-7}, 10^{-3} d_{\mathrm{out}}\}$. Compare the resulting matrix with J^{-1}. The log–log slope is an ordinary least-squares fit restricted in advance to $q_2 < 0.1$.

K.4 Additional checks of map complexity

The script `code/analyze_map_complexity.py` computes d_{cal} from the conditional expected errors of the existing 4,000 calibration trials; no additional fitted models are counted. It independently regenerates the seed-zero design for all 20 (n, N) settings and recomputes Eq. (15), agreeing with the recorded values to the displayed precision. Table 10 reports two sample sizes, with all settings in the accompanying CSV. These design statistics are distinct from the noisy realized prediction errors.

The same script checks the two explicit 2R manipulation rules from Appendix B with $\ell_1 = 1$, $\ell_2 = 0.7$, seed 2592026, 10,000 start-target pairs, and 21 equally spaced path samples per pair. Start and target radii are sampled independently and uniformly in area on the truncated annulus with $|q_2| \in [0.2, \pi - 0.2]$; the starting elbow sign is uniform. Six pairs replace the random angles by coincident, antipodal, and near-cut directions. The selector uses the short-angle rule when the

relative-direction cosine is at least $-1/2$, and the positive-angle rule otherwise. It is a selector between two rules, not a single globally continuous planner.

The maximum initial configuration error in unit-circle chordal coordinates is 8.1×10^{-16} or less; maximum endpoint task error is below 1.4×10^{-15}, and unit-circle norm error is below 6.7×10^{-16}. Every sampled path stays in its elbow sheet and truncated annulus. This is a numerical check of the analytic construction, not a learned-planning benchmark or a numerical proof that two domains are minimal. Minimality follows from the category bound.

L SPATIAL 7R PROTOCOL AND ADDITIONAL VALIDATIONS

Forward geometry and target construction. For joint j, the standard DH transform is

$$\boldsymbol{T}_j(q_j) = \mathrm{Rot}_z(q_j)\,\mathrm{Trans}_z(d_j)\,\mathrm{Trans}_x(a_j)\,\mathrm{Rot}_x(\alpha_j), \qquad \boldsymbol{T}(\boldsymbol{q}) = \prod_{j=1}^{7} \boldsymbol{T}_j(q_j). \tag{62}$$

Parameters appear in Section 5.1 and `results/spatial7r_config.json`. These are synthetic parameters, not a manufacturer-verified robot. Joints are treated as periodic and unconstrained. If $\boldsymbol{T}(\boldsymbol{q}_0) = (\boldsymbol{R}_0, \boldsymbol{x}_0)$, a target is $(\exp([\boldsymbol{\omega}]_\times)\boldsymbol{R}_0, \boldsymbol{x}_0 + \Delta\boldsymbol{x})$, with the six perturbation coordinates sampled uniformly and independently within the stated half-widths. This is uniform in the chosen local coordinates, not Haar-uniform in $\mathrm{SE}(3)$. A separate geometry pilot (seed 7123) selected the local box and redundancy offsets before generating training and test data. A larger trial offset of 0.65 rad did not give reliable local solves; the retained offset is 0.3 rad.

Inverse labels and baseline solver. For each of the two fixed final-joint values, Newton updates solve the first six Jacobian columns against world-frame position and rotation-vector residuals. Steps are capped at norm 0.4 rad, with at most 35 iterations and a combined numerical stopping tolerance 10^{-10}. This combined tolerance is an implementation convention on coordinates in metres and radians, not an invariant pose metric. It implies the separate residual bounds reported in the main text. No samples are omitted. For the unconstrained known-geometry baseline, the update is $\Delta\boldsymbol{q} = \boldsymbol{J}^\top(\boldsymbol{J}\boldsymbol{J}^\top + 10^{-7}\boldsymbol{I}_6)^{-1}\boldsymbol{e}$ with the same cap, tolerance, and common initial configuration, and an iteration cap of 20. The damping is in the raw coordinate convention. The separate position and geodesic orientation metrics are used for reporting both solvers and learners. This local test favors a good common initial guess and says nothing about numerical IK from arbitrary starts.

Data, targets, and optimization. For every N and seed $s \in \{0, 1, 2, 3, 4\}$, the generator seed is $73000 + s$. Training/validation sets are nested prefixes across N; different seeds give independent streams. Splitting uses seed $74000 + s$ and a 15% validation fraction by unique pose, so the two solutions to a held-out pose never leak into the fitting set. Inputs are divided by the six box half-widths. Labelled models append -1 or 1. Angle responses are $\boldsymbol{q}/\pi$; periodic responses concatenate $(\cos\boldsymbol{q}, \sin\boldsymbol{q})$ and decode with pairwise $\mathrm{atan2}$. The losses therefore differ intentionally. Neither representation, by itself, supplies a choice between redundant solutions.

The four parameter counts are 5,063 (pooled angles), 5,518 (pooled periodic), 5,127 (labelled angles), and 5,582 (labelled periodic). All use tanh hidden layers $(64, 64)$, Adam learning rate 0.002, regularization 10^{-5}, and batches of 128. Manual early stopping tracks mean squared validation response error, requires an improvement of 10^{-7}, waits 35 epochs, and runs at least 60 and at most 400 epochs. The best validation weights are restored. Initialization seeds are s. The test generator seed is 78192, independently generating 8,192 poses, each with two targets. No test scores determine epochs, hyperparameters, branch offsets, or which runs are reported. The complete 60-fit run took 189 seconds in the recorded CPU environment. Runtime is reported for reproducibility only. Position RMSE is $\sqrt{\mathbb{E}\|\widehat{\boldsymbol{x}} - \boldsymbol{x}\|_2^2}$. Orientation RMSE is the root mean square of the geodesic rotation angle, measured with the rotation vector of $\boldsymbol{R}\widehat{\boldsymbol{R}}^\top$. These expectations are empirical averages over the fixed held-out set. Success requires both a position error at most 5 mm and an orientation error at most one degree. Dispersion is sample standard deviation over five fits, not a population confidence interval. The raw CSV also retains 95th percentiles, maxima, joint-selection error, epochs, and elapsed times. A pooled model repeats the same prediction for the two labels; this is intentional. The label-conditioned comparison has more information available at inference. The oracle joint-mean calculation uses the exact two labels and does not require fitting a model.

Table 11: Spatial 7R learning at all sample sizes. Mean ± sample standard deviation over five seeds. Success requires position error ≤ 5 mm and orientation error $\leq 1^\circ$. Every test pose is evaluated with both designated solutions.

N	Method	Position (mm)	Orientation (°)	Success (%)
256	Pooled / angles	79.70 ± 0.96	7.19 ± 0.25	0.00 ± 0.00
256	Pooled / periodic	76.28 ± 0.99	6.62 ± 0.10	0.00 ± 0.00
256	Labelled / angles	16.99 ± 0.96	2.10 ± 0.22	5.72 ± 2.09
256	Labelled / periodic	6.45 ± 0.48	0.82 ± 0.06	52.70 ± 5.13
1024	Pooled / angles	79.64 ± 4.56	6.65 ± 0.74	0.00 ± 0.00
1024	Pooled / periodic	78.45 ± 4.86	6.69 ± 0.61	0.00 ± 0.00
1024	Labelled / angles	8.47 ± 1.33	1.05 ± 0.15	29.20 ± 8.83
1024	Labelled / periodic	3.42 ± 0.93	0.41 ± 0.11	86.98 ± 14.97
4096	Pooled / angles	77.36 ± 8.25	6.44 ± 0.65	0.00 ± 0.00
4096	Pooled / periodic	75.52 ± 3.56	6.45 ± 0.43	0.00 ± 0.00
4096	Labelled / angles	6.41 ± 1.85	0.77 ± 0.30	48.36 ± 18.08
4096	Labelled / periodic	3.02 ± 0.82	0.36 ± 0.11	91.05 ± 7.50

Derivative and redundancy checks. The geometric Jacobian column is $[\boldsymbol{z}_j \times (\boldsymbol{x}_{\mathrm{ee}} - \boldsymbol{o}_j); \boldsymbol{z}_j]$, where $\boldsymbol{z}_j$ and $\boldsymbol{o}_j$ are the joint axis and origin in world coordinates. Central differences with step 10^{-6} test all seven columns at 1,000 independently sampled torus configurations (seed 77001). All sampled weighted Jacobians have rank six at threshold 10^{-9}; the smallest sampled singular value is 1.67×10^{-4} m/rad. This is not a uniform rank certificate. Periodicity checks change one joint by 2π and give maximum transform-entry error 1.14×10^{-15}. The maximum combined column error of 2.98×10^{-10} bounds separately the translational component error (m/rad) and angular component error (rad/rad) numerically in these units.

Independent section differentiation uses 128 held-out poses and both labels, giving 256 points and 1,536 tangent-direction checks. The perturbation step is 10^{-5} in weighted task coordinates, with label solves tightened to 10^{-13}. The largest relative discrepancy with Eq. (17) is 4.08×10^{-8}, and $\|\boldsymbol{K}\boldsymbol{D}\boldsymbol{s} - \boldsymbol{I}_6\|_F$ stays below 1.99×10^{-15}. All 64 box corners and both labels also solve, with maximum combined residual 8.14×10^{-11}. Corners and random points remain finite checks, not a proof that every interior target belongs to a single regular chart.

The weighted pseudoinverse obeys $\|\boldsymbol{K}^\dagger\|_2 \sigma_6(\boldsymbol{K}) = 1$ to 1.34×10^{-15}. Adding a null-space term $\beta \boldsymbol{v}_0 \boldsymbol{u}_6^\top$ produces algebraic right inverses with median norm ratios $1.00, 1.00, 1.27, 7.96$ for $\beta = 0, 1, 10, 100$. These matrices demonstrate the pointwise freedom; arbitrary matrix fields of this form need not integrate to sections. Figure 4 also follows $\boldsymbol{q} = t(0.3, -0.8, 0.5, 1, -0.4, 0.7, 0.2)$ towards the straight-arm singularity. Neither plot establishes a global section or a global manipulation planner.

Reproduction and audit. Run `code/run_spatial7r.py` for data and fitting, `code/verify_spatial7r.py` for independent weight replay and section checks, and `code/make_spatial7r_figures.py` for tables and plots. The model geometry is in `code/spatial7r.py`. All 60 saved weight sets reproduce the recorded test metrics exactly in the independent NumPy evaluator. The original 30 planar weight sets also pass their retained replay check. Finite numerical agreement corroborates implementations, whereas the proofs establish the general claims under their stated assumptions.

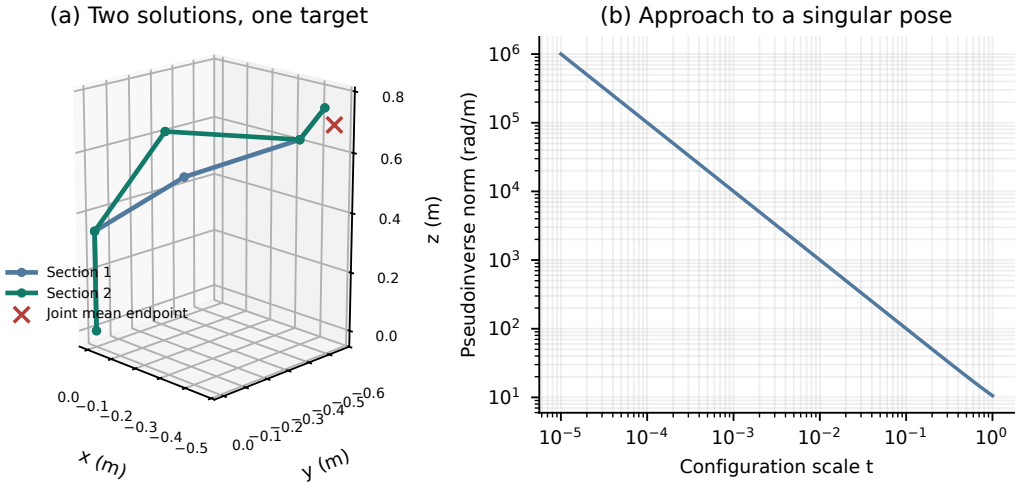

Figure 4: Spatial geometry checks. Left: two joint configurations for the same held-out pose; the mean joint vector misses that pose. Right: the weighted pseudoinverse norm along a prescribed approach to a singular configuration. The fixed translation–rotation conversion is 0.25 m/rad.