EXPRESSIVITY BARRIERS IN CATEGORICAL EQUIVARIANT NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Categorical equivariant networks process features linked by restriction maps—rules that transfer a feature between related objects—and require outputs to commute with these transfers. We ask when such networks can approximate every continuous target obeying this rule, with arbitrarily small uniform error on compact sets. For a specified layer class, we give two complementary results that remain valid as depth and hidden width increase. First, in a simple graph example, every network output is affine, so the target $z \mapsto \|z\|^2 z$, which obeys the same rule, has a positive minimum approximation error. Input-dependent scalar gates restore nonlinear dependence but may leave output directions unreachable. For graph systems whose cycle symmetries form a finite group, a target can be approximated arbitrarily well exactly when its value at each input lies in the available directions. Independent inputs can make deficient spans rare, reducing average error while leaving nonzero worst-case error at deficient inputs. We formalize the algebraic parts in Cubical Agda and test the obstruction and a repair in a controlled experiment. These results guide whether to add scalar information, independent inputs, or vector features when depth and width cannot help.

1 INTRODUCTION

Adding layers or channels is a natural way to make a network more flexible. That strategy can fail if the permitted operations never produce the kind of output a task requires. We study such failures in networks that attach features to vertices, edges, or other related objects and must process them consistently. Before making a network larger, we ask whether its chosen features and layers allow the required nonlinear response at all. We require the network to commute with every restriction map in its feature diagram, a property we call *categorical equivariance*. Our barriers concern this layerwise requirement; they do not apply to message passing or sheaf diffusion in general (Appendix H).

Figure 1 makes this failure concrete. The edge e and its endpoints v, w each carry a real number. The two edge-to-endpoint maps in the figure are identities, and all three objects map to the zero vector space at $\perp$ (for e, by composition through either endpoint). Therefore, for any continuous $f : \mathbb{R} \to \mathbb{R}$ with $f(0) = 0$, the component assignment $F_e = F_v = F_w = f$ and $F_\perp(0) = 0$ commutes with every arrow in the figure. In particular, $f(x) = x^3$ is a valid target. The permitted layers defined in Section 2 consist of natural affine maps, bottom-based scalar gates, and nonlinear maps applied to collected restricted copies. In this example, their scalar gates see only the constant value at $\perp$, and nonlinear operations on collected restricted copies cannot transmit a nonlinear dependence back to the edge or endpoint outputs. Every resulting output is therefore affine, $x \mapsto ax + b$, so it cannot approximate $x \mapsto x^3$ arbitrarily well. The affine-barrier theorem below proves that this limitation survives arbitrary hidden features, branches, and repeated nonlinear processing of collected features.

A standard approximation strategy provides a useful comparison. Group averaging turns ordinary approximators into equivariant ones (Yarotsky, 2022; Ravanbakhsh, 2020). Its categorical counterpart first collects each feature and its restricted copies in an arrow bundle, applies nonlinear maps to this temporary collection, and then reconstructs the output (Maruyama, 2025). In the edge example, the failure occurs at this final reconstruction step: although the bundle can contain nonlinear values, the permitted natural maps back to the output diagram cannot carry that nonlinear dependence into the output. The resulting map is therefore affine, which is the affine barrier established below.

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.

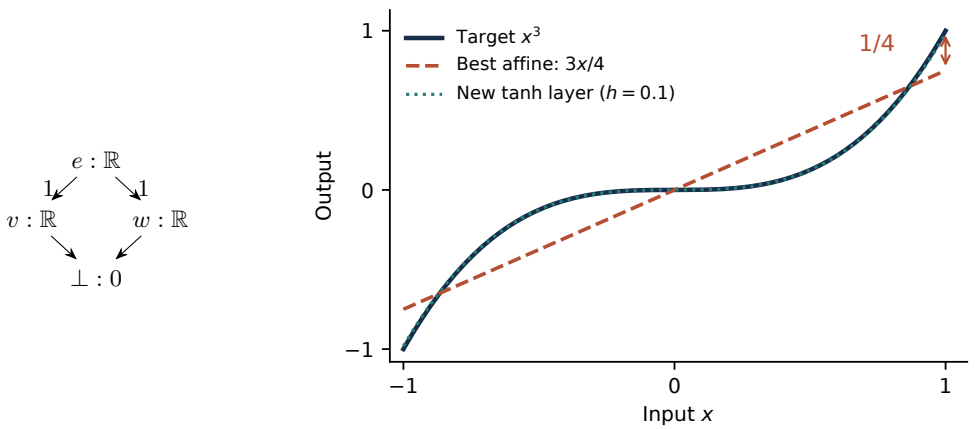


Figure 1: Edge example and approximation. Left: edge and endpoint features with identity restrictions and a zero bottom stalk; the permitted class yields affine outputs. Right: the cubic target, its optimal affine approximant, and a two-unit tanh map from the enlarged class. The best affine uniform error is $1/4$; sampled tanh errors are reported in Appendix F.

Our first criterion rules out approximation of targets that require nonlinear variation in directions the layers cannot supply. It computes, from linear compatibility equations, a subspace containing all nonlinear variation the layers can produce on inputs with the same bottom restriction. Outside this subspace, every generated map is affine on inputs sharing the same bottom restriction (Theorem 3.2). On graph local systems this gives an exact affine reachable class and best uniform cubic error $1/4$ (Theorem 4.2). A learning task tests this obstruction and its repair (Section 5).

The first obstruction motivates richer scalar channels. These can restore nonlinear responses, but can still leave vector directions unavailable. Our second criterion checks whether each target output lies in the vector subspace available at that input. On finite-holonomy local systems, this condition is necessary and sufficient for approximation by the enlarged class (Section 6). The criterion identifies when scalars suffice and when a vector feature is required. It also explains the gap between average and worst-case error: independent inputs can make deficient spans rare, while a fixed deficient input still imposes a nonzero uniform error. Copying features into wider hidden layers cannot change that span.

The ingredients of the positive results are established: invariant scalar coefficients on equivariant features, input-span descriptions of reachable outputs, completeness criteria from a spanning basis, and the planar dihedral covariants (Yarotsky, 2022; Villar et al., 2021; Cen et al., 2025; Solomon, 1963). What is new is that the available directions are *derived* from a categorical generator set rather than assumed, and that the derivation survives arbitrary hidden diagrams, nested bundles, branches, and unbounded depth and width. The feature and scalar diagrams are themselves substantive choices: even adding an empty bottom object can preserve every target map while making the original scalar gates input-independent (Section 6).

To express these compatibility conditions uniformly, we use homotopy type theory (HoTT). It keeps features and reversible symmetry transport together, and indexes network expressions by input and output diagrams. Induction over these expressions yields the two invariants; Agda checks their algebraic proofs. Noninvertible restrictions remain explicit diagram maps. Section 8 explains the language; Appendix K gives a worked introduction.

2 Feature diagrams and permitted layers

Let C be a finite category. A feature diagram is a functor $X : C^{\mathrm{op}} \to \mathrm{Vect}_{\mathrm{fd}}$ over $\mathbb{R}$. An arrow $u : b \to a$ induces a linear restriction $X(u) : X(a) \to X(b)$. A continuous natural map $F : X \to Y$ is a family of continuous maps satisfying

$$F_b X(u) = Y(u) F_a.$$

In words, processing and then restricting must agree with restricting and then processing. We call such a map *natural*. Write $\mathrm{Nat}_{\mathrm{cts}}(X, Y)$ for these maps and $\mathrm{Nat}_{\mathrm{lin}}(X, Y)$ when every component is linear. For specified compact sets $K_a \subset X(a)$, approximation means uniform convergence on each K_a; since C is finite, the error is the maximum of the objectwise errors. When we say *compact-open convergence or closure*, we mean uniform convergence on every compact subset of each stalk.

Arrow bundles. For an object b and a vector space A, define

$$I_b(A)(a) = \prod_{u:b\to a} A, \qquad (I_b(A)(w)h)_u = h_{w\circ u}. \tag{1}$$

Here $w : a \to c$, $h \in I_b(A)(c)$, and $u : b \to a$. An empty product is the zero vector space. Set

$$DZ = \bigoplus_{b\in C} I_b(Z(b)), \qquad (\Delta_Z)_a(z)_u = Z(u)z. \tag{2}$$

Thus $DZ(a)$ has one slot for every arrow into a, labelled by its source. The identity-arrow slot makes Δ_Z injective. Any family of globally Lipschitz maps $H_b : U(b) \to V(b)$ induces a natural map

$$H_\downarrow : DU \to DV, \qquad (H_\downarrow h)_u = H_b(h_u) \quad (u : b \to a). \tag{3}$$

Reindexing changes the arrow and preserves its source label, which proves naturality even when the maps H_b are nonlinear.

For the obstruction, take $C = P$ to be a finite poset with least object $\bot$. Write $r^X_{da} = X(d \le a)$ and $J_d = I_d(\mathbb{R})$. Then $J_d(a) = \mathbb{R}$ for $d \le a$ and is zero otherwise; $J_\bot$ is the constant scalar diagram.

Definition 2.1 (Layer class)**.** Fix a continuous scalar activation $\alpha : \mathbb{R} \to \mathbb{R}$. Let $\mathcal{A}(X, Y)$ consist of the maps generated by finite composition and finite parallel concatenation from:

1. natural affine maps $L + c : Z \to Z'$, where L is a natural linear map and the bias satisfies $Z'(u)c_a = c_b$ for each $u : b \to a$;
2. scalar gates $z \mapsto \alpha(s_a(z))z$, where $s : Z \to J_\bot$ is a continuous natural map;
3. the lifts $\Delta_Z : Z \to DZ$ and the componentwise maps $H_\downarrow : DU \to DV$ in (3).

All intermediate diagrams are finite-dimensional and may themselves be iterated arrow bundles. Parallel concatenation uses objectwise direct sums.

In the finite point-base setting with counting measures, this class contains the CENN generators of Maruyama (2025, Definitions 4–10). We allow all natural affine maps and even continuous natural scalar channels, so the negative theorem also covers more restrictive choices of these generators. Conversely, every natural affine map can be implemented by an identity-arrow convolution; Appendix A gives the matching. We do not permit an arbitrary natural nonlinear map on an arbitrary diagram: that would include the target cubic itself. This requirement differs from equivariance under automorphisms of a full graph configuration, as in the OENN setting of Maruyama (2026). Our results concern the restriction-commuting generators above, including the categorical CENN generators cited here. A message-passing layer may combine independently stored neighboring features. Our generators process a feature together with its specified restrictions, requiring the result to agree before and after restriction. This difference in information flow explains the scope stated in the introduction.

3 A TRACE THAT CONFINES NONLINEAR DEFECTS

The trace collects directions that natural linear maps can receive from non-bottom bundle slots. Its quotient retains the remaining directions; these will be forced to behave affinely on inputs with the same bottom restriction. Both spaces are computed from linear naturality alone. Define the *non-bottom injective trace* and its quotient by

$$\mathsf{T}(Z) = \sum_{d\neq\bot} \sum_{\eta\in\mathrm{Nat}_{\mathrm{lin}}(J_d,Z)} \mathrm{im}\,\eta, \qquad \mathsf{Q}Z = Z/\mathsf{T}(Z). \tag{4}$$

Images and sums are taken objectwise. The sum is a subfunctor; bases of the finite-dimensional natural-transformation spaces suffice to compute it. Let $\pi_Z : Z \to \mathsf{Q}Z$ denote the quotient.

Lemma 3.1. *Every natural linear map preserves* T*, and*

$$\mathsf{T}(Z \oplus W) = \mathsf{T}(Z) \oplus \mathsf{T}(W), \qquad \mathsf{T}(DZ) = \bigoplus_{d \neq \bot} J_d \otimes Z(d). \tag{5}$$

In particular, $\mathsf{Q}(DZ) \cong J_\bot \otimes Z(\bot)$.

Proof. Preservation follows by composing maps out of each J_d. Projections and inclusions prove the direct-sum identity. Every non-bottom summand of DZ lies in its trace. For the reverse inclusion, consider a natural linear map $\eta : J_d \to J_\bot \otimes U$ with $d \neq \bot$. Its bottom component is zero because $J_d(\bot) = 0$. Naturality to the bottom, where every target restriction is the identity, forces all its components to be zero. Thus a trace map has no component in the bottom summand. □

Theorem 3.2 (Affine barrier)**.** *Assume* $C = P$ *is a finite poset with least object* $\bot$*, and all diagrams are finite-dimensional. For every* $\Phi \in \mathcal{A}(X, Y)$ *and every object* a*,* $\pi_{Y,a}\Phi_a$ *is affine on each fibre (the set of inputs with the same restriction to* $\bot$*) of* $r^X_{\bot a}$*. This property also holds for every compact-open limit of such networks.*

Proof. Fix the input X, an object a, and $x, y \in X(a)$ with $r^X_{\bot a}x = r^X_{\bot a}y$. For a partial network $F : X \to Z$ and $t \in [0, 1]$, consider its affine defect

$$\delta_t F = F_a((1-t)x + ty) - (1-t)F_a(x) - tF_a(y).$$

We prove $\delta_t F \in \mathsf{T}(Z)(a)$ by induction on the network expression. The identity has zero defect. Postcomposition with $L + c$ sends the defect to $L_a \delta_t F$, which lies in $\mathsf{T}(Z')(a)$ by Lemma 3.1.

For a scalar gate, naturality gives

$$s_a(F_a(z)) = s_\bot(F_\bot(r^X_{\bot a}z)).$$

The gate therefore has the same scalar multiplier at x, y, and their convex combination. Its output defect is that scalar times $\delta_t F$. The lift Δ_Z is already a natural linear map.

Now consider $H_\downarrow : DU \to DV$ after *any* preceding natural map $F : X \to DU$. In the quotient $\mathsf{Q}(DV)(a)$ only the bottom slot remains. Naturality of F identifies that slot with

$$H_\bot\big(F_\bot(r^X_{\bot a}z)\big),$$

using $DU(\bot) = U(\bot)$. This expression is constant on the chosen fibre. The output defect consequently belongs to $\mathsf{T}(DV)(a)$, regardless of how F was constructed or how often U was previously bundled. Parallel concatenation follows from the direct-sum identity in the lemma. These cases exhaust Definition 2.1.

The resulting affine identities pass to compact-open limits by evaluating at the three points $x, y, (1-t)x + ty$. Continuity then gives affinity on each affine fibre. □

The bundle case allows arbitrary nesting: a componentwise nonlinearity need not immediately follow the canonical lift. No natural splitting of the quotient is required. The theorem gives a necessary condition; it does not characterize density on arbitrary diagrams.

Corollary 3.3 (Finite error certificate)**.** *Equip* $\mathsf{Q}Y(a)$ *with the quotient norm. If* x, y *lie in one restriction fibre and a target* $F \in \mathrm{Nat}_{\mathrm{cts}}(X, Y)$ *has quotient midpoint defect*

$$d = \big\|\pi_{Y,a}F_a((x+y)/2) - \tfrac{1}{2}\pi_{Y,a}F_a(x) - \tfrac{1}{2}\pi_{Y,a}F_a(y)\big\| > 0,$$

then every network has uniform error at least $d/2$ *on those three points.*

Proof. Subtract the zero quotient defect of the network. The quotient projection has norm at most one, and the sum of the three absolute coefficients is two. □

4 EXACT GRAPH BARRIERS AND A CHANGE OF GENERATORS

Let Γ be a finite connected graph with at least one edge and no self-loops. Its augmented face poset consists of vertices, edges, and $\perp$, with $\perp < v < e$ for each incidence. Let $Y(\perp) = 0$ and let every incidence restriction $Y(e) \to Y(v)$ be an isomorphism. Such data form a local system on the non-bottom incidence diagram.

Lemma 4.1. *For this diagram,* $\mathsf{T}(Y) = 0$.

Proof. A natural map $J_e \to Y$ is zero at the endpoints. Naturality and an invertible incidence restriction force its edge component to vanish. For a map $J_v \to Y$, take an incident edge $e = (v, w)$. Since $J_v(w) = 0$, the equation at the other endpoint w forces the edge component to vanish. The equation at v then forces the vertex component to vanish. This applies to every incident edge and every generator in (4). □

Assume the incidence restrictions are orthogonal for chosen inner products. Choose a reference feature space (or stalk) V and a spanning tree of the non-bottom incidence graph. Transport along tree paths gives isometries $\tau_a : V \to Y(a)$. Transport around a cycle may return a vector in changed coordinates. The *holonomy group* records these changes: let $H \subset O(V)$ be generated by $\tau_d^{-1} r_{da}^Y \tau_a$ over all incidences. Write $C_H(V, V)$ for the continuous H-equivariant maps, $\mathrm{End}_H(V)$ for the linear ones, and V^H for the vectors fixed by H.

Theorem 4.2 (Reachable maps and best approximation error)**.** *For a positive-rank orthogonal local system as above, reference-stalk evaluation identifies*

$$\mathrm{Nat}_{\mathrm{cts}}(Y, Y) \cong C_H(V, V), \qquad \mathcal{A}(Y, Y) \cong \{z \mapsto Az + c : A \in \mathrm{End}_H(V),\ c \in V^H\}. \tag{6}$$

For $F_a(z) = \|z\|^2 z$ *at non-bottom objects and* $F_\perp(0) = 0$*,*

$$\inf_{\Phi \in \mathcal{A}(Y,Y)} \max_{a \neq \perp} \sup_{\|z\| \leq 1} \|F_a(z) - \Phi_a(z)\| = \frac{1}{4}. \tag{7}$$

Proof. Naturality along tree paths forces $F_a = \tau_a f \tau_a^{-1}$. The remaining equations say precisely that f commutes with the generators of H, hence with H. Bottom naturality is automatic. By Theorem 3.2 and Lemma 4.1, every network is affine on each full stalk. An affine map commutes with H exactly under the two conditions in (6); conversely, every such family is a permitted affine layer.

Orthogonal restrictions commute with the radial cubic, so F is natural. Restrict an affine approximant to a diameter $z = tv$, with v unit, and project onto v. Taking its odd part cannot increase the error on $[-1, 1]$, leaving a scalar approximation bt to t^3. For $e(t) = t^3 - bt$, the identity $e(1) - 2e(1/2) = 3/4$ gives $\|e\|_\infty \geq 1/4$. The natural map $\Phi = 3\mathrm{id}_Y/4$ attains this error because $\max_{0 \leq r \leq 1} r|r^2 - 3/4| = 1/4$. □

The zero bottom stalk is not merely a way to hide zero-dimensional inputs. For $\widetilde{Y} = (J_\perp \otimes \mathbb{R}^m) \oplus Y$, $m > 0$, every stalk is nonzero and $\mathsf{T}(\widetilde{Y}) = 0$. On the fibre with global coordinate $v = 0$, the natural target $(v, z) \mapsto (v, \|z\|^2 z)$ has the same lower bound. An independent global channel does not remove the obstructed local directions. The trace calculation uses invertible incidences and an incident edge at every vertex, as connectivity guarantees; Appendix C gives boundary cases.

Restoring density on the same diagram. When H is finite, add the following *direct* nonlinear maps on Y:

$$f_H(z) = \frac{1}{|H|} \sum_{h \in H} h^{-1} m(hz), \qquad N_a = \tau_a f_H \tau_a^{-1}, \quad N_\perp(0) = 0, \tag{8}$$

where $m : V \to V$ is an ordinary MLP with a continuous, globally Lipschitz, nonpolynomial activation. A change of variables makes f_H equivariant. To approximate a natural target, transport the objectwise compacts to V, take their finite H-saturation, and approximate the reference map there by m. Averaging and isometric transport do not increase the error. Thus these new layers are

dense in $\mathrm{Nat}_{\mathrm{cts}}(Y, Y)$. This is the finite-group case of Yarotsky (2022, Proposition 2.2), applied through (6). The feature functor is unchanged; the layer class is strictly enlarged. The distinction is forced by (7).

5 Learning a nonlinear response on the fixed diagram

The preceding obstruction is independent of optimization. We test its consequence for learning a specified nonlinear response, using the same input and output local system for the original and repaired models. Take a rank-two triangle with quarter-turn holonomy $H = C_4$. At a reference stalk, for $z = (x, y)$ uniform in the unit disk, set

$$F_\beta(z) = \|z\|^2 z + \beta(x^3 - 3xy^2,\ y^3 - 3x^2y). \tag{9}$$

This is the gradient of the fourfold-symmetric energy $[(x^2 + y^2)^2 + \beta(x^4 - 6x^2y^2 + y^4)]/4$. We use $\beta = 0$ for the radial target and $\beta = 1/4$ for an anisotropic response. Transport by the fixed local frames gives a natural target on the entire diagram. This is a synthetic response-learning task; the data are individual stalk vectors.

Writing J for the quarter turn, Theorem 4.2 identifies the entire original reachable class as $z \mapsto az + bJz$. Angular orthogonality gives the exact population risk

$$\mathbb{E}\,\|F_\beta(z) - az - bJz\|^2 = \frac{(a - 2/3)^2 + b^2}{2} + \frac{1}{36} + \frac{\beta^2}{4}. \tag{10}$$

Thus no original network can have population vector RMSE below $1/6$ or $5/24$ on the two tasks. These distributional optima differ from the uniform approximation error in Theorem 4.2.

Table 1: Test vector RMSE (mean $\pm$ sample standard deviation, five seeds). The exact original-class population optima are $1/6 \approx 0.1667$ and $5/24 \approx 0.2083$. The unrestricted MLP is a nonequivariant control. The original stack and Reynolds MLP have comparable parameter counts.

Model	Parameters	Radial RMSE	Anisotropic RMSE
Affine oracle	2	0.1669 ± 0.0001	0.2088 ± 0.0002
Original stack	1,250	0.1668 ± 0.0001	0.2088 ± 0.0003
Reynolds MLP	1,218	0.0038 ± 0.0011	0.0025 ± 0.0005
Unrestricted MLP	1,218	0.0088 ± 0.0012	0.0081 ± 0.0015

We compare a least-squares affine oracle, a permitted three-layer stack with multiplicities $1, 24, 24, 1$ and two learned constant scalar gates, a direct Reynolds MLP, and an unrestricted MLP. Both MLPs have widths $2, 32, 32, 2$ and tanh activations. The oracle covers the entire original class, including hidden diagrams beyond those in the trained stack.

Each seed uses 1,024 training and 1,024 validation samples, 2,000 Adam steps, validation-based checkpoint selection, and a shared independent 20,000-sample test set. The original stack approaches the analytic floor; the Reynolds model learns both nonlinear responses (Table 1). Both satisfy the incidence equations to numerical precision. Appendix G gives the complete protocol, residuals, learning curves, and checkpoint audit. This tests the predicted mechanism on a fixed diagram; broader learning efficiency remains open.

6 Designing scalar channels and vector features

The Reynolds experiment adds nonlinear vector layers. Could changing only the scalar channels suffice? A seemingly harmless diagram choice matters here: adjoining a zero bottom feature preserves all continuous natural target maps, but putting $\mathbb{R}$ there in the constant scalar diagram forces every scalar channel to be input-independent. Appendix H proves this distinction. The scalar diagram must be chosen along with the features.

Instead set $S_0(\perp) = 0$ and $S_0(a) = \mathbb{R}$ elsewhere, with identity incidence restrictions. On a zero-bottom local system, natural channels to S_0 are holonomy-invariant functions, including $z \mapsto \|z\|^2$.

These allow nonlinear gates. Whether they suffice depends on the output directions available from the vector features.

Let $\mathcal{A}_{\mathrm{sc}}$ retain all the original generators and admit every continuous natural scalar rescaling $G_a(z) = \lambda_a(z)z$. It permits arbitrary hidden diagrams, finite branches, and nested bundles.

Theorem 6.1 (Scalar-gate closure)**.** *Let Γ be a finite connected loopless graph with at least one edge, and let X, Y be zero-bottom orthogonal local systems on its augmented face poset. Choose reference spaces V, W and isometric tree frames, and let the finite joint holonomy group $H \subset O(V) \times O(W)$ be generated by the paired transports around graph cycles. An element $h = (h_V, h_W)$ acts on V by h_V and on W by h_W. Put $E(x) = \mathrm{Hom}_H(V, W)x + W^H$. Then*

$$\overline{\mathcal{A}_{\mathrm{sc}}(X, Y)} \cong \{f \in C_H(V, W) : f(x) \in E(x) \text{ for all } x\}.$$

Writing $H_x = \{h \in H : h_V x = x\}$, the class is dense exactly when $E(x) = W^{H_x}$ for every x.

Here $E(x)$ is the set of values supplied by affine equivariant maps at x; W^{H_x} consists of vectors fixed by every h_W with $h \in H_x$. It is the set allowed by symmetry alone. Density requires the available directions to exhaust the symmetry-allowed ones.

Proof idea. For a fixed original input X, track the subspace $\{L_a x : L \in \mathrm{Nat}_{\mathrm{lin}}(X, Z)\} + \{c_a : c \text{ a natural bias}\}$. Natural affine maps and scalar gates preserve membership; for bundle targets this subspace is the entire stalk. This proves an upper bound through arbitrary nesting, reducing to $E(x)$ at the output. Conversely, invariant scalar coefficients multiplying a basis of affine equivariants approximate every continuous map taking values in $E(x)$. Regularized linear algebra handles points where its dimension changes. Finite averaged scalar MLPs compute the coefficients through S_0 channels. Agda checks the pointwise upper-bound induction. Full analytic proofs and the irreducible test are in Appendix I. □

For standard planar C_4 features, $E(z) = \mathrm{span}\{z, Jz\}$; scalar gates with sufficiently rich invariant channels therefore suffice. For the order-eight group D_4, the commutant is $\mathbb{R}I$, so $E(z) = \mathbb{R}z$: scalar gates cannot create a component transverse to the input.

Missing directions determine the best error. For any continuous equivariant target F, put $d_F(x) = \mathrm{dist}(F(x), E(x))$. On a compact input set K, with any probability measure μ supported there, Appendix J proves

$$\inf_{N \in \mathcal{A}_{\mathrm{sc}}} \mathbb{E}_\mu \|F - N\|^2 = \mathbb{E}_\mu d_F^2, \qquad \inf_{N \in \mathcal{A}_{\mathrm{sc}}} \sup_K \|F - N\| = \sup_K d_F. \tag{11}$$

Thus the same missing directions have different consequences under different input distributions.

Independent inputs change average approximation. Since $E(x)$ depends on the input vectors, additional independent inputs can make deficient spans a measure-zero event and thereby reduce average error, without removing every deficient input. For finite irreducible representations, the criterion gives an exact threshold. Write $\mathbb{D} = \mathrm{End}_H(U)$ and $n = \dim_{\mathbb{R}} U / \dim_{\mathbb{R}} \mathbb{D}$; by real Schur theory, $\mathbb{D}$ is $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$.

Corollary 6.2 (Input-multiplicity thresholds)**.** *For an irreducible real representation U of a finite group, take reference input U^m and output U^k, $m, k \geq 1$. On the full unit input ball, the scalar-enlarged class approximates every continuous equivariant target uniformly exactly when $n = 1$, and in L^p under normalized Lebesgue measure, $1 \leq p < \infty$, exactly when $m \geq n$.*

The reason is that, for nontrivial U, $E(x)$ consists of k copies of the $\mathbb{D}$-span of the input vectors. It fills the output almost everywhere when $m \geq n$. Yet at $(u, 0, \ldots, 0)$ with u fixed only by the action's kernel, it remains proper if $n > 1$. Regularized coefficients give the positive result; a target outside that span gives the obstruction. Appendix J proves both directions, including the separate bias case for trivial U. Thus square symmetry needs two independent inputs for average universality, and cube rotations need three, while uniform universality fails at every multiplicity in both examples.

In particular, for D_4, $m \geq 2$, and $F(x) = \overline{x_1}^3$,

$$\inf_{N \in \mathcal{A}_{\mathrm{sc}}} \mathbb{E} \|F - N\|^2 = 0, \qquad \inf_{N \in \mathcal{A}_{\mathrm{sc}}} \sup_{\|x\| \leq 1} \|F - N\| = 1. \tag{12}$$

At $x = (e^{i\pi/8}, 0, \dots, 0)$, the target is unit and perpendicular to every allowed output. The zero network attains uniform error one. Population MSE nevertheless approaches zero. Here multiplicity counts independently varying input vectors; adding hidden width or copying a vector does not repair its span. Figure 3 in Appendix J uses explicit regularized approximators to show how the region of large error narrows around a fixed failing input. A distribution concentrated on deficient spans can retain average error even above the multiplicity threshold.

A vector feature repairs the uniform obstruction. To remove a uniform obstruction of this kind we must supply the missing directions. Already for a single planar D_4 input, adding $p(x, y) = (x^3 - 3xy^2, y^3 - 3x^2y)$ suffices for density. The features $z, p(z)$ span the stabilizer-fixed space at every input, and $\phi(x, y) = (x^2 + y^2, x^2y^2)$ separates D_4 orbits. Consequently

$$f_\theta(z) = a_\theta(\phi(z))z + b_\theta(\phi(z))p(z)$$

with ordinary MLP coefficients is dense in all continuous D_4-equivariant maps. Each term is an explicit scalar gate followed by a natural projection; the p branch is a newly added direct feature. Invariant-coefficient approximation is established (Yarotsky, 2022, Proposition 2.4); input-span and cross-product constructions appear in Villar et al. (2021, Propositions 4–5), and these polynomial covariants are classical (Solomon, 1963). The categorical result identifies exactly which directions the original affine and bundle operations, even with enlarged scalar gates, cannot supply. The same argument gives a two-feature repair for every planar dihedral group (Appendix J).

The design choices are separable in training. Table 2 compares four readout classes on $F_{1/4}$ of (9) over a two-cycle D_4 local system, trained under the protocol of Section 5 (Appendix G.4). Each measured error tracks its analytic floor, so constant coefficients, insufficient scalar information, and a missing vector direction are separable by measurement; only the added covariant removes the error.

Table 2: Test vector RMSE on the D_4 response $F_{1/4}$ (mean $\pm$ sample standard deviation, five seeds), against the exact population infima of Appendix I.4. The first row is a least-squares fit over the whole original reachable class, not a trained network.

Readout class	Analytic infimum	Test RMSE
Affine oracle $z \mapsto az$	$5/24$ (0.2083)	0.2087 ± 0.0003
Norm-only gate $a(\|z\|^2)z$	$1/8$ (0.1250)	0.1251 ± 0.0002
All natural scalar rescalings $a(\phi(z))z$	$1/(8\sqrt{2})$ (0.0884)	0.0889 ± 0.0004
Two-feature readout $a(\phi(z))z + b(\phi(z))p(z)$	0 (0.0000)	0.0021 ± 0.0008

The category matters as well as its holonomy. The same D_4 representation over a one-object group category admits a Reynolds-compiled map whose output is not parallel to its input. The poset bound therefore cannot be inferred from group symmetry alone (Appendix I).

7 Reconstruction and Noninvertible Actions

The obstructions above motivate a sufficient condition for reconstruction to succeed. Let C be a finite category and work over a field.

Proposition 7.1 (Reconstruction criterion)**.** *There is a natural linear map $R_Y : DY \to Y$ satisfying $R_Y \Delta_Y = \mathrm{id}_Y$ if and only if Y is injective in $[C^{\mathrm{op}}, \mathrm{Vect}_{\mathrm{fd}}]$. Over $\mathbb{R}$, maps $R_Y H_\downarrow \Delta_X$ are then compact-open dense in $\mathrm{Nat}_{\mathrm{cts}}(X, Y)$, already when each H_b is a finite MLP with a fixed continuous, globally Lipschitz, nonpolynomial activation.*

Proof idea. Coinduction makes DY injective. Its retracts are injective; conversely, injectivity extends id_Y across Δ_Y. Approximate the target in each bundle slot, then apply R_Y. The finite operator norm of R_Y controls the error. Appendix B gives the adjunction and the full approximation argument. □

The approximation mechanism is the finite equivariant compiler condition of Maruyama (2025), with classical injective-functor ingredients. Finite-poset injectives are direct sums of the J_d, with

variance reversed from the convention of Brown & Draganov (2021). For finite groups over $\mathbb{R}$, reconstruction is Reynolds averaging (Appendix B). However, failure of injectivity alone does not prove non-density; that requires an invariant such as Theorem 3.2. Noninvertible actions separate this criterion from monoid averaging: the conditions of Nasu & Maruyama (2027, pp. 33–35) force invertible output actions, while a categorical reconstruction can exist for a singular idempotent (Appendix D).

8 Dependent types as a language for equivariance

Our language connects two uses of dependent types. HoTT keeps feature spaces and reversible symmetries together: a dependent family assigns a space to each base point, and a path transports values invertibly. A vector-space family over the classifying type BG describes a representation; a section of its linear-map family is an intertwiner (Bezem et al., 2026). A G-set is a family $P : BG \to \mathrm{Set}$ whose total type presents the action groupoid. Noninvertible restrictions remain explicit diagram maps, not paths. Network expressions are indexed by their input and output diagrams; induction over this syntax establishes the two invariants. Thus the language links symmetry-compatible types to proofs about permitted layers. Appendix K gives a worked introduction.

Cubical Agda checks the algebraic network inductions over abstract modules (Vezzosi et al., 2019); closures, quantitative error bounds, and measure-theoretic thresholds are proved analytically. Appendix E records the scope and the dependent block-kernel proposition. Its formal version uses globally indexed sites; the general orbit/stabilizer coordinate derivation is a pen-and-paper result. These calculations support the language and are not additional expressivity theorems.

9 Discussion

Position relative to prior work. The compiler and its conditional approximation theorem originate with Maruyama (2025). Architecture-dependent lower bounds also occur for set networks (Wagstaff et al., 2022); ours permits arbitrary hidden dimensions and concerns restriction-commuting generators. Spanning covariants admit pointwise invariant-scalar decompositions (Cen et al., 2025). Our uniform closure also handles changing ranks; its analytic ingredient has a classical function-module interpretation (Chen, 2024). Exact closures and depth-dependent universality appear for permutation networks (Pacini et al., 2025; 2026), including a depth-independent one: pooling-free convolutional filters of width one reach exactly the coordinatewise maps at every depth $d \geq 2$ (Pacini et al., 2026, Example 3). Our barrier is likewise depth-independent, but differs in kind, since our generators leave an affine class.

Changing the feature diagram can remove the obstruction: independent endpoint features restore pointwise nonlinear operations, whereas restriction-commuting maps still cannot use neighboring inputs. Bundle neural networks (Bamberger et al., 2025) diffuse features over a learned flat bundle on a graph and approximate arbitrary feature transformations given injective positional encodings; their node mixing and frame-wise activation lie outside the present class. The detailed comparison is in Appendix H.

Limitations and next questions. Our complete characterization applies to the stated graph feature diagrams; outside this setting, the criteria do not provide a complete characterization. The experiments test responses at individual feature spaces, so performance on full graph-prediction tasks remains unestablished, and the input thresholds depend on the data distribution. The exact scalar-gate characterization also requires finite cycle holonomy; learned transports may generate infinite symmetry groups. Approximation results concern expressivity, not whether training finds good weights.

Conclusion. Feature and scalar choices shape what these networks can express, and our criteria guide the changes that depth and width cannot replace: richer scalar information restores nonlinearity, while independent inputs or new vector features supply missing directions. Yet better average approximation can still leave unavoidable error at particular inputs, so average error alone cannot certify small maximum error on compact sets.

REPRODUCIBILITY STATEMENT

Complete proofs and the generator correspondence are given in the main text and appendices. The supporting artifact contains the Agda sources, a checker that builds against pinned dependencies in a fresh temporary directory, exact linear-algebra checks, learning code and selected checkpoints, numerical protocols, and figure generators. Appendix E states precisely which results are machine-checked; Appendix F lists the reproduction commands.

AI USE STATEMENT

In this work, we used generative AI tools to generate synthetic data sets, to provide feedback on our research methodology and experiments, and to implement methods. Concretely, the Cubical Agda development, the exact verification scripts, the synthetic response-learning data and training code, and the figure generators were written largely with AI assistance, and AI tools were used to scrutinize and attempt to refute our proofs and to review the experimental protocol.

We have not used generative AI tools to help develop the theoretical models or conceptual frameworks, to formulate the mathematical claims, to provide critical ingredients for proving those claims, to assist in the writing of the proofs, to propose or refine hypotheses, or to interpret the results; these are the authors' own work. Assisting with translation, cleaning and reformatting data sets, and supporting qualitative and thematic data analysis are not applicable to this work.

Additionally, we used generative AI tools for tasks with recommended disclosure: creating and editing software code, creating artifacts, creating and modifying figures, suggesting experimental parameters, identifying, summarizing and analyzing existing literature, formatting references, and editing the manuscript to improve readability. The prose of the paper was written by the authors.

We have reviewed all AI-assisted work. The authors verified every theorem statement and proof in this paper by hand, and checked every bibliography entry against its primary source. The Agda development type-checks under `--cubical --safe` with no postulates against a pinned library commit, and Appendix E states exactly which statements that covers; the exact algebraic claims are independently recomputed by rational-arithmetic scripts; every archived response-learning checkpoint is reloaded and its recorded metrics recomputed by a checkpoint audit; and the D_4 design comparison is rerun deterministically from its script (Appendix F, Appendix G). We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Jacob Bamberger, Federico Barbero, Xiaowen Dong, and Michael Bronstein. Bundle neural networks for message diffusion on graphs. In *International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2405.15540v2. Comparison uses arXiv version 2, April 2025.

Marc Bezem, Ulrik Buchholtz, Pierre Cagne, Bjørn Ian Dundas, and Daniel R. Grayson. Symmetry, 2026. URL https://unimath.github.io/SymmetryBook/book.pdf. Book version 2c8c912, September 10.

Cristian Bodnar, Francesco Di Giovanni, Benjamin Paul Chamberlain, Pietro Liò, and Michael M. Bronstein. Neural sheaf diffusion: A topological perspective on heterophily and oversmoothing in GNNs. In *Advances in Neural Information Processing Systems*, 2022. URL https://arxiv.org/abs/2202.04579v4.

Adam Brown and Ondrej Draganov. Computing minimal injective resolutions of sheaves on finite posets, 2021. URL https://arxiv.org/abs/2112.02609v2.

Jiacheng Cen, Anyi Li, Ning Lin, Tingyang Xu, Yu Rong, Deli Zhao, Zihe Wang, and Wenbing Huang. Universally invariant learning in equivariant GNNs. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://arxiv.org/abs/2510.13169v1.

Deliang Chen. The Machado–Bishop theorem in the uniform topology. *Journal of Approximation Theory*, 304:106085, 2024. doi: 10.1016/j.jat.2024.106085. URL https://arxiv.org/abs/2407.19925. Comparison uses arXiv:2407.19925, Corollary 2.10.

Pim de Haan, Taco Cohen, and Max Welling. Natural graph networks. In *Advances in Neural Information Processing Systems*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/2517756c5a9be6ac007fe9bb7fb92611-Abstract.html.

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

Bowen Jing, Stephan Eismann, Patricia Suriana, Raphael John Lamarre Townshend, and Ron Dror. Learning from protein structure with geometric vector perceptrons. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2009.01411.

Yoshihiro Maruyama. Categorical equivariant deep learning: Category-equivariant neural networks and universal approximation theorems, 2025. URL https://arxiv.org/abs/2511.18417v2.

Yoshihiro Maruyama. Foundations of equivariant deep learning: Unifying graph and sheaf neural networks, 2026. URL https://arxiv.org/abs/2607.03798v4. ICML 2026 paper; comparison uses arXiv version 4, August 2026.

Ryo Nasu and Yoshihiro Maruyama. Mathematical foundations of monoid-equivariant neural networks. In *Pattern Recognition, ICPR 2026*, volume 16826 of *Lecture Notes in Computer Science*, pp. 31–47. Springer, 2027. doi: 10.1007/978-3-032-31933-3_3. URL https://link.springer.com/chapter/10.1007/978-3-032-31933-3_3. First published online 3 August 2026.

Marco Pacini, Gabriele Santin, Bruno Lepri, and Shubhendu Trivedi. On universality classes of equivariant networks. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://arxiv.org/abs/2506.02293v2. Comparison uses arXiv version 2, February 2026.

Marco Pacini, Mircea Petrache, Bruno Lepri, Shubhendu Trivedi, and Robin Walters. On universality of deep equivariant networks. In *International Conference on Learning Representations*, 2026. URL https://arxiv.org/abs/2510.15814v2. Comparison uses arXiv version 2, June 2026.

Sejun Park, Chulhee Yun, Jaeho Lee, and Jinwoo Shin. Minimum width for universal approximation. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=O-XJwyoIF-k.

Siamak Ravanbakhsh. Universal equivariant multilayer perceptrons. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 7996–8006, 2020. URL https://proceedings.mlr.press/v119/ravanbakhsh20a.html.

Zakhar Shumaylov, Peter Zaika, James Rowbottom, Ferdia Sherry, Melanie Weber, and Carola-Bibiane Schönlieb. Lie algebra canonicalization: Equivariant neural operators under arbitrary Lie groups. In *International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2410.02698.

Louis Solomon. Invariants of finite reflection groups. *Nagoya Mathematical Journal*, 22:57–64, 1963. doi: 10.1017/S0027763000011028.

Andrea Vezzosi, Anders Mörtberg, and Andreas Abel. Cubical Agda: A dependently typed programming language with univalence and higher inductive types. *Proceedings of the ACM on Programming Languages*, 3(ICFP):87:1–87:29, 2019. doi: 10.1145/3341691.

Soledad Villar, David W. Hogg, Kate Storey-Fisher, Weichi Yao, and Ben Blum-Smith. Scalars are universal: Equivariant machine learning, structured like classical physics. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://arxiv.org/abs/2106.06610.

Edward Wagstaff, Fabian B. Fuchs, Martin Engelcke, Michael A. Osborne, and Ingmar Posner. Universal approximation of functions on sets. *Journal of Machine Learning Research*, 23(151): 1–56, 2022. URL https://jmlr.org/papers/v23/21-0730.html.

Yinshuang Xu, Jiahui Lei, Edgar Dobriban, and Kostas Daniilidis. Unified Fourier-based kernel and nonlinearity design for equivariant networks on homogeneous spaces. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, 2022. URL https://proceedings.mlr.press/v162/xu22e/xu22e.pdf.

Dmitry Yarotsky. Universal approximations of invariant maps by neural networks. *Constructive Approximation*, 55:407–474, 2022. doi: 10.1007/s00365-021-09546-1. URL https://arxiv.org/abs/1804.10306.

A MATCHING THE FINITE CATEGORICAL ARCHITECTURE

This appendix specifies the correspondence used when applying Theorem 3.2 to the CENN generators of Maruyama (2025, Definitions 4–10). All base spaces are singletons and all arrow sets carry counting measure. Thus a feature section at a is simply a vector in $X(a)$, integrals are finite sums, and continuity is ordinary finite-dimensional continuity. The comparison concerns the integrated naturality condition in that model; it does not impose an additional, stronger, pointwise condition on each individual arrow kernel.

The augmented graph setting is also explicit in that source's Appendix B.3: its face category includes a least object, permits singleton bases, and allows zero bottom features. Thus the zero-bottom example is a stated model instance. Its failure of reconstruction places it outside the hypotheses of the source's conditional universality theorem.

Natural convolutions. A categorical convolution, including its natural bias, is a natural affine map and is therefore covered by Definition 2.1. In the reverse direction, suppose $L : X \to Y$ is natural and linear. At a, use L_a as the coefficient of the identity-arrow slot and zero as the coefficients of all other slots. The integrated operator is exactly L_a. Its required naturality equation becomes $Y(u)L_a = L_b X(u)$, which holds by assumption. Adding a natural section c gives $L + c$. This proves the reverse inclusion used for the exact affine class in Theorem 4.2.

Scalar nonlinearities. A natural scalar channel lands in the constant scalar diagram. Because $\perp$ is least, any such channel is determined on a restriction fibre by its bottom component. The proof of Theorem 3.2 permits continuous scalar channels, so it includes the linear-channel special case. The gate is a scalar multiplier of its input vector. An unrestricted coordinatewise nonlinearity on an arbitrary hidden diagram is a different generator and is not included in the statement.

Bundle lifts and nonlinear maps. The finite form of the arrow bundle is (2). The componentwise nonlinearity applies the same map H_b to every slot whose arrow has source b, exactly as in (3). There are finitely many slots. Consequently globally Lipschitz component maps give a uniform bundle Lipschitz bound with respect to the maximum norm. Ordinary finite MLPs with a globally Lipschitz activation satisfy this condition. Polynomials such as x^3 serve as continuous targets, not as admissible componentwise MLP primitives.

Maps out of bundles and arbitrary nesting. For a natural linear $L : DU \to Z$, write $\iota_u : U(b) \to DU(a)$ for the injection into the slot indexed by $u : b \to a$. Finite direct-sum expansion gives

$$L_a(h) = \sum_{u:b\to a} (L_a \iota_u) h_u.$$

Thus every such map is a finite bundle convolution with the integrated naturality of L. The same applies to natural affine maps. Intermediate diagrams in this argument can be $D^r U$, their direct sums, or any other finite-dimensional diagrams. In a finite feed-forward expression, each node is a natural map from the original input. Keeping that input fixed is what lets the defect proof handle skip branches and arbitrary preceding maps into a bundle.

B COINDUCTION, INJECTIVITY, AND APPROXIMATION

B.1 THE ADJUNCTION, WITH BOTH INVERSE IDENTITIES

For a finite category over a field, coinduction gives

$$\mathrm{Nat}_{\mathrm{lin}}(Z, I_b(U)) \cong \mathrm{Hom}_k(Z(b), U). \tag{13}$$

For $u : b \to a$, the restriction formula in (1) is well typed: if $w : a \to c$, then $w \circ u : b \to c$ indexes a coordinate of $I_b(U)(c)$. Identity and composition laws follow from reindexing. For a natural linear $\eta : Z \to I_b(U)$ define

$$\epsilon(\eta) = \pi_{\mathrm{id}_b} \eta_b.$$

For $f : Z(b) \to U$ define

$$(\theta f)_a(z)_u = f(Z(u)z).$$

Contravariant functoriality proves that θf is natural: for $w : a \to c$ its u coordinate after restriction is $fZ(w \circ u)z = fZ(u)Z(w)z$. The identity coordinate gives $\epsilon\theta f = f$. For the other composite, naturality of η at $u : b \to a$ gives

$$\eta_a(z)_u = \pi_{\mathrm{id}_b}\eta_b(Z(u)z).$$

Hence $\theta\epsilon\eta = \eta$, proving (13). The construction is natural in both arguments. The same equations prove the continuous adjunction: each map is continuous because all component restrictions and finite products are continuous.

B.2 Why a reconstruction is equivalent to injectivity

The category of finite-dimensional diagrams on a finite category is abelian; kernels, cokernels, and exactness are computed objectwise. Evaluation $Z \mapsto Z(b)$ is exact. Since every vector space over a field is injective, $\mathrm{Hom}_k(-, U)$ takes short exact sequences to exact sequences with reversed arrows. Through (13), the same is true of $\mathrm{Nat}_{\mathrm{lin}}(-, I_b(U))$, so $I_b(U)$ is injective. Finite direct sums of these objects are injective.

The map Δ_Z is natural because $Z(u)Z(w) = Z(w \circ u)$, and it is monic because it contains the identity coordinate. Thus DY is an injective object containing Y. If a natural retraction R_Y exists, Y is an injective direct summand. If Y is injective, its identity map extends across Δ_Y to precisely such a retraction. This proves the algebraic part of Proposition 7.1.

For a finite poset, every injective diagram has the standard decomposition of the form $Y \cong \bigoplus_d J_d \otimes U_d$. Brown & Draganov (2021) use covariant sheaves and downset injectives; applying their convention to P^{op} gives the upset supports used here. In these coordinates $Y(a) = \bigoplus_{d \le a} U_d$. An explicit reconstruction selects the U_d component from the $Y(d)$ slot and places it in the corresponding summand of $Y(a)$. It must not include all coordinates of each $Y(d)$, since those slots can contain overlapping grades.

B.3 The compact approximation estimate

Let $F : X \to Y$ be continuous and natural and let $K_a \subset X(a)$ be compact for every object. Define

$$K_b^+ = \bigcup_{\substack{a \in C \\ u : b \to a}} X(u)K_a.$$

There are finitely many arrows, so K_b^+ is compact. The identity arrow ensures $K_b \subset K_b^+$. Choose ordinary MLPs $H_b : X(b) \to Y(b)$ with error below δ on K_b^+, using a fixed continuous globally Lipschitz nonpolynomial activation. Such finite MLPs are globally Lipschitz, and the standard Euclidean universal approximation property supplies the required approximation; this is also the nonequivariant step in the averaging argument of Yarotsky (2022).

For $x \in K_a$ and $u : b \to a$, naturality gives $F_b(X(u)x) = Y(u)F_a(x)$. Every coordinate of $H_\downarrow\Delta_X(x) - \Delta_Y F(x)$ therefore has norm below δ. With the maximum norm on the finite collection of slots,

$$\sup_{x \in K_a} \|(R_Y H_\downarrow \Delta_X)_a(x) - F_a(x)\| \le \|R_{Y,a}\|\,\delta.$$

Choose $\delta < \varepsilon/(1 + \max_a \|R_{Y,a}\|)$. This proves simultaneous approximation on all the specified compacts, including the case of zero-dimensional outputs.

B.4 The group case and variance

For a one-object group category use multiplication $gh = g \circ h$. A contravariant feature functor corresponds to the ordinary representation $X(g) = \rho_X(g^{-1})$. The arrow bundle has restriction $DY(w)(h)_g = h_{wg}$, and $\Delta_Y(y)_g = \rho_Y(g^{-1})y$. For a finite group with invertible order in the field, set

$$R_Y(h) = \frac{1}{|G|}\sum_g \rho_Y(g)h_g.$$

Then $R_Y \Delta_Y = \mathrm{id}$, and substituting $k = wg$ gives $R_Y DY(w) = \rho_Y(w^{-1}) R_Y$. Compiling a component map H and replacing g by g^{-1} produces

$$(R_Y H_\downarrow \Delta_X)(x) = \frac{1}{|G|} \sum_g \rho_Y(g)^{-1} H(\rho_X(g)x).$$

This is Reynolds symmetrization. The same argument componentwise on a finite groupoid reduces to its isotropy groups. It is the invertibility of their orders, not merely the existence of arrows, that is used by this averaging construction.

C Further examples and boundary cases

C.1 A three-object chain

On $\perp < a < t$, let $S(a) = \mathbb{R}$ and $S(\perp) = S(t) = 0$. A natural map $J_a \to S$ must vanish: its component at a equals the restriction of its zero component at t. A map $J_t \to S$ is also zero. Thus $\mathsf{T}(S) = 0$, and Theorem 3.2 makes every network affine on $S(a)$. Naturality at $a \leq t$ additionally forces the intercept to vanish. Conversely, every scalar linear map is a natural layer. The generated self-maps are exactly $x \mapsto bx$. The cubic is natural because it vanishes at zero, and the same calculation as in Theorem 4.2 shows that its best uniform approximation error is exactly $1/4$.

The finite checker computes $\mathrm{Nat}_{\mathrm{lin}}(D^r S, S) = 0$ for $r = 1, 2, 3$. These computations illustrate failed pure bundle compilation but do not prove the arbitrary-nesting theorem. In particular, adding an identity skip branch recovers all linear maps even though the pure bundle compiler spaces are zero.

C.2 An edge and a quotient of an injective

Let K be the rank-one edge diagram in Figure 1. Natural continuous self-maps are arbitrary common functions $f : \mathbb{R} \to \mathbb{R}$, while the generated self-maps are exactly common affine maps. For $f(x) = x^2$, any affine approximant with uniform error ε satisfies

$$1 = \left| f(0) - \frac{f(-1) + f(1)}{2} \right| \leq 2\varepsilon.$$

The constant $1/2$ attains the bound, so the best uniform approximation error is exactly $1/2$.

Let $S_\perp$ be supported only at $\perp$. There is an objectwise exact sequence of diagrams

$$0 \longrightarrow S_\perp \longrightarrow J_\perp \longrightarrow K \longrightarrow 0.$$

Thus a quotient of an injective by a bottom-supported subfunctor need not have a dense generated class. An independent global feature cannot be created naturally from this input: by the continuous adjunction,

$$\mathrm{Nat}_{\mathrm{cts}}(K, J_\perp \otimes U) \cong C(K(\perp), U) = C(\{0\}, U).$$

Every such natural map is constant across the input.

C.3 Nonzero bottom stalks and the role of assumptions

For any diagram Y, the trace of $J_\perp \otimes U$ is zero by the adjunction, and Lemma 3.1 gives

$$\mathsf{T}((J_\perp \otimes U) \oplus Y) = \mathsf{T}(Y).$$

Thus adding an independent constant summand to the graph example preserves zero trace because that example has $\mathsf{T}(Y) = 0$. Restriction to the bottom forgets the local coordinate and retains only the U coordinate. On each fibre the affine barrier still applies. The natural target $(v, z) \mapsto (v, \|z\|^2 z)$ has best uniform error $1/4$ on the local unit ball with $v = 0$, achieved by $(v, z) \mapsto (v, 3z/4)$.

If an isolated vertex v is present, a map $J_v \to Y$ can carry a nonzero vector at v without an incident edge forcing it to vanish. If an incidence restriction is singular, the other-endpoint argument can

leave a nonzero kernel direction. Both cases give explicit nonzero traces in the checker. Orthogonality ensures that the radial cubic is natural for the chosen inner products. If one keeps standard coordinates with the non-isometric rank-one restriction $r(x) = 2x$, then

$$(rx)^3 - r(x^3) = 6x^3.$$

Thus the same coordinate cubic is not natural in those coordinates; this illustrates dependence on the chosen inner products rather than a necessity of orthogonality as an abstract property. Finally, without a least object, scalar channels need not be determined by a single bottom value. For example, norms give nonconstant natural scalar channels on an orthogonal local system. Such a diagram is outside Theorem 3.2.

C.4 Holonomy and the finite averaging repair

Remove the bottom object before forming tree transports. Inverting the zero restrictions to the bottom would collapse the nonzero feature data. On the remaining incidence graph, the restrictions are isomorphisms and can be inverted along paths. The spanning tree gives identifications τ_a; the transport equation along every off-tree incidence is equivalent to commutation with $h_{da} = \tau_d^{-1} r_{da}^Y \tau_a$. This proves the bijection in (6) and also shows that a different reference frame conjugates its coordinate description. No preferred frame is part of the intrinsic natural map.

Suppose H is finite. For $k \in H$, reindexing $g = hk$ proves

$$f_H(kz) = \frac{1}{|H|} \sum_h h^{-1} m(hkz) = k \frac{1}{|H|} \sum_g g^{-1} m(gz) = k f_H(z).$$

For a Lipschitz constant L of m, orthogonality gives a Lipschitz constant at most L for f_H and every transported layer. Given objectwise compacts K_a, set $K = \bigcup_{a \neq \perp} \tau_a^{-1} K_a$ and $K^H = \bigcup_{h \in H} hK$. For $f \in C_H(V, V)$ and $\|m - f\| < \varepsilon$ on K^H,

$$\|f_H(z) - f(z)\| \leq \frac{1}{|H|} \sum_h \left\| h^{-1}(m(hz) - f(hz)) \right\| < \varepsilon \qquad (z \in K).$$

This proves density after transport, using the known averaging argument. For infinite $H \subset O(V)$, continuity extends equivariance to its compact closure. Haar integration supplies the corresponding mathematical approximation (Yarotsky, 2022). Exact realization by a prescribed finite network requires a further argument and is not claimed.

D The monoid reconstruction comparison

D.1 What follows from the two source assumptions

The hypotheses used here are the coefficient covariance A2 and Sandwich Solvability on printed pages 33–35 of Nasu & Maruyama (2027). The mean acts entrywise on bounded matrix-valued functions. For a fixed s, uniform bounds on L and ρ make $m \mapsto L(ms) Q_* \rho(m)$ bounded. Multiplying its mean by $\rho(s)$ and using A2 gives

$$\rho(s) \mu_m[L(ms) Q_* \rho(m)] = \mu_m[L(m) Q_* \rho(m)] = I_V.$$

Hence $\rho(s)$ is surjective and therefore invertible. If $\|L(m)\| \leq C_L$ and $\|\rho(m)\| \leq C_\rho$, positivity and normalization of the mean also give $\left\|\rho(s)^{-1}\right\| \leq C_L \|Q_*\| C_\rho$.

There is an equivalent image argument. For every m, s, A2 implies $\operatorname{im} L(m) \subseteq \operatorname{im} \rho(s)$. Every vector obtained by the averaging reconstruction belongs to this finite-dimensional closed subspace. An identity reconstruction on all of V therefore forces $\operatorname{im} \rho(s) = V$. The identity required by Sandwich Solvability is on the ambient output space, which can be stronger than reconstruction only on the range of a particular target.

For finite M, the image of ρ is a finite submonoid of $GL(V)$ and hence a group: a positive power of every image element is the identity. If M has an absorbing zero, its image is an invertible idempotent and therefore the identity; the equation $\rho(0)\rho(m) = \rho(0)$ then forces the entire output representation to be trivial. These consequences concern the output representation under the combined assumptions. They leave open noninvertible input actions and do not challenge the validity of the conditional approximation theorem.

D.2 Finite reconstruction and the opposite-category convention

To regard an ordinary representation $\rho : M \to \mathrm{End}(V)$ as a contravariant feature functor, take $C = M^{\mathrm{op}}$. Then the arrow bundle acts by right reindexing, $D(s)(h)_m = h_{ms}$, and $\Delta(y)_m = \rho(m)y$. For $R(h) = \mu_m[L(m)Q_*h_m]$, apply right invariance to the bounded function $m \mapsto L(m)Q_*h_m$, then use A2, to obtain

$$\rho(s)R(h) = \mu_m[\rho(s)L(ms)Q_*h_{ms}] = \mu_m[L(m)Q_*h_{ms}] = RD(s)(h).$$

The first displayed equality in this last line uses right invariance on $m \mapsto L(m)Q_*h_m$ before multiplying by $\rho(s)$. The identity condition gives $R\Delta = I_V$. Thus for finite M the special averaging assumptions imply the categorical compiler condition.

D.3 A singular idempotent with a compiler

For $M = \{1, e\}$, $e^2 = e$, the monoid is commutative, so the variance distinction does not affect the explicit calculation. Let $E^2 = E$. Define

$$D(e)(h_1, h_e) = (h_e, h_e), \quad \Delta(y) = (y, Ey), \quad R(h_1, h_e) = (I - E)h_1 + (2E - I)h_e. \tag{14}$$

This reconstruction satisfies

$$\begin{aligned} R(y, Ey) &= (I - E)y + (2E - I)Ey = y, \\ RD(e)(h_1, h_e) &= Eh_e, \\ ER(h_1, h_e) &= E(I - E)h_1 + E(2E - I)h_e = Eh_e. \end{aligned}$$

The algebraic reconstruction works for every idempotent, including zero, singular orthogonal projections, and singular nonorthogonal projections. It also agrees with the semisimplicity of $k[M] \cong k \times k$.

For $E = \mathrm{diag}(1, 0)$, the unique normalized right-invariant mean on this finite monoid is evaluation at e. A2 gives $L(1) = L(e) = EL(e)$, so

$$L(e) = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix}, \qquad L(e)Q_*E = \begin{pmatrix} aq_{11} + bq_{21} & 0 \\ 0 & 0 \end{pmatrix}.$$

Here q_{ij} are the entries of Q_*. The Sandwich Solvability expression cannot equal the identity. The categorical compiler nonetheless exists by the preceding equations.

For any ordinary map $H : \mathbb{R}^2 \to \mathbb{R}^2$,

$$\Phi_H(x) = (I - E)H(x) + (2E - I)H(Ex)$$

commutes with E because $\Phi_H(Ex) = EH(Ex) = E\Phi_H(x)$. If F is equivariant, the same expression with $H = F$ equals F. When H approximates F to error ε on $K \cup EK$, the orthogonal choice $E = \mathrm{diag}(1, 0)$ has Euclidean operator norms $\|I - E\| = \|2E - I\| = 1$, giving error at most 2ε on K. For a nonorthogonal idempotent, the corresponding bound is $(\|I - E\| + \|2E - I\|)\varepsilon$. For a concrete globally Lipschitz nonlinear example, take

$$H(x, y) = (\tanh(x + y), \tanh(x + 2y)).$$

Then

$$\Phi_H(x, y) = (\tanh x,\ \tanh(x + 2y) - \tanh x),$$

which is non-affine and commutes with the singular projection.

E Dependent kernels: structure and formal scope

E.1 Biproduct kernels and transport

Let B be a 1-type, let $P, Q : B \to \mathrm{FinSet}$ be site families, and let $V_{b,p}, W_{b,q}$ be vector-space families. Put $X_b = \bigoplus_{p:P(b)} V_{b,p}$ and $Y_b = \bigoplus_{q:Q(b)} W_{b,q}$. The common pair base is

$$D_{Q,P} = \sum_{b:B} (Q(b) \times P(b)) \simeq \left(\sum_b Q(b) \right) \times_B \left(\sum_b P(b) \right).$$

The fibre product makes source and target sites share the same symmetry.

Proposition E.1 (Dependent block kernels). *There is a linear equivalence*

$$\prod_{b:B} \mathrm{Hom}_k(X_b, Y_b) \simeq \prod_{b:B} \prod_{q:Q(b)} \prod_{p:P(b)} \mathrm{Hom}_k(V_{b,p}, W_{b,q}), \tag{15}$$

which preserves identity and composition. Extraction and assembly are

$$K_{b,q,p} = \pi_q L_b \iota_p, \qquad (L_b x)_q = \sum_p K_{b,q,p} x_p.$$

For a third site family $R : B \to \mathrm{FinSet}$ with fibres $U_{b,r}$, put $Z_b = \bigoplus_{r:R(b)} U_{b,r}$. If M is the block kernel of a map $Y \to Z$, composition is $(MK)_{b,r,p} = \sum_{q:Q(b)} M_{b,r,q} K_{b,q,p}$.

The finite direct sum $\bigoplus_p V_p$ has injections ι_p and projections π_p, with $\sum_p \iota_p \pi_p = \mathrm{id}$, $\pi_p \iota_p = \mathrm{id}$, and $\pi_q \iota_p = 0$ for $p \neq q$. For $K_{q,p} : V_p \to W_q$, assembly gives $L = \sum_{q,p} \iota_q K_{q,p} \pi_p$. Extraction of this sum returns $K_{q,p}$ by the injection/projection identities. Conversely,

$$\sum_{q,p} \iota_q (\pi_q L \iota_p) \pi_p = \left(\sum_q \iota_q \pi_q \right) L \left(\sum_p \iota_p \pi_p \right) = L.$$

Both operations preserve addition and scalar multiplication. Inserting the middle biproduct identity into a composite gives

$$\pi_r M L \iota_p = \sum_q (\pi_r M \iota_q)(\pi_q L \iota_p).$$

These equations are invariant under finite relabelling and linear isomorphisms of the summands. They therefore apply to families over B and yield the section equivalence of Proposition E.1.

For a set-valued family A over a connected 1-type with base point b_0, a section determines an element of $A(b_0)$ fixed by transport along every loop. Conversely, transport of such an element to any point is independent of the chosen path, because changing paths acts by a loop. The fibre being a set ensures that the coherence proofs carry no further higher-dimensional data. This is the section interpretation used in Bezem et al. (2026). On finitely many components one obtains a product of fixed-point sets after choosing component representatives.

For $B = BG$, the resulting orbit/stabilizer parameterization is

$$\mathrm{Hom}_G(X, Y) \cong \prod_O \mathrm{Hom}_{H_O}(V_{p_O}, W_{q_O}), \qquad \dim \mathrm{Hom}_G(X, Y) = \sum_O \dim \mathrm{Hom}_{H_O}(V_{p_O}, W_{q_O}). \tag{16}$$

For the pair action groupoid of a finite group, the loops at (q, p) are the simultaneous stabilizer $H_{q,p}$. Transport on the Hom fibre is $A \mapsto \rho_W(h) A \rho_V(h)^{-1}$, whose fixed elements are the intertwiners. This proves (16). Reynolds averaging is a retraction to this fixed subspace when the stabilizer order is invertible. Finite-dimensionality supplies the dimension sum. A constructive extraction of orbit coordinates must also specify the finite presentation and the representative selection; the intrinsic section theorem does not need representatives.

E.2 Biases, masks, and changes of presentation

Biases are sections of Y and parallel branches use biproducts of families. A dependent decidable relation $R_b(q, p)$ defines kernels supported on its allowed pairs, with zero extension elsewhere. If K is supported on R and M on S, the composition formula gives support contained in

$$(S \circ R)_b(r, p) \iff \exists q,\ S_b(r, q) \text{ and } R_b(q, p).$$

If no such q exists, each block product is zero. This is an upper bound on support; cancellation can make the support smaller, and it is not a universality theorem for local layers.

For a base equivalence $e : B' \simeq B$, pullback gives an equivalence of dependent section spaces. A quasi-inverse and its homotopies give the inverse operation. Since extraction, assembly, and composition use transported biproduct structure, they commute with this pullback. The statement concerns intrinsic spaces of layers, not equivalence of particular parameter coordinates or optimization procedures. Ordinary HoTT can encode noninvertible categories explicitly; identity paths themselves supply only invertible transport. Likewise Euclidean continuity needs explicit topological structure.

E.3 WHAT IS MACHINE-CHECKED

The project has thirteen Agda modules and uses the existing Cubical library. The algebraic development works with lawful `LeftModule` structures and bundled module homomorphisms, rather than postulating the biproduct identities or applying them to unstructured functions. Finite products serve as finite sums; the module fibres need not be finite-dimensional or carry a basis. The formalized arrow bundle is the finite-poset special case of (1)–(2), where each hom-set is a proposition and therefore contributes at most one slot. The following table identifies representative declarations. Paths are relative to `formalization/Equivariant/`.

Table 3: Checked declarations. The full module-qualified names and source hashes are recorded in `formalization/result.json`.

Mathematical statement	File and declaration
Section transport coherence	`Core.agda`: `section-coherence`
Sections over the total pair type	`Core.agda`: `PairSections.pair-section-equiv`
Classifying-type sections as loop-fixed data	`Core.agda`: `FixedPoints.section-fixed-equiv`
Reconstruction from finite summands	`LinearAlgebra.agda`: `Algebra.reconstruct`
Both kernel inverse identities	`Kernel.agda`: `Blocks.kernel-equiv`
Evaluation of the assembled kernel	`Kernel.agda`: `Blocks.assemble-eval`
Composition and block multiplication	`Kernel.agda`: `Composition.extract-compose`, `assemble-multiply`
Identity, diagonal, off-diagonal	`Kernel.agda`: `Identity.assemble-identity`, `identity-diagonal`, `identity-off-diagonal`
Dependent labelled kernel sections	`Kernel.agda`: `Dependent.total-kernel-equiv`
Addition and scalar preservation	`Kernel.agda`: `Blocks.extract-add`; `KernelLinear.agda`: `LinearBlocks.extract-scale`, `assemble-add`, `assemble-scale`
Trace under linear maps	`Trace.agda`: `Traces.linear-preserves-trace`
Bundle trace equals bottom-slot kernel	`Trace.agda`: `Traces.bundle-trace-to-zero`, `bundle-zero-to-trace`
Finite-sum trace identity	`Trace.agda`: `Traces.sum-trace-to-components`, `components-to-sum-trace`
Network substitution semantics	`TraceNetwork.agda`: `Networks.evaluate-plug`
Trace-valued network affine defect	`TraceNetwork.agda`: `Networks.Weighted.network-affine-barrier`
Pointwise direction bound	`Direction.agda`: `Directions.network-direction`
Original generator embedding preserves evaluation	`DirectionEmbedding.agda`: `Embedding.evaluate-embed`

For example, the implementation of extraction is the following literal source equation, under the parameterized module that defines the source and target module families:

```
extract f q p = compose (project W q) (compose f (inject V p))
```

The inverse identities are equalities of bundled homomorphisms, including their linearity laws. Scalar linearity is proved over a commutative ring by preservation equations. The Hom spaces have not additionally been packaged into a single Cubical module-equivalence record.

The dependent formal theorem permits varying module fibres over an arbitrary base but uses the globally labelled site families $\mathrm{Fin}(n(b))$ and $\mathrm{Fin}(m(b))$. This excludes general permutation monodromy of the sites. The classifying-type fixed-point theorem is checked separately; its combination with arbitrary unlabelled finite-site kernels and orbit coordinate extraction remains to be formalized. The trace development uses a finite decidable poset with a chosen least object and lawful module-valued diagrams over a small commutative ring. `TraceDiagram.agda` constructs coinduction,

the arrow bundle, its bottom projection, natural component lifts, and finite sums. `Trace.agda` defines trace membership as the propositional truncation of finite witnesses built from zero, images of maps from non-bottom J_d, addition, scalar multiplication, and equality. This constructs the actual span; its bundle identity is proved rather than assumed.

`Affine.agda` proves identities for weighted affine defects. `TraceNetwork.agda` defines a typed finite network syntax with arbitrary hidden diagrams and proves naturality of its evaluation. Substitution represents composition and has a checked semantic equation. For any scalars $s+t=1$, structural induction proves trace membership of the defect on each bottom-restriction fibre. The bundle case proves the stronger fact that any natural map into a bundle has trace-valued defect. The syntax allows arbitrary component functions; the paper's continuous and Lipschitz choices are a subclass.

`Direction.agda` uses the same diagram infrastructure on finite decidable posets without assuming a least object. It constructs a global natural section extending any chosen bundle stalk vector (`bundleSection-at`). A second typed syntax admits every natural scalar rescaling. The theorem `network-direction` constructs a linear natural map L and natural bias c with $L_a x + c_a = N_a(x)$ for each network, object, and input. The witnesses depend on (a, x); this is membership in (17), not global affinity. Its affine, scalar, bundle, and parallel cases, and substitution semantics, are proved. `DirectionEmbedding.agda` embeds the original syntax and proves equality of its objectwise evaluation with the enlarged syntax.

`Multiplicity.agda` extends the finite kernel development to repeated modules equipped with a family of linear operators. The declaration `Repeated.Blocks.block-equiv` identifies intertwiners with matrices of commuting endomorphisms, including both inverse equations and equivariance laws. `reach-to-span` and `span-to-reach` prove the two inclusions of the pointwise image. `pad-collapse`, `padded-obstruction`, and `full-span-reaches` check zero-padding persistence and arbitrary output multiplicity. No Schur classification or rank theorem is assumed in these algebraic statements.

`Regularization.agda` proves covariance of inverses in the `Ridge` module, under the name `inverse-intertwines`. Its `Approximation` submodule checks image membership and residual identities, including `residual-identity` and `persistent-kernel`. Inverses are explicit hypotheses, and the backward map is arbitrary; the real application chooses $B^\top$. These proofs check the ridge algebra used in Appendix J. They do not establish inverse existence, positive definiteness, spectral bounds, or convergence. The full multiplicity threshold and exact-error theorems remain analytic pen-and-paper results.

The algebraic inductions do not require finite-dimensional modules. The real quotient topology and compact-open closure argument, the graph trace calculation, and the error bounds have not been formalized. Neither have masks, reconstruction, the monoid propositions, or Reynolds density, the analytic regularized coefficient lemma, or the stabilizer/commutant criterion. No machine-checked real-number instantiation is included. The artifact is supporting evidence at its declared scope, not a formal verification of the complete paper.

E.4 Pinned environment and reproducible checking

The successful environment is Agda 2.6.4.3 and Cubical commit

`d69d74c8da03c70325718f96d578d51b622e1dbf`

from the `agda/cubical` repository, corresponding to the Cubical 0.7 development. All project modules specify `--cubical --safe`. The checker validates the compiler version, dependency commit, and absence of dependency-source modifications; it then copies sources into a fresh temporary directory and ignores global library settings. It does not reuse compiled dependency interfaces or download dependencies.

From the artifact root, after installing the pinned compiler and checking out the dependency at that commit, run:

```
python formalization/verify.py --cubical /path/to/cubical \
  --output formalization/result.json
```

The checker imports `Equivariant.Everything`, which imports every project module. The result contains the environment, checked declaration names, and SHA-256 hashes of the project Agda sources. A clean build can take several minutes.

F Computational verification protocol

F.1 Exact algebra

The verification scripts represent a finite diagram by its stalk dimensions and restriction matrices. A natural transformation is found by introducing one unknown matrix per object, imposing every equation $Y(u)L_a - L_b X(u) = 0$, and solving the resulting rational linear system. For $\mathsf{T}(Y)(a)$, the code evaluates bases of $\mathrm{Nat}_{\mathrm{lin}}(J_d, Y)$ at a, concatenates their images, and computes the column span. Compiler existence adds the linear equations $R\Delta = I$ to the naturality equations.

Table 4: Exact rational calculations. The trace entry is zero at every object. In the quarter-turn example the commutant consists of matrices $aI + bJ$, where J is rotation through $\pi/2$.

Diagram	Rank	$\dim \mathrm{End}_H(V)$	$\dim V^H$	$\dim \mathsf{T}(Y)(a)$
Single edge, identity restrictions	1	1	1	0
Triangle, sign holonomy	1	1	0	0
Triangle, quarter-turn holonomy	2	2	0	0

Additional exact checks include the following.

- A C_2 action on three sites with two-dimensional fibres gives a global equivariant matrix space of dimension 18. The five pair-orbit contributions are $2, 4, 4, 4, 4$. A star support mask gives dimension 14; composition and its support bound are checked by exact matrices.
- The middle simple of a chain has no reconstruction, whereas an injective control has one. Three successive pure bundle compiler spaces into that simple vanish, and an identity skip branch recovers its one-dimensional linear endomorphism space.
- On a twice-nested edge bundle, a natural input map followed by coordinate clipping produces a nonzero midpoint defect. Clipping is globally Lipschitz and is evaluated with rational inputs. Every map in a seven-dimensional basis of natural compilers sends the defect into the output trace; four basis images are nonzero.
- Isolated-vertex and singular-incidence examples have nonzero trace. The nonorthogonal cubic defect is the polynomial $6x^3$.
- The idempotent monoid calculations verify a right-invariant mean, failure of the special identity reconstruction for a singular output, and the general natural reconstruction for zero, orthogonal singular, and nonorthogonal singular idempotents.

These are exact finite certificates. The universally quantified statements about all diagrams and all network depths follow from the proofs, not from these computations.

F.2 Deterministic numerical illustrations

The illustrations in this subsection use no training, data split, or stochastic optimizer. The separate learning task is in Appendix G. For $h > 0$, define the odd finite-difference map

$$m_h(x) = \frac{2\tanh(hx) - \tanh(2hx)}{2h^3}.$$

For the cubic illustration, evaluate m_h and x^3 on $x_i = -1 + 2i/2000$, $i = 0, \ldots, 2000$. Report the maximum absolute difference. The values are:

h	Maximum error on the 2,001-point grid
0.20	0.0748951128
0.10	0.0196654875
0.05	0.0049788368

The map is odd, hence equivariant for rank-one signed holonomy. Since $\tanh t = t - t^3/3 + 2t^5/15 + O(t^7)$,

$$m_h(x) = x^3 - 2h^2x^5 + O(h^4)$$

uniformly for x in a fixed bounded interval as $h \to 0$. This expansion explains convergence but the table itself is only a sampled error measurement. The explicit output weights grow as h^{-3}, so this construction does not establish good conditioning or ease of training.

Other floating-point checks compare a compiled finite-group map with its Reynolds formula and test a quarter-turn equivariant nonlinear layer. Residuals near machine precision check the implementations; they do not replace the symbolic equivariance proofs.

F.3 Commands and outputs

From the artifact root, run:

```
python verification/check_research.py \
  --output verification/results.json
python verification/check_audit.py \
  --output verification/audit-results.json
python verification/check_scalar_design.py \
  --output verification/scalar-design-results.json
python verification/check_breadth.py \
  --output verification/breadth-results.json
python verification/check_correspondence.py \
  --output verification/correspondence-results.json
python verification/make_figures.py --output reproduced/figures
python verification/make_breadth_figure.py \
  --output reproduced/figures/breadth.pdf
```

The recorded algebra environment is Python 3.13.5, NumPy 2.3.3, and SymPy 1.14.0. The JSON files record the actual versions used. Figure generation additionally uses Matplotlib and emits a vector PDF. The package includes a dependency list, a top-level reproduction script, and a manifest of source hashes. It contains the Python and Agda sources needed to rerun every check; the manuscript and its LaTeX sources are submitted separately and are not part of it.

G Response-learning protocol and population error

G.1 Diagram, target, and loss

The augmented triangle has vertices v_0, v_1, v_2, edges $e_0 = (v_0, v_1), e_1 = (v_1, v_2), e_2 = (v_2, v_0)$, and a zero bottom stalk. Every non-bottom stalk is $\mathbb{R}^2$. In reference coordinates each incidence is the identity except $r_{v_0e_2} = J$, rotation through $\pi/2$. The code uses fixed local frames τ_a with angles $(0, 0.31, -0.47, 0.83, -0.22, 0.64)$ for $(v_0, v_1, v_2, e_0, e_1, e_2)$, respectively. Thus the actual restrictions are $\tau_v h_{ve} \tau_e^{-1}$, with $h_{ve} = J$ at the exceptional incidence and I otherwise. The predicted local map is $\tau_a f \tau_a^{-1}$. All six incidence equations, including the cycle-closing equation, are evaluated numerically. These orthogonal frames leave the reference-stalk loss unchanged.

The target is (9) for $\beta \in \{0, 1/4\}$. The added term is the complex conjugate cubic: in polar coordinates it is $r^3(\cos 3\theta, -\sin 3\theta)$, whereas the radial target is $r^3(\cos\theta, \sin\theta)$. This proves C_4 equivariance. For $\beta \neq 0$ the response is not equivariant under all rotations. The polynomial energy interpretation motivates a nonlinear anisotropic response; there is no measured physical dataset in this experiment.

Write $z = r(\cos\theta, \sin\theta)$. Uniform disk sampling uses independent $\theta \sim U[0, 2\pi)$ and $r = \sqrt{U}$ with $U \sim U[0, 1]$. The anisotropic term is angularly orthogonal to z, Jz, and the radial cubic, and has squared norm r^6. Consequently

$$\begin{aligned}\mathbb{E}\,\|F_\beta(z) - az - bJz\|^2 &= (1+\beta^2)\mathbb{E}r^6 - 2a\mathbb{E}r^4 + (a^2+b^2)\mathbb{E}r^2 \\ &= \frac{1+\beta^2}{4} - \frac{2a}{3} + \frac{a^2+b^2}{2}.\end{aligned}$$

Completing the square gives (10). The risk is mean squared Euclidean vector error, with no division by the two output coordinates. Its square root gives population RMSE $1/6$ and $5/24$. The audit script also derives the formula by symbolic integration in polar coordinates.

G.2 Models, optimization, and selection

The affine oracle fits $az + bJz$ by least squares on the training set. It represents the entire reachable class in Theorem 4.2. The trained original stack has three complex-linear layers, represented over $\mathbb{R}$, with multiplicities $1 \to 24 \to 24 \to 1$. A linear block is $a_{oc}I + b_{oc}J$, where o, c index output and input channels. Between layers are two learned gates $\tanh(\gamma_i)$; they are input-independent, as required by naturality to the constant scalar diagram when the bottom stalk is zero. The stack has $1{,}250$ parameters. No nonzero output bias commutes with C_4. The theorem and oracle cover all permitted hidden diagrams and bundles; the trained stack is one concrete member of that class.

The ordinary MLP has widths $2 \to 32 \to 32 \to 2$, biases, and tanh hidden activations, for $1{,}218$ parameters. The Reynolds MLP averages this map over the four rotations as in (8). The unrestricted control uses the same architecture without averaging and is not constrained to be natural. The two MLPs share initial base weights within each seed. Parameter counts are comparable, not exactly matched. The Reynolds layer changes the generators and preserves the input/output feature functor.

The fixed seeds are $11, 23, 37, 53, 71$. Each gives $1{,}024$ training and $1{,}024$ independent validation points. A separate test set of $20{,}000$ points uses seed 900001 and is shared across all runs. Training uses Adam with learning rate 0.003, minibatch size 256, and $2{,}000$ updates. Models share the minibatch sequence within a seed. Validation is evaluated at steps $0, 100, \ldots, 2000$; the lowest-loss checkpoint is retained. All five main seeds are reported. A 500-step pilot preceded the fixed main protocol; test results are not used for checkpoint selection.

All runs use CPU, one thread, float64, and PyTorch deterministic algorithms. The recorded PyTorch build is 2.9.1+cu128, with CPU execution. Mean training times per seed are about 10.3 seconds for the original stack, 7.7 seconds for the Reynolds MLP, and 3.7 seconds for the unrestricted MLP on the execution machine. These timings include periodic validation and depend on these particular implementations; they are not a scaling comparison.

G.3 Recorded evidence and reproduction

The artifact records all 40 evaluations (ten affine fits and thirty trained models), learning histories, 30 selected checkpoints, parameter counts, training times, test errors, and naturality residuals. The reported test RMSEs are the mean and sample standard deviation across seeds, not confidence intervals over independent test datasets. The training-fitted oracle need not minimize loss on the finite test set; a validation-selected original stack can therefore have slightly smaller test error without contradicting the population bound.

Sampled maximum errors use a grid of 129 radii and 256 angles; they are not certified supremum norms. Equivariance and naturality are tested on $1{,}024$ held-out points. For the original and Reynolds models, residuals of all six incidence equations are below 1.3×10^{-15}. The unrestricted control's maximum residuals across seeds are 0.048 and 0.036 on the two tasks. The construction proves equivariance; these residuals check its implementation. The checkpoint audit reloads every selected state, verifies its validation loss against the recorded minimum, recomputes the affine fit, and reproduces all stored metrics with zero difference in the recorded environment. It also checks that every seed/model/task combination appears exactly once. This audit validates the archived computations; it does not rerun optimization.

With the additional PyTorch dependency installed, run from the artifact root:

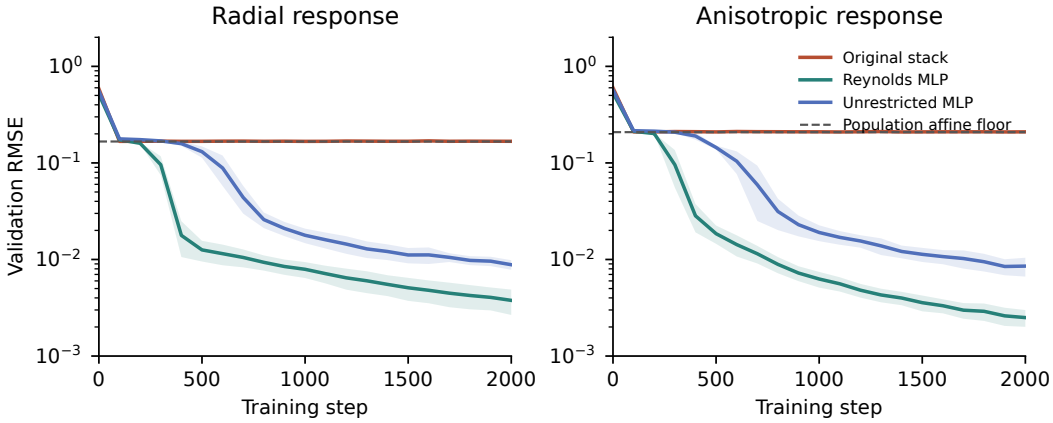


Figure 2: Validation vector RMSE over training, with mean and one standard deviation over five seeds. Dashed lines are the exact original-class population optima. Validation samples fluctuate around the population distribution; their error is not bounded below by that population optimum. Test results use the checkpoint chosen by validation, not necessarily the final training iterate.

```
python experiments/audit_checkpoints.py \
  --output reproduced/learning-audit.json
python experiments/summarize.py
python experiments/learn_response.py --output reproduced/learning
```

The first command checks the archived results without training. The second regenerates the summary, table, and vector figure. The last reruns the complete training protocol in a separate directory, leaving the archived evidence intact. The experiment demonstrates the expressivity mechanism on a fixed synthetic task; noisy data, larger groups, and applications with graph-configuration inputs require further study.

G.4 The D_4 design comparison

Table 2 uses a second diagram, chosen so that the missing direction of Section 6 is present rather than vacuous. Two triangles are glued at a common vertex, giving two independent cycles; the exceptional incidence transports are the quarter turn and a reflection, so the joint holonomy is D_4 and $\mathrm{End}_{D_4}(\mathbb{R}^2) = \mathbb{R}I$. Every non-bottom stalk is $\mathbb{R}^2$, incidence restrictions are orthogonal, and the bottom stalk is zero. The target is F_β of (9) with $\beta = 1/4$, which is D_4-equivariant because z and $p(z)$ are both D_4-covariants.

Three readouts are trained: $a(\|z\|^2)z$, with the coefficient network seeing only the norm; $a(\phi(z))z$, seeing both invariants; and $a(\phi(z))z + b(\phi(z))p(z)$, which adds the covariant feature. Each coefficient network is a two-hidden-layer tanh MLP of width 32 with an affine skip on the invariants. All three are D_4-equivariant and natural by construction, since the coefficients are invariant and $z, p(z)$ are covariant; the run confirms this, with equivariance and all twelve incidence residuals below 10^{-10}. The first row of the table is a least-squares fit of $z \mapsto az$, which is the entire original reachable class here, so it is a capacity bound rather than an optimization result. The oracle, norm-only, scalar-gated, and two-feature rows have 1, 1,155, 1,188, and 1,224 parameters, respectively.

The protocol is the one above: seeds $11, 23, 37, 53, 71$, 1,024 training and 1,024 validation points, a shared 20,000-point test set, Adam with learning rate 0.003, minibatch 256, 2,000 updates, and validation-selected checkpoints. Each measured mean lies within 0.0006 of the corresponding analytic infimum, except the two-feature row, whose infimum is zero and whose measured error is 0.0021. Individual seeds may fall marginally below a population infimum, since the test set is finite. The artifact stores all fifteen validation-selected neural checkpoints, the five oracle coefficients, learning histories, and metrics. `audit_d4_checkpoints.py` reloads the states, verifies the validation selections and parameter counts, and recomputes the test, grid, equivariance, and incidence

metrics. The comparison separates the three failure modes; it is not a claim that these readouts are preferable to any published architecture. From the artifact root, with PyTorch installed, run:

```
python experiments/learn_d4_design.py \
  --output reproduced/d4-design.json \
  --table reproduced/design-table.tex
python experiments/audit_d4_checkpoints.py \
  --results reproduced/d4-design.json \
  --output reproduced/d4-audit.json
```

G.5 Sampled maximum errors

Table 5 complements the vector RMSE results with the maximum Euclidean error observed on the fixed polar grid. Each run uses 129 radii and 256 angles, including the unit-circle boundary. These maxima are lower estimates of the true uniform error, not certified supremum norms. The models were selected by validation MSE, so they need not minimize uniform error: even for the radial target, the population MSE-optimal affine map is $\frac{2}{3}\mathrm{id}$, with uniform error $1/3$, whereas the minimax affine map is $\frac{3}{4}\mathrm{id}$, with error $1/4$.

Table 5: Sampled maximum errors on the unit disk, reported as mean $\pm$ sample standard deviation over five seeds. These values use the same fits and selected checkpoints as Tables 1 and 2.

Holonomy	β	Readout	Sampled maximum error
C_4	0	Affine oracle	0.3330 ± 0.0088
C_4	0	Original stack	0.3319 ± 0.0076
C_4	0	Reynolds MLP	0.0168 ± 0.0034
C_4	0	Unrestricted MLP	0.0417 ± 0.0025
C_4	0.25	Affine oracle	0.5839 ± 0.0134
C_4	0.25	Original stack	0.5839 ± 0.0133
C_4	0.25	Reynolds MLP	0.0166 ± 0.0026
C_4	0.25	Unrestricted MLP	0.0426 ± 0.0056
D_4	0.25	Affine oracle	0.5838 ± 0.0134
D_4	0.25	Norm-only gate	0.2608 ± 0.0067
D_4	0.25	Invariant-scalar gate	0.2501 ± 0.0003
D_4	0.25	Two-feature readout	0.0099 ± 0.0034

The script `experiments/summarize_grid_errors.py` generates this table from the archived metrics; both checkpoint audits recompute those metrics from the saved models.

H Configuration diagrams and the scalar-channel choice

These comparisons locate the obstruction more precisely. They do not extend it to arbitrary graph neural networks.

H.1 Independent node values on the same edge poset

On the augmented edge poset $\bot < v, w < e$, the rank-one local system in Figure 1 has dimensions $(0, 1, 1, 1)$ and identity incidence restrictions. In contrast, define a configuration diagram Z by

$$Z(\bot) = 0, \quad Z(v) = Z(w) = \mathbb{R}, \quad Z(e) = \mathbb{R}^2, \qquad r_{ve}(x, y) = x, \quad r_{we}(x, y) = y.$$

Then $Z = J_v \oplus J_w$, so $\mathsf{T}(Z) = Z$ and Z is injective. The affine quotient condition is vacuous on this diagram. An explicit reconstruction selects the vertex slots:

$$R_e(h_v, h_w, h_e) = (h_v, h_w), \quad R_v(h_v) = h_v, \quad R_w(h_w) = h_w.$$

It is natural and satisfies $R\Delta = \mathrm{id}_Z$. Taking $H_v = H_w = \tanh$ and $H_e = 0$ in the original bundle construction gives $(RH_{\downarrow}\Delta)_e(x, y) = (\tanh x, \tanh y)$. Thus zero bottom features and noninvertible arrows alone do not imply affine expressivity.

More generally, let P be the subset poset of a finite vertex set and $Z(U) = \bigoplus_{v \in U} V_v$, with coordinate restrictions. Then

$$Z \cong \bigoplus_v J_{\{v\}} \otimes V_v.$$

Every summand is a non-bottom elementary injective, so $\mathsf{T}(Z) = Z$ and linear reconstruction exists. This is a decomposition of the configuration presheaf, not an identification of a particular message-passing architecture.

H.2 Restriction compatibility and neighbor dependence

If a natural family $F_U : Z(U) \to Z(U)$ commutes with all restrictions to singleton subsets, then

$$p_v F_U(x) = F_{\{v\}}(x_v).$$

The output coordinate at v can depend only on x_v. Conversely, any fixed family of coordinatewise maps has this naturality. The condition permits nonlinear pointwise processing while excluding neighbor-dependent updates.

For example, on two scalar nodes let $L = \left(\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right)$. The diffusion step $B = I - L$ sends (x, y) to (y, x). The inputs $(0, 0)$ and $(0, 1)$ have the same first coordinate but distinct first outputs, so no singleton component can make B natural under coordinate deletion. Neural Sheaf Diffusion operates on the full node cochain space through a sheaf Laplacian and may learn feature-dependent restrictions (Bodnar et al., 2022). Those operations require a separate correspondence. Naturality with respect to graph or neighborhood isomorphisms (de Haan et al., 2020) is a different condition from commuting with every such deletion.

Bundle neural networks. Bundle neural networks (Bamberger et al., 2025) attach a copy of $\mathbb{R}^d$ to each node of a graph and transport features by orthogonal maps, so their feature object is a discrete vector bundle resembling the local systems of Section 4. Their bundles are flat: the maps are attached to nodes and transport factors as $O_v^\top O_u$, which the authors note gives path independence. The holonomy is therefore trivial, so the missing-direction criterion of Appendix I is vacuous for them. The affine barrier is not: with trivial holonomy Theorem 4.2 still gives exactly the affine self-maps, and the radial cubic still has best uniform error $1/4$. The difference lies instead in the operations. Their encoder step $x_v \mapsto O_v^\top W O_v x_v$ is a natural linear map in our sense, with $\tau_v = O_v^\top$ and $A = W$. Three further operations are outside Definition 2.1: the transports are learned from the graph and positional encodings rather than fixed, and in the general model depend on the current features; the diffusion step mixes features across all nodes through a bundle heat kernel, which the preceding paragraph excludes; and the activation is applied to stalk coordinates once the local frames are discarded, rather than as a natural map. Their approximation theorem further assumes injective positional encodings (Bamberger et al., 2025, Theorem 5.3), which separate the nodes. That universality result and Theorem 4.2 therefore concern different operation classes on similar objects, and do not conflict.

H.3 Zero extension preserves targets but changes scalar gates

Let Q be the non-bottom incidence category and Q^+ its augmentation by a least object. Extending a diagram by the zero vector space at that object gives a bijection between continuous natural-map spaces before and after extension: the extra component is unique and all new equations have zero target. Nevertheless, extending the scalar target as a *constant* scalar diagram puts $\mathbb{R}$ at the new bottom object. Naturality then makes every scalar channel on a zero-extended input constant on each full stalk.

To see the effect of this choice, retain the augmented orthogonal local system Y, but replace the scalar target by S_0, with $S_0(\bot) = 0$, $S_0(a) = \mathbb{R}$ at non-bottom objects, identity incidence restrictions, and zero restrictions to the bottom. With continuous scalar channels, $s_a(z) = \|z\|^2$, $s_\bot(0) = 0$ is natural to S_0. The resulting gate

$$G_a(z) = \tanh(\|z\|^2) z, \qquad G_\bot(0) = 0$$

commutes with the orthogonal restrictions and is nonlinear: its third derivative along a unit diameter at the origin is 6. It is excluded by the original constant-scalar type. This is a change of generators,

not a counterexample to Theorem 3.2, and it does not establish density for the modified gate class. An interpretation of the bottom as an empty region therefore needs an explicit choice of scalar data there.

The exact edge traces, reconstruction, nonlinear witness, diffusion restriction defect, and scalar-gate identities are checked by `verification/check_correspondence.py`; the recorded results are in `verification/correspondence-results.json`.

I SCALAR DIAGRAMS AND REACHABLE OUTPUT DIRECTIONS

Allowing an input-dependent scalar gate removes one obstruction to nonlinearity. It need not remove every obstruction to approximation. This section gives a necessary condition for arbitrary finite-poset diagrams and an exact criterion for finite-holonomy graph local systems. The positive construction uses established invariant-scalar approximation; the additional restriction comes from tracking which vector directions can pass through the categorical generators.

I.1 A BOUND THAT SURVIVES ARBITRARY NATURAL SCALAR GATES

Let P be a finite poset, and let all feature diagrams be contravariant diagrams of finite-dimensional real vector spaces. Write $J_d(a) = \mathbb{R}$ when $d \le a$, and $J_d(a) = 0$ otherwise, with the identity restrictions wherever both spaces are nonzero. The arrow bundle is $DZ = \bigoplus_d J_d \otimes Z(d)$. Its restrictions forget slots, and its lift is $\Delta_Z(z) = (r^Z_{da} z)_{d \le a}$. A component nonlinearity $DU \to DV$ applies a chosen globally Lipschitz map $H_d : U(d) \to V(d)$ to each slot labelled by d. Here a natural bias is a section $c = (c_a)_a$ with $r^Z_{da} c_a = c_d$. For a fixed input diagram X, an object a, and $x \in X(a)$, define the following vector subspace of $Z(a)$:

$$E_{X,Z}(a,x) = \{L_a x : L \in \mathrm{Nat}_{\mathrm{lin}}(X,Z)\} + \{c_a : c \text{ is a natural bias of } Z\}. \tag{17}$$

Both terms are linear subspaces. Equivalently, this is the set of values at x of all natural affine maps $X \to Z$. A least object is not needed for this bound.

Define $\mathcal{A}_{\mathrm{sc}}$ using natural affine maps, bundle lifts, component nonlinearities, and every continuous natural map of the form

$$G_a(z) = \lambda_a(z) z,$$

where the coefficients are scalar functions and the displayed family must be natural. The coefficients need not be natural to the constant scalar diagram. Arbitrary hidden diagrams, finite composition, and parallel branches are permitted. When a least object exists, this class contains the original constant-scalar generators. Gating separate natural direct summands is included through projections and parallel branches. This definition does not admit arbitrary coordinatewise activations on an undecomposed feature stalk.

Theorem I.1 (Reachable directions)**.** *Every $N \in \mathcal{A}_{\mathrm{sc}}(X,Z)$ satisfies $N_a(x) \in E_{X,Z}(a,x)$. The same condition holds in its compact-open closure (uniform convergence on every compact subset). In particular, any continuous natural target $F : X \to Z$ satisfies*

$$\|F_a(x) - N_a(x)\| \;\ge\; \mathrm{dist}\big(F_a(x), E_{X,Z}(a,x)\big).$$

Proof. Use structural induction on a network with fixed original input X. The input case uses $L = \mathrm{id}_X$. A natural linear map $M : Z \to W$ sends the two terms of (17) into the corresponding terms for W: L becomes ML, and a natural bias becomes Mc. Adding a natural bias preserves membership. A scalar gate preserves membership because $E_{X,Z}(a,x)$ is a vector subspace; no constancy of its coefficient along input fibres is used.

For a bundle target, the condition holds for every possible output value:

$$E_{X,DZ}(a,x) = DZ(a). \tag{18}$$

Indeed, for any $h = (h_d)_{d \le a} \in DZ(a)$, choose $u_d = h_d$ if $d \le a$, and $u_d = 0$ otherwise. The family $c_t = (u_d)_{d \le t}$ is a natural bias of DZ and satisfies $c_a = h$: bundle restrictions simply forget slots. Thus both a bundle lift and an arbitrary component nonlinearity have output in the required subspace, regardless of their input. Finally, $E_{X,\oplus_i Z_i}(a,x) = \oplus_i E_{X,Z_i}(a,x)$: maps and sections

can be chosen independently in each summand. This proves the parallel case and completes the induction. Each space in (17) is finite-dimensional and closed, so evaluation at a fixed (a, x) proves the closure statement and the distance bound. □

The bundle step explains why unrestricted hidden diagrams do not evade the result. Every vector in a bundle stalk extends to a natural bias of that bundle. A natural linear output map sends it to a direction already available from an output bias. Subsequent scalar rescaling and linear mixing preserve the subspace tracked by the induction. This observation uses the single slot per comparable pair in a poset; it does not establish the same bound for categories with multiple parallel arrows.

I.2 When the scalar diagram carries invariant information

Now take an augmented face poset of a finite connected loopless graph with at least one edge. Let X and Y have zero bottom stalks and orthogonal incidence restrictions. Fix a common spanning tree, reference spaces V, W, and isometric frames τ_a^X, τ_a^Y. Assume the joint holonomy image

$$H \subset O(V) \times O(W)$$

is finite. The word joint matters: the input and output actions come from the same graph loops. Tree transport identifies natural maps $X \to Y$ with H-equivariant maps $V \to W$. Consequently (17) at a reference stalk becomes

$$E(x) = \{Ax : A \in \mathrm{Hom}_H(V, W)\} + W^H. \tag{19}$$

Define $S_0(\bot) = 0$ and $S_0(a) = \mathbb{R}$ at every non-bottom object, with identity incidence restrictions and zero restrictions to the bottom. A continuous natural channel $q : X \to S_0$ is exactly a continuous H-invariant scalar function on V, transported by the chosen frames. Multiplication gives a natural bilinear action $S_0 \times X \to X$. At the bottom both feature and coefficient spaces are zero; no scalar action on a nonzero bottom feature is being asserted. In contrast, a natural channel $X \to J_\bot$ is constant, because $X(\bot) = 0$.

The invariant functions need not be supplied as unknown target-dependent oracles. For finite H, an ordinary scalar-output tanh MLP m gives

$$q_m(x) = \frac{1}{|H|} \sum_{h \in H} m(h_V x). \tag{20}$$

These channels are dense in the continuous invariant scalar functions on compact sets, by ordinary approximation followed by averaging (Yarotsky, 2022, Proposition 2.1).

I.3 An exact closure criterion

We first give the approximation lemma that will be used with the affine maps in (19). A covariant below means a continuous equivariant vector-valued map; it need not be polynomial.

Pointwise spanning alone does not provide continuous exact coefficients. For example, $F(v) = v/\sqrt{\|v\|}$ for $v \neq 0$, with $F(0) = 0$, is a continuous orthogonally equivariant map. Its unique coefficient relative to the single feature $B(v) = v$ diverges at zero. The set-theoretic and polynomial representations in Villar et al. (2021, Proposition 4 and Appendix B) are distinct statements; the former does not promise continuous coefficients. The following closure argument explicitly handles this distinction.

Lemma I.2 (Invariant coefficients and pointwise span)**.** *Let H be finite with orthogonal representations V, W, let $B_1, \dots, B_m : V \to W$ be continuous equivariant maps, and put $B(x) = [B_1(x) \ \cdots \ B_m(x)]$. For scalar channels (20), let each q_j below range over continuous H-invariant scalar functions (the averaged MLP channels are dense in this class). The compact-open closure of*

$$\mathcal{M}_B = \left\{ \sum_{j=1}^{m} q_j(x) B_j(x) \right\}$$

is precisely $\{F \in C_H(V, W) : F(x) \in \mathrm{im}\, B(x)$ for every $x\}$. This class equals $C_H(V, W)$ if and only if

$$\mathrm{im}\, B(x) = W^{H_x} \quad \text{for every } x \in V, \qquad H_x = \{h \in H : h_V x = x\}. \tag{21}$$

Proof. Pointwise membership and its preservation under limits are immediate. For the reverse inclusion, suppose $F(x) \in \operatorname{im} B(x)$. For $\epsilon > 0$, define regularized coefficients

$$c_\epsilon(x) = B(x)^\top \big(B(x)B(x)^\top + \epsilon I_W\big)^{-1} F(x).$$

They are continuous. Since $B(h_V x) = h_W B(x)$ and h_W is orthogonal, each coefficient is H-invariant. Write $A(x) = B(x)B(x)^\top$. The error is

$$F(x) - B(x)c_\epsilon(x) = \epsilon\big(A(x) + \epsilon I_W\big)^{-1} F(x).$$

At each x it tends to zero, since $F(x) \in \operatorname{im} A(x)$; its norm decreases as ϵ decreases, by diagonalizing the positive semidefinite matrix $A(x)$. These norms are continuous. On a fixed compact set, Dini's theorem applied to $\epsilon = 1/n$ therefore gives uniform convergence. For a fixed ϵ, approximate its finitely many invariant coefficients by (20). The columns of B are bounded on the compact set, so the resulting vector error is arbitrarily small. This argument includes points where the rank of B changes; it assumes neither a continuous pseudoinverse nor bounded exact coefficients.

Every equivariant value at x belongs to W^{H_x}, which proves sufficiency of (21). For necessity, take $w \in W^{H_x}$. Choose a continuous bump β centred at x, with $\beta(x) = 1$ and support disjoint from the other points of its finite orbit. The globally defined map

$$F(y) = \frac{1}{|H_x|} \sum_{h \in H} \beta(h_V^{-1} y) h_W w$$

is equivariant and satisfies $F(x) = w$. Thus failure to span any such value prevents density. This also covers an orbit consisting of a single point. □

Theorem I.3 (Scalar design criterion for local systems)**.** *Let Γ be a finite connected loopless graph with at least one edge, and let X, Y be zero-bottom orthogonal local systems on its augmented face poset, with finite joint holonomy $H \subset O(V) \times O(W)$ in isometric tree frames. Then, writing $E(x) = \{Ax : A \in \operatorname{Hom}_H(V, W)\} + W^H$,*

$$\overline{\mathcal{A}_{\mathrm{sc}}(X, Y)} \cong \{F \in C_H(V, W) : F(x) \in E(x) \text{ for all } x\}. \tag{22}$$

In particular, the enlarged generators are dense in all continuous natural maps if and only if

$$\{Ax : A \in \operatorname{Hom}_H(V, W)\} + W^H = W^{H_x} \quad \text{for every } x \in V. \tag{23}$$

The lower inclusion already uses scalar channels to S_0 computed by finite averaged MLPs, natural affine maps, and finite branches.

Proof. Theorem I.1 gives the upper inclusion. Choose as B_j a basis of linear intertwiners applied to x, together with a basis of constant maps valued in W^H. These are natural affine maps after transport. Each term $q_j(x)B_j(x)$ has the following exact layer realization:

$$x \longmapsto (x, B_j(x)) \longmapsto q_j(x)(x, B_j(x)) \longmapsto q_j(x)B_j(x).$$

The middle map is a natural scalar rescaling on $X \oplus Y$, using an S_0-channel that reads its X coordinate. The other maps are natural affine maps. Summing the finitely many branches realizes $\mathcal{M}_B$. Lemma I.2 proves the lower inclusion and the criterion. For compact families on all stalks, transport to the reference space and take the finite H-saturation of their union before approximating. □

This is an equality of closures, not an assertion that every continuous map has an exact finite neural representation. It also is not a criterion for all finite-poset diagrams: the reverse inclusion uses invertible non-bottom restrictions and finite joint holonomy. Scalar channels to S_0 are a genuine change from the original constant-scalar generators.

Corollary I.4 (Irreducible feature test)**.** *For self-maps with finite irreducible holonomy $H \subset O(V)$, the scalar-enlarged class is dense if and only if*

$$\dim_{\mathbb{R}} \operatorname{End}_H(V) = \dim_{\mathbb{R}} V.$$

Proof. The trivial irreducible representation is one-dimensional and satisfies the criterion. Otherwise $V^H = 0$. Every nonzero equivariant endomorphism is invertible: its kernel and image are invariant subspaces. Hence the evaluation map $A \mapsto Ax$ is injective for every $x \neq 0$, and $\dim E(x) = \dim \mathrm{End}_H(V)$. Equality of the two dimensions makes $E(x) = V$ for all $x \neq 0$, while $E(0) = V^H = 0$, so (23) holds. If the dimensions differ, choose x outside the finite union of fixed subspaces of nonidentity elements of H. Its stabilizer is trivial, and $E(x)$ is a proper subspace of V. □

Thus scalar gates suffice for the standard planar C_4 representation, whose commutant is $\mathrm{span}\{I, J\}$, but fail for the standard planar D_4 representation, whose commutant is $\mathbb{R}I$. This failure persists under all the hidden diagrams and bundle operations in the enlarged class, not just stacks of a fixed representation.

I.4 A COMPLETE PLANAR EXAMPLE AND ONE ADDITIONAL VECTOR FEATURE

Write D_4 for the order-eight symmetry group of the square, generated by $J(x, y) = (-y, x)$ and $R(x, y) = (x, -y)$. Define

$$p(z) = (x^3 - 3xy^2,\ y^3 - 3x^2y), \qquad z = (x, y).$$

The map p is D_4-equivariant and is the real form of $\bar{z}^3$. Because $\mathrm{End}_{D_4}(\mathbb{R}^2) = \mathbb{R}I$ and $(\mathbb{R}^2)^{D_4} = 0$, Theorem I.3 says that the scalar-enlarged closure consists exactly of continuous equivariant maps whose output is parallel to z at every $z \neq 0$, and whose output at zero is zero.

The two vector features z and $p(z)$ satisfy

$$\det[z\ p(z)] = 4xy(y^2 - x^2).$$

Away from the axes and diagonals, they span $\mathbb{R}^2$ and the stabilizer is trivial. On those lines, at a nonzero point, the stabilizer is a reflection and its fixed space is $\mathbb{R}z$. At zero both features vanish and the full group has fixed space zero. Thus these two features span precisely the stabilizer-allowed values everywhere.

Two scalar features suffice:

$$\phi(z) = (s, t) = (x^2 + y^2,\ x^2y^2).$$

They separate D_4 orbits. Indeed, x^2, y^2 are the unordered roots of $u^2 - su + t = 0$, and D_4 comprises all signed permutations of the two coordinates. On any invariant compact set, ϕ therefore identifies the orbit space homeomorphically with its image. Every continuous invariant coefficient factors continuously through that image. The image is compact, hence closed in $\mathbb{R}^2$; a continuous extension to $\mathbb{R}^2$ followed by ordinary MLP approximation supplies the scalar network.

Corollary I.5 (Two vector features suffice)**.** *For ordinary two-output MLPs (a_θ, b_θ) with tanh hidden units, linear outputs, and widths allowed to increase, the family*

$$f_\theta(z) = a_\theta(\phi(z))z + b_\theta(\phi(z))p(z) \tag{24}$$

is dense in all continuous D_4-equivariant maps $\mathbb{R}^2 \to \mathbb{R}^2$ on compact sets. On a graph local system with this holonomy, its transported maps are natural. It is realized by adding the one direct covariant feature p to the S_0 scalar construction and using the branch realization in the proof of Theorem I.3.

Proof. Apply Lemma I.2 to the two columns just analyzed, replacing the averaged MLP coefficients by MLPs on the orbit-separating map ϕ. The same regularized coefficients factor continuously through ϕ on each invariant compact set, so the proof is unchanged. □

The new vector feature is necessary relative to the scalar-enlarged class: without any additional direction-producing operation, Theorem I.1 prevents density. Adding this one feature suffices. This is not a minimal-parameter claim or a claim that this is the only possible repair. The scalar MLP may include an affine skip, without changing density; then choosing $a(s, t) = s$ and $b(s, t) = \beta$ realizes the response below exactly. The two features are fixed independently of a training target.

Exact errors distinguish three design choices. Consider the fourfold-symmetric response

$$F_\beta(z) = \|z\|^2 z + \beta p(z). \tag{25}$$

For $z = r(\cos\theta, \sin\theta)$, its radial and tangential components are $r^3(1 + \beta\cos 4\theta)$ and $-\beta r^3 \sin 4\theta$. The best parallel-output approximant is the pointwise orthogonal projection

$$F_\parallel(z) = q_*(z)z, \qquad q_*(z) = r^2(1 + \beta\cos 4\theta), \quad q_*(0) = 0.$$

This coefficient is continuous and invariant; equivalently, $q_*(z) = s + \beta(s^2 - 8t)/s$ for $s > 0$. Its error has norm $|\beta|r^3|\sin 4\theta|$. Consequently the scalar-enlarged class has best uniform error exactly $|\beta|$ on the unit disk and best population vector mean squared error $\beta^2/8$ for uniform disk inputs. The latter uses $\mathbb{E}r^6 = 1/4$ and $\mathbb{E}\sin^2(4\theta) = 1/2$. The projection belongs to the broad class; finite scalar MLPs approach both infima by uniform approximation.

For comparison, restricting the shallow readout to $a(\|z\|^2)z$ gives best population error $\beta^2/4$: its coefficient cannot depend on angle. Its minimizer is $a(s) = s$. This is a statement about that specified norm-only readout, not about every possible hidden-diagram architecture with some norm operations. For the original constant-scalar generators the best population error is $1/36 + \beta^2/4$, attained by $2z/3$. These yield the following exact population infima on the same D_4 input/output local system:

Class	Vector MSE	RMSE when $\beta = 1/4$
Original generators	$1/36 + \beta^2/4$	$5/24$
Norm-only readout $a(\|z\|^2)z$	$\beta^2/4$	$1/8$
All natural scalar rescalings, with nesting	$\beta^2/8$	$1/(8\sqrt{2})$
Two-feature readout (24)	0	0

These are mathematical approximation limits, not newly measured training results. They separate constant coefficients, insufficient scalar information, missing vector directions, and a sufficient feature choice.

I.5 Novelty, implementation, and remaining scope

Invariant scalar coefficients multiplying covariants are established equivariant approximation machinery (Yarotsky, 2022, Propositions 2.3–2.4). Villar et al. (2021) describe scalar-based universal constructions for several physical symmetry groups. The polynomial ingredients of (24) are classical reflection-group invariant theory: z and $p(z)$ are scaled gradients of basic invariants $x^2 + y^2$ and $\mathrm{Re}(x + iy)^4$ (Solomon, 1963). The limitation of direction-preserving nonlinearities also predates this work (Xu et al., 2022). The contribution here is the full categorical generator restriction and its consequences, rather than the invariant-scalar machinery or the polynomial features themselves. Cen et al. (2025, Theorem 3.2) also give a completeness criterion from invariant scalars and a spanning steerable basis.

The categorical results established here are the finite-poset bound (17), including arbitrary hidden diagrams and nonlinear bundle operations, and its combination with scalar approximation to obtain the exact closure (22). The irreducible test makes that criterion a small commutant calculation.

Exact checks are in `verification/check_scalar_design.py`. They solve naturality on a graph with two independent cycles and holonomy D_4, and verify the commutants, trace, stabilizer examples, polynomial identities, and exact risk calculations. The reference implementation `experiments/scalar_design.py` implements (24), including its actual graph restrictions and the affine/gate branch correspondence. Its audit checks nontrivial local frames and finite neural coefficients. The Appendix G.4 comparison measures these infima by training; no performance advantage over existing architectures is claimed. The Agda module `Equivariant/Direction.agda` constructs bundle bias extensions and proves the pointwise direction bound by typed network induction, without requiring a least object. A separate checked embedding preserves evaluation of the original generators. Compact-open closure, regularized coefficient approximation, and the stabilizer criterion remain pen-and-paper results, supported by the finite checks above. General non-invertible feature diagrams still have only the necessary condition in Theorem I.1. The analysis

concerns fixed stalkwise response models; extending it to full graph-prediction tasks is outside the present scope. The criterion gives density, not an approximation rate or bounds on neural weights near points where the dimension of the available span changes.

Why the poset assumption matters. Over the one-object category of D_4, regular-bundle Reynolds compilation admits $(x, y) \mapsto (\tanh x, \tanh y)$. This map commutes with every signed coordinate permutation, and applying it before the Reynolds compiler returns the same map. At $(1, 2)$ its value is not parallel to $(1, 2)$, since $\tanh 2 - 2\tanh 1 = -2(\tanh 1)^3/(1+(\tanh 1)^2) \neq 0$. Thus the finite-poset direction bound would be false for general finite categories. In a group arrow bundle, natural biases need not span the whole stalk: different arrows have the same source and impose relations between its slots. The proof uses thinness, not merely the holonomy group.

J Input multiplicity and average approximation

The exact direction class has consequences beyond a single representation. We first compute its best approximation error for arbitrary targets and input measures. We then specialize to repeated irreducible features, where the commutant gives explicit thresholds. Throughout, the graph, zero bottom stalks, orthogonal restrictions, finite joint holonomy, and generator class are those of Theorem I.3.

J.1 Exact error from the available directions

Let X, Y have reference representations V, W, and let $E(x) = \mathrm{Hom}_H(V, W)x + W^H$. Choose a matrix $B(x)$ whose columns are the values of a real basis of $\mathrm{Hom}_H(V, W)$ and a basis of W^H as constant covariants. Thus $\mathrm{im}\, B(x) = E(x)$. For $F \in C_H(V, W)$ put $d_F(x) = \mathrm{dist}(F(x), E(x))$.

Lemma J.1 (Exact approximation errors)**.** *Let $K \subset V$ be nonempty and compact, and let μ be any finite Borel measure supported on K. Identifying networks with their reference maps, for $1 \leq p < \infty$,*

$$\inf_{N \in \mathcal{A}_{\mathrm{sc}}(X,Y)} \|F - N\|_{L^p(\mu)} = \|d_F\|_{L^p(\mu)}\,, \tag{26}$$

$$\inf_{N \in \mathcal{A}_{\mathrm{sc}}(X,Y)} \sup_{x \in K} \|F(x) - N(x)\| = \sup_{x \in K} d_F(x). \tag{27}$$

The same infima hold using finite averaged scalar MLP coefficients in the branch construction of Theorem I.3.

Proof. The pointwise direction bound gives both lower bounds. For the reverse bounds, use the regularized coefficients from Lemma I.2:

$$c_\varepsilon(x) = B(x)^\top [B(x)B(x)^\top + \varepsilon I]^{-1} F(x), \qquad F_\varepsilon(x) = B(x)c_\varepsilon(x). \tag{28}$$

These coefficients are continuous and invariant: $B(hx) = h_W B(x)$, $F(hx) = h_W F(x)$, and h_W is orthogonal. Each F_ε is an allowed finite scalar-branch map. Its residual is

$$F(x) - F_\varepsilon(x) = \varepsilon [B(x)B(x)^\top + \varepsilon I]^{-1} F(x).$$

In an orthonormal eigenbasis of $BB^\top$, the multipliers are $\varepsilon/(\lambda + \varepsilon)$, $\lambda \geq 0$. The residual norm decreases as $\varepsilon \downarrow 0$, is bounded by $\|F(x)\|$, and converges to the norm of the projection onto $\ker B(x)^\top$, namely $d_F(x)$. Dominated convergence proves (26).

For (27), take $\varepsilon_j \downarrow 0$ and $t > \sup_K d_F$. The compact sets

$$C_j = \{x \in K : \big\|F(x) - F_{\varepsilon_j}(x)\big\| \geq t\}$$

are nested and have empty intersection. Compactness implies some C_j is empty, giving the uniform upper bound as $t \downarrow \sup_K d_F$. This argument does not assume continuity of d_F or of the image projector where rank changes. For each fixed ε, approximate its finitely many scalar coefficients uniformly on the finite H-saturation of K. Since B is bounded there, the finite scalar MLP realizations approach both infima as well. □

The linear algebra is standard. The uniform identity also follows from classical function-module localization (Chen, 2024, Corollary 2.10): on the compact H-saturation of K, invariant scalars separate orbits, and an equivariant residual has constant norm on each orbit. The module's values at x are exactly $E(x)$. The proof above supplies one explicit regularization for both error notions. Its architectural use depends on the full-generator direction bound and the permitted scalar-branch realization. Neither infimum is asserted to be attained. The result also clarifies the role of the input distribution: average error measures how much target mass lies outside $E(x)$, while uniform error retains every point of K.

J.2 An exact classification for repeated irreducible features

Let H act faithfully and irreducibly on a nonzero real space U. A nonfaithful action reduces to this case by replacing H with its image. Define

$$\mathbb{D} = \mathrm{End}_H(U), \qquad d = \dim_{\mathbb{R}} \mathbb{D}, \qquad n = \dim_{\mathbb{D}} U = \frac{\dim_{\mathbb{R}} U}{d}. \tag{29}$$

By Schur's lemma and Frobenius's classification, $\mathbb{D}$ is a real division algebra isomorphic to $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$. It acts on U on the left by evaluation. Take $V = U^m$ and $W = U^k$, with diagonal actions and $m, k \geq 1$.

For nontrivial U, $W^H = 0$ and block extraction gives

$$\mathrm{Hom}_H(U^m, U^k) = \mathrm{Mat}_{k\times m}(\mathbb{D}), \qquad E(x) = \big(\mathrm{span}_{\mathbb{D}}\{x_1, \ldots, x_m\}\big)^k. \tag{30}$$

If the input span has $\mathbb{D}$-dimension r, then $\dim_{\mathbb{R}} E(x) = kdr$. This argument uses left coefficients and does not assume commutativity of $\mathbb{D}$. For trivial irreducible $U = \mathbb{R}$, fixed biases span the output and the class is uniformly dense; the span formula without biases must not be used at zero in that case.

Theorem J.2 (Multiplicity thresholds)**.** *Let K be the closed unit ball in U^m and μ its normalized Lebesgue measure. The scalar-enlarged categorical networks from U^m to U^k approximate every restriction of a continuous H-equivariant target arbitrarily well:*

1. *uniformly on K if and only if $n = 1$;*
2. *in $L^p(\mu)$, for any fixed $1 \leq p < \infty$, if and only if $m \geq n$.*

The uniform classification also holds for compact-open density on all of U^m. Neither threshold depends on $k \geq 1$.

Proof. The trivial representation was handled above. Suppose U is nontrivial. If $n = 1$, any nonzero input component spans U over $\mathbb{D}$, so $E(x) = W$ for $x \neq 0$. At zero, equivariance forces $F(0) = 0$. Theorem I.3 gives compact-open density.

If $n > 1$, select a nonzero u with trivial stabilizer. Such vectors lie outside the finite union of proper fixed subspaces of nonidentity elements of H. Scale it so $x_0 = (u, 0, \ldots, 0)$ is inside K. Then $E(x_0) = (\mathbb{D}u)^k$ is proper, although $W^{H_{x_0}} = W$. Choose $w \notin E(x_0)$. A continuous bump β with $\beta(x_0) = 1$, supported in a ball excluding all other points of Hx_0, gives a global equivariant target

$$F(y) = \sum_{h\in H} \beta(h^{-1}y) h_W w, \qquad F(x_0) = w.$$

The pointwise error bound rules out uniform density for every m.

For the finite-p positive direction, if $m \geq n$ then the input $\mathbb{D}$-span equals U almost everywhere. To prove this for all three division algebras, choose a real basis $\delta_1, \ldots, \delta_d$ of $\mathbb{D}$ and form the real matrix with columns $\delta_a x_i$. Its image is the underlying real input span. A maximal real minor is a nonzero polynomial, as seen by choosing a tuple containing a $\mathbb{D}$-basis of U. The rank-deficient locus is contained in its zero set, which has Lebesgue measure zero. Thus $d_F = 0$ almost everywhere and Lemma J.1 proves finite-p density.

If $m < n$, the maximal input rank is m over $\mathbb{D}$. Its locus is open dense by the same real-minor argument. Intersect it with the open dense free-orbit locus and the interior of K, and choose x_0

there. Again select $w \notin E(x_0)$ and construct the equivariant bump target. On a neighborhood of x_0, the matrix B has constant rank, so its image projector is continuous. Hence $d_F \geq \eta > 0$ on some open ball $A \subset K$, which has positive measure. Every network then satisfies $\|F - N\|_{L^p(\mu)} \geq \eta\mu(A)^{1/p} > 0$. □

Here m counts independently varying input vectors. Copying one input into hidden channels does not change its span. The positive finite-p result holds for every finite compactly supported measure absolutely continuous with respect to Lebesgue measure. Necessity was proved for the full input ball with uniform measure; singularly supported distributions can retain an average-error obstruction even above the threshold. Under full-support ball measure, the essential supremum of a continuous residual equals its supremum, so the uniform obstruction also holds in L^∞. Finiteness is essential to the free-orbit argument: the standard $O(3)$ action on one vector has a nontrivial generic stabilizer, and these necessity thresholds do not transfer to it.

Feature representation	d	n	Uniform density	Finite-p density
One-dimensional sign	1	1	Every $m \geq 1$	Every $m \geq 1$
Planar C_4 rotations	2	1	Every $m \geq 1$	Every $m \geq 1$
Left Q_8 action on $\mathbb{R}^4$	4	1	Every $m \geq 1$	Every $m \geq 1$
Planar D_4 square symmetry	1	2	No m	Exactly $m \geq 2$
Cube rotations on $\mathbb{R}^3$	1	3	No m	Exactly $m \geq 3$
Tensor $C_4 \times D_4$ on $\mathbb{R}^4$	2	2	No m	Exactly $m \geq 2$
Tensor $Q_8 \times D_4$ on $\mathbb{R}^8$	4	2	No m	Exactly $m \geq 2$

The last two rows use the faithful image of the tensor action, of order 16 and 32 respectively. They test the complex and quaternionic thresholds with $n > 1$. Any finite representation can be realized as graph holonomy. For generators $h_1, \ldots, h_s$, take s triangles sharing one vertex, identity restrictions along a spanning tree, and one exceptional restriction per cycle realizing h_i. The resulting graph is finite, connected, and loopless. For the trivial group, an edge suffices.

J.3 A TARGET WITH ZERO MSE INFIMUM AND UNIFORM ERROR ONE

Corollary J.3 (A fixed target separates the two errors)**.** *Let D_4 act diagonally on $m \geq 2$ planar inputs, and let $p(z) = \overline{z}^3$. For $F(x) = p(x_1)$ on the unit ball $K \subset \mathbb{R}^{2m}$ with normalized Lebesgue measure,*

$$\inf_{N \in \mathcal{A}_{\mathrm{sc}}} \mathbb{E}\,\|F - N\|^2 = 0, \qquad \inf_{N \in \mathcal{A}_{\mathrm{sc}}} \sup_K \|F - N\| = 1. \tag{31}$$

Proof. Here $d = 1$, $n = 2$, so the finite-p theorem gives the first equality. At the unit vector $u = e^{i\pi/8}$, $p(u)$ is unit and perpendicular to u. At $x_0 = (u, 0, \ldots, 0)$, every network value belongs to $E(x_0) = \mathbb{R}u$, giving error at least one. The zero network attains one because $\|p(x_1)\| = \|x_1\|^3 \leq 1$ on K. □

For (28), the error at this same x_0 is exactly one for every $\varepsilon > 0$: $B(x_0)^\top p(u) = 0$. Average error can therefore tend to zero while that fixed error persists. No continuous network attains zero MSE on this full-support distribution, since it would then equal F throughout K. The statement is about an infimum and gives no optimization, parameter-size, or sample-complexity guarantee.

J.4 A REPAIR FOR EVERY PLANAR DIHEDRAL GROUP

For D_N of order $2N$, $N \geq 3$, acting by $z \mapsto \omega z$, $\omega = \exp(2\pi i/N)$, and $z \mapsto \overline{z}$, the commutant is $\mathbb{R}I$. Define

$$p_N(z) = \overline{z}^{\,N-1}, \qquad \phi_N(z) = (|z|^2, \operatorname{Re} z^N).$$

Since $\overline{\omega}^{N-1} = \omega$, p_N is equivariant, and ϕ_N separates orbits. The determinant is

$$\det[z, p_N(z)] = -r^N \sin(N\theta).$$

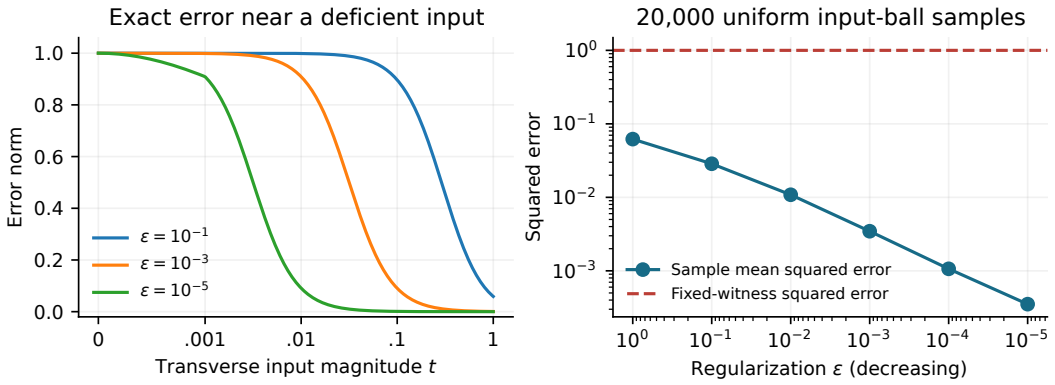


Figure 3: Regularized approximators for the two-input D_4 cubic target; no training is performed. Left: the exact error along $x(t) = (u, tp(u))/\sqrt{1+t^2}$, with $u = e^{i\pi/8}$, is $(1+t^2)^{-3/2}\varepsilon/(t^2/(1+t^2)+\varepsilon)$. Error stays one at $t = 0$ while its neighborhood narrows. Right: mean squared errors on 20,000 fixed samples from the unit ball in $\mathbb{R}^4$, together with the exact squared error at the fixed witness. The population limit is proved in Corollary J.3; the sampled means do not certify that limit.

Off the reflection axes the columns span the plane; on an axis the nonzero column z spans the fixed line; at zero both the columns and full-group fixed space vanish. Lemma I.2 therefore gives density of

$$a_\theta(\phi_N(z))z + b_\theta(\phi_N(z))p_N(z)$$

with ordinary scalar MLP coefficients. These are classical dihedral covariants (Solomon, 1963), used here to repair the precise categorical restriction.

For target p_N, projecting onto $\mathbb{R}z$ leaves tangential error $r^{N-1}|\sin(N\theta)|$. The scalar projection coefficient $r^{N-2}\cos(N\theta)$ extends continuously by zero at the origin. Thus the scalar-enlarged single-input class has best uniform error one on the disk and best uniform-disk vector MSE

$$\left(\int_0^1 2r^{2N-1}\,dr\right)\left(\frac{1}{2\pi}\int_0^{2\pi}\sin^2(N\theta)\,d\theta\right) = \frac{1}{2N}.$$

J.5 Prior work and verification boundary

Yarotsky (2022, Sections 2.2–2.3) use generating polynomial covariants and polarization to handle multiplicities; those nonlinear features are available in their construction. Our theorem instead classifies the directions available from the specified affine, scalar, and bundle operations. Cen et al. (2025, Theorem 3.2 and Appendix A.3) give pointwise invariant-scalar decompositions using a spanning steerable basis. That statement does not itself assert continuous coefficients at rank changes or the closure of our full syntax. Villar et al. (2021, Propositions 4–5) give input-span descriptions for physical symmetry groups; their orthogonal-group stabilizers can force exactly that span, unlike the free finite-group inputs used here. The GVP scalar approximation theorem assumes three independent input vectors (Jing et al., 2021). Uniform and L^p universality already have different thresholds for width-constrained ReLU networks (Park et al., 2021). Continuity obstructions and exceptional inputs also appear in canonicalization (Dym et al., 2024; Shumaylov et al., 2025). These precedents do not by themselves supply the full categorical generator invariant or its necessary-and-sufficient input thresholds. The contribution is that characterization and its consequences; the underlying coefficient, rank, and approximation principles are established.

`verification/check_breadth.py` checks representative commutants, input ranks, the exact D_4 witness, dihedral identities for $N = 3, \ldots, 8$, and a deterministic regularization illustration. These checks are not training results. Agda's `Multiplicity.agda` proves equivariant block extraction/assembly, both inclusions of the evaluated span, and persistence under zero padding. `Regularization.agda` proves the residual identity, inverse covariance, and persistence on the transpose kernel, with inverses supplied as hypotheses. The Schur/Frobenius classification, real rank and null-set arguments, spectral estimates, integration, and approximation limits remain pen-

and-paper proofs. The new modules do not constitute a machine verification of Theorem J.2 in its entirety.

K A SHORT GUIDE TO THE DEPENDENT-TYPE LANGUAGE

This appendix explains the notation used in Section 8 without assuming prior knowledge of homotopy type theory. The purpose of the language is to keep a feature, its symmetry transport, and the compatibility of its maps together. It also gives a precise way to specify which network expressions are permitted. The expressivity proofs still require algebra about those expressions.

K.1 A FAMILY IS A COLLECTION OF COMPATIBLE FEATURE SPACES

A type can be read as a collection of objects. A *dependent family* $A : B \to \mathcal{U}$ assigns a type $A(b)$ to each $b : B$; $\mathcal{U}$ denotes a universe of types. The word "dependent" means that the type of a value may change with its index. For this paper, the fibres are usually finite sets, vector spaces, or spaces of linear maps.

In HoTT an identification $\gamma : b = b'$ acts on a family by *transport*, $T^A_\gamma : A(b) \simeq A(b')$. Think of this as a reversible change of coordinates. A 1-type is a type whose identification spaces are sets; it can be interpreted as a groupoid. Thus it records objects and reversible arrows, including their composition, with no further nontrivial higher identification data. This use of a path is structural: it is not a trajectory followed by a numerical input.

For an ordinary group G, its *classifying type* BG has one connected component and a chosen point $*$. Its loops at $*$ form G. A vector-space family $\mathcal{V} : BG \to \mathrm{Vect}_{\mathbb{R}}$ therefore specifies a representation: its underlying space is $V = \mathcal{V}(*)$ and transport along g is $\rho_V(g)$. Conversely, the representation gives this family in the univalent presentation. This is the action-as-family viewpoint of Bezem et al. (2026, Chapter 5). Euclidean topology and continuity are additional structure; they do not follow from the type alone.

K.2 A DEPENDENT FUNCTION CARRIES THE COMPATIBILITY EQUATION

A *section* $s : \prod_{b:B} A(b)$ chooses a value in each fibre. The product symbol here denotes a dependent function type, not a numerical multiplication. Since s is a function on the base type, its values satisfy

$$T^A_\gamma(s(b)) = s(b') \qquad (\gamma : b = b').$$

For the family of linear maps between $\mathcal{V}$ and $\mathcal{W}$, transport is conjugation:

$$T^{\mathrm{Hom}}_g(L) = \rho_W(g) L \rho_V(g)^{-1}.$$

Consequently a section of this family is exactly an equivariant linear map. The familiar equation $\rho_W(g)L = L\rho_V(g)$ is the section's compatibility condition. It has not been discarded: the dependent type packages it, and a construction of a section must supply it.

Here is the connection to the planar examples. A quarter turn acts by $J = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. For input and output $\mathbb{R}^2$, a linear-map section over BC_4 must satisfy $LJ = JL$, hence $L = aI + bJ$. Adding the reflection $R = \mathrm{diag}(1, -1)$ requires $LR = RL$, forcing $b = 0$. Thus the same section construction yields

$$\mathrm{End}_{C_4}(\mathbb{R}^2) = \mathrm{span}_{\mathbb{R}}\{I, J\}, \qquad \mathrm{End}_{D_4}(\mathbb{R}^2) = \mathbb{R}I.$$

At a nonzero input z, the first gives all planar directions $az + bJz$; the second gives only the line $\mathbb{R}z$. The direction theorem extends this elementary calculation to the full scalar-enlarged network syntax. It is that induction, rather than the notation alone, which rules out recovering a missing direction through hidden bundles or additional depth.

K.3 SITES, PAIRS, AND STABILIZERS USE THE SAME CONSTRUCTION

The type $\sum_{b:B} P(b)$ consists of pairs (b, p) with $p : P(b)$; this is the *dependent sum*, not a vector-space sum. For a G-set $P : BG \to \mathrm{Set}$, it presents the action groupoid: its objects are sites, and an

arrow from p to p' is a symmetry taking p to p'. A feature bundle is a vector-space family over this total type. The two constructions needed for a kernel are therefore

$$\underbrace{\sum_{b:BG}(Q(b)\times P(b))}_{\text{target–source pair base}}, \qquad \underbrace{\prod_{b:BG}\prod_{q:Q(b)}\prod_{p:P(b)}\mathrm{Hom}(V_{b,p},W_{b,q})}_{\text{compatible kernel blocks}}.$$

The common b ensures that both sites use the same symmetry. A loop at a pair (q,p) is a symmetry fixing both sites. Its group is the pair stabilizer $H_{q,p}$. A kernel block at that pair must be fixed by its Hom transport, so it is an $H_{q,p}$-intertwiner. Choosing one pair from each orbit gives the orbit formula in (16). This is a rederivation of the known parameterization, with the compatibility data handled by one construction.

K.4 Why the noninvertible part stays explicit

A restriction such as $\mathbb{R} \to 0$ loses information and has no inverse. It cannot be transport along an identity path. We therefore retain the poset, its linear restriction maps, and their composition laws explicitly. HoTT can describe such structured diagrams, but does not turn their arrows into reversible paths. After reducing a graph local system to holonomy, the invertible part can again use the family presentation above.

The other useful dependent type indexes a network expression by its input and output diagrams: $\mathrm{Net}(X,Y)$. Composition takes an expression in $\mathrm{Net}(X,Z)$ and one in $\mathrm{Net}(Z,Y)$, so intermediate feature types must agree. Affine layers, scalar gates, branches, and bundle lifts are constructors of this expression type. Induction over these constructors then has two uses in this paper:

- For the original generators, quotienting the output by $T(Y)$ makes every expression affine along a bottom-restriction fibre.
- With all natural scalar rescalings admitted, each evaluated output has a natural affine witness: $N_a(x) \in E_{X,Y}(a,x)$.

The witness in the second statement may depend on x; it does not make the network globally affine. The Agda files `TraceNetwork.agda` and `Direction.agda` check these inductions; `DirectionEmbedding.agda` checks that the original expressions embed without changing evaluation. Appendix E gives their exact scope.

The benefit is a common language for lawful maps and for the expressions built from selected maps. It supports reuse of the kernel algebra and precise induction over architectures. Neither barrier requires univalence, and both have ordinary categorical proofs. The dependent presentation is valuable as an organizing and formalization tool, not an additional universality theorem or a change to the numerical learning algorithm.