Differential Topology of Parametrized Neural Networks

Jet representations, decision boundaries, and reconstruction

November 2026

Abstract. A smooth neural architecture represents a space of weights by evaluations on a data manifold. We give architecture-level conditions for controlling first jets, decision boundaries, and reconstruction. An affine readout of immersed features is submersive on first jets, while an arbitrary smooth feature map with free final biases still gives generic transverse classification boundaries. We compute a fold discriminant for a quadratic readout on the circle and a regular family whose decision boundaries have nontrivial monodromy. For equivariant layers the stabilizer representation can force an excess intersection even when the total evaluation is submersive. An explicit torus calculation exhibits the associated obstruction complex. For autoencoders, exact ambient recovery requires an embedding, and a spherical example gives a sharp uniform error bound below the required latent dimension. Relative mapping objects and descent specify how these results apply to local parameter and feature presentations.

Keywords. Parametrized networks, transversality, representation theory, singularities, monodromy, autoencoders.

1 Networks as families of smooth maps

A parametrized neural network is a family of maps realized by a specified architecture. Differential topology studies these maps through their derivatives, their fibers, and their behavior along paths of parameters. This makes it possible to ask which local perturbations a layer actually supplies and which global topological constraints remain even if local perturbations are plentiful. A basic example is an encoder with a smooth, trainable readout. The readout can have enough derivatives to move a decision boundary, while a narrow intermediate space still obstructs reconstruction on the whole data manifold.

The starting points are Hirsch's treatment of jets and transversality [8, Chs. 2–4] and their extension to relative families, derived pullbacks, and equivariant representations in [1, Chs. 18–22, 25–28]. Earlier categorical models of learning represent layers by parametrized maps, including their compositions and shared weights [6, 2]. Here the map from a prescribed architecture to the function object is subjected to differential-topological tests. Related empirical work investigates changes in data topology across network layers [10]. Our statements concern smooth realizations and their included perturbations.

1.1 Relative evaluation and representations

Let $\mathcal{E}$ be the ∞-topos of sheaves of spaces on smooth test manifolds, and let B be a coefficient or geometry base. For $X, Y \to B$ write $[X,Y]_B$ for the internal mapping object in $\mathcal{E}_{/B}$. A model with weights $P \to B$ is an evaluation

$$F : P \times_B X \longrightarrow Y$$

over B, or its transpose $P \to [X,Y]_B$. For example, if $B = *$ and X, Y are represented by smooth manifolds, an evaluation on P is a smooth map $P \times X \to Y$ on ordinary representable parameter charts. The classifying map records the realizations available to that architecture. Two different points of P can specify the same evaluation, and a change of coordinates acts on the presenting parameter space. A parameter shared by multiple factors of the network is a pullback along the corresponding diagonal into the product of their parameter spaces.

Internal exponentials commute with base change in B. For an effective epimorphism $U \to B$ with Čech nerve $U_\bullet$, we have the descent equivalence

$$\mathcal{E}_{/[X,Y]_B} \simeq \lim_{[n]\in\mathbf{\Delta}} \mathcal{E}_{/[X|_{U_n}, Y|_{U_n}]_{U_n}}$$

from [1, Ch. 18]. Thus compatible evaluations and parameter presentations on geometry charts glue. On charts of smooth bundles, a change of input and output coordinates transforms the vertical derivative by the derivative of the output change on the left and the inverse input change on the right. It preserves rank, so rank strata and regularity conditions are intrinsic to the glued model. For an action, the same formula is conjugation by the linear representations on tangent spaces. Its restriction to stabilizers becomes a concrete selection rule in Section 3.

1.2 Which kind of realization is measured

A differentiable classifier with real logit f_p has a decision boundary $f_p^{-1}(0)$ and sublevel sets $f_p^{-1}((-\infty, c])$. A scalar strictly monotone link, such as the logistic link, preserves the boundary when the corresponding threshold is transported. A classifier with several logits admits a stratification by the indices realizing the maximum. An autoencoder, by contrast, is a factorization of an approximation to the identity on a data manifold. Its topology depends on the encoder's image in the latent space and

on the codomain of the decoder. We use smooth activations for all jet and regular-value assertions. The uniform reconstruction obstruction requires only continuous realizations, so it also applies to continuous piecewise-linear activations.

2 Readouts and the available jet directions

Let M be an n-dimensional smooth manifold without boundary. Fix a smooth feature extractor $\phi : M \to \mathbb{R}^d$ and use the affine readout

$$f_{A,b}(x) = A\phi(x)+b, \qquad (A,b) \in \mathcal{A} = \mathrm{Hom}(\mathbb{R}^d, \mathbb{R}^q)\times\mathbb{R}^q. \tag{1}$$

The choice to hold the feature extractor fixed specifies a subfamily of whatever full architecture realizes ϕ. This separates the derivatives supplied by the readout from those provided by perturbing upstream layers.

Theorem 2.1 (First-jet control by an affine readout)**.** *If ϕ is an immersion, then $J : \mathcal{A} \times M \to J^1(M, \mathbb{R}^q)$, $(A, b, x) \mapsto j^1_x f_{A,b}$, is a submersion. Without the immersion assumption, the corresponding map to $J^0(M, \mathbb{R}^q)$ remains a submersion. In either case, for a finite or countable family of smooth jet submanifolds at the indicated order, the parameters for which the corresponding jet section is transverse to all of them form a conull residual set.*

Proof. At fixed x, use jet coordinates (y, L) with $L : T_xM \to \mathbb{R}^q$. The differential in the readout directions is

$$(\dot{A}, \dot{b}) \longmapsto \big(\dot{A}\phi(x) + \dot{b},\ \dot{A}D\phi_x\big).$$

Given any $L_0 : T_xM \to \mathbb{R}^q$ and immersive ϕ, extend $L_0(D\phi_x)^{-1}$ linearly from $D\phi_x(T_xM)$ to $\mathbb{R}^d$ to obtain $\dot{A}$. Then any value variation y_0 is realized by $\dot{b} = y_0 - \dot{A}\phi(x)$. This proves surjectivity in the vertical jet coordinates. Base coordinates are covered because J lies over the projection to M. For J^0 only the value variation is needed, and free biases supply it at every x.

A submersion is transverse to each smooth jet submanifold S. Its preimage is smooth, and regular values of the projection of the preimage to $\mathcal{A}$ are exactly the parameters whose jet extensions are transverse to S. This is the parametric transversality argument [1, Ch. 6]. Sard's theorem gives a null exceptional set. The critical locus has a countable compact exhaustion because it is closed in a second-countable manifold. Its images under the projection are compact null sets and therefore nowhere dense. Taking the countable intersection of the good parameter sets proves the claim. □

For a binary classifier, Theorem 2.1 implies that generically the decision boundary is empty or a codimension-one smooth submanifold, even if ϕ has singular points. The free final bias alone suffices. A fixed-bias family can have additional constraints. For example, if a feature identically vanishes on a region, a bias-free readout may vanish on that entire region. For vector outputs, immersion of ϕ supplies every first-order normal perturbation at a given input. It makes no assertion that independent perturbations at multiple inputs can be realized by the same $\dot{A}$.

2.1 Several classes and several points

Consider q logits and the open part of the tie stratum where a subset I of r classes is exactly the set of maximizers. Fix $i_0 \in I$. Equalities $y_i - y_{i_0} = 0$ for $i \in I \setminus \{i_0\}$ are $r-1$ independent linear equations, and strict inequalities separate the other coordinates. This is a smooth codimension-$(r-1)$ submanifold of $\mathbb{R}^q$. Consequently, the inverse image for a generic readout is empty or has codimension $r-1$. No r-way maximum tie occurs if $r > n+1$. The conclusion also holds when the bias space is restricted to vectors with coordinate sum zero. The differences of any bias vector are unchanged by subtracting a constant vector so the variation still surjects onto the normal directions to each tie. Thus the ordinary redundancy of adding the same constant to every logit can be removed without losing the relevant perturbations.

There is a complementary global result when ϕ is an embedding. For different points x, x' the map $(A, x, x') \mapsto A(\phi(x)-\phi(x'))$ is a submersion to $\mathbb{R}^q$. Indeed, $\phi(x)-\phi(x')$ is nonzero and a linear map can assign it any value. The source in the input variables is $M\times M\setminus\Delta$, of dimension $2n$. When $q > 2n$, parametric transversality forces absence of off-diagonal coincidences for a conull residual set of A. By the first-jet argument, r-rank derivative strata of codimension $(n-r)(q-r)$ are avoided when that codimension exceeds n. This formula follows by choosing an invertible r-minor and using block elimination: the remaining $(q-r)\times(n-r)$ block after elimination must be zero. In particular for $q \geq 2n$ all nonimmersive strata are avoided. For compact M and $q \geq 2n+1$ the generic readout of an embedding is thus an embedding. An affine bias has no effect on this two-point test. This is a version of the projection argument for embeddings applied to the allowed neural readouts.

2.2 A two-parameter boundary discriminant

Use the unit circle as a data manifold with coordinate $x = (\cos t, \sin t)$. A network with one square unit and affine readout can realize

$$F(a, b, t) = \cos t + a\cos(2t) - b$$

because $\cos(2t) = 2(\cos t)^2 - 1$. Its zero locus in (a, b, t) is smooth since the last bias derivative is -1. Projection to (a, b) is proper because the circle is compact. A parameter

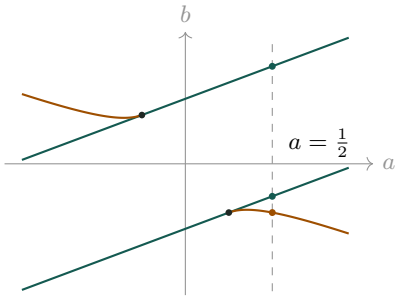


Figure 1: Boundary discriminant of the quadratic circle score. The lines come from $t = 0, \pi$ and the curved branches from $\cos t = -1/(4a)$. The black points mark degenerate critical points. The dashed slice gives the regular counts in the text.

belongs to the discriminant when a root also satisfies

$$\partial_t F = -\sin t(1 + 4a\cos t) = 0.$$

Thus the discriminant consists of the two lines $b = a + 1$ and $b = a - 1$ and the curve $b = -a - 1/(8a)$ for $|a| \geq 1/4$. At $|a| = 1/4$ the latter joins a line with a degenerate critical point. At $a = 1/4$ and $t = \pi$ the first nonzero variation of $\cos t + a\cos 2t$ in $t - \pi$ is quartic, with coefficient $1/8$.

For example, fix $a = 1/2$. With $u = \cos t$, the logit equals $u^2 + u - 1/2 - b$. The minimum as a function of $u \in [-1, 1]$ occurs at $u = -1/2$, while the endpoint values before subtracting b are $-1/2$ and $3/2$. Every interior root in u yields two roots on the circle and an endpoint root yields one. The number of boundary points at regular thresholds is therefore

Regular threshold b	Boundary points
$(-\infty, -3/4)$	0
$(-3/4, -1/2)$	4
$(-1/2, 3/2)$	2
$(3/2, \infty)$	0

At the lower threshold two critical points enter simultaneously because the network score is even in t. At the other thresholds there is a single nondegenerate critical point. Each isolated nondegenerate event has the fold normal form, so it adds or removes an opposite-sign pair of local roots [1, Ch. 20]. The unsigned boundary topology changes, while the signed count at every regular threshold is zero.

More generally, for a compact M and a scalar evaluation $F : P \times M \to \mathbb{R}$ with 0 a regular value, the projection of the zero locus is a proper submersion off its discriminant. It is a smooth bundle over each regular chamber in its image [1, Ch. 20]. Over the other chambers the fibers are empty. Local triviality concerns a whole chamber. It leaves open the possibility of monodromy around a loop in that chamber, even if the point count is fixed. Section 4 computes that effect in a different network.

3 A stabilizer obstruction to regular readouts

A model with moving parameters can be equivariant even when most of its fixed-parameter maps have few symmetries. Let a compact Lie group G act smoothly on manifolds P, X, and linearly on a finite-dimensional real space V. Suppose a smooth readout $F : P \times X \to V$ satisfies

$$F(gp, gx) = gF(p, x).$$

A single realization $F_p(x) = F(p, x)$ is equivariant for $K = G_p$. At a zero x of F_p, the subgroup governing its derivative is $H = K_x = G_p \cap G_x$. The general relation between stabilizers and normal derivatives is developed in [1, Chs. 25–26].

Proposition 3.1 (Regularity at a parameter stabilizer)**.** *If F_p is transverse to zero at x, then $T_xX/T_x(Kx)$ admits a surjective H-linear map to V. In particular its σ-isotypic summand has dimension at least that of V_σ for each real irreducible σ of H. Conversely, a surjection of these representations is realizable as the normal derivative of a smooth K-equivariant map defined near the orbit Kx.*

Proof. For $h \in H$, differentiate $F_p(hx') = hF_p(x')$ at x. The result is $D_xF_p \circ D_xh = h \circ D_xF_p$. The derivative vanishes on $T_x(Kx)$ since F_p is zero along the orbit. Surjectivity is preserved when the derivative factors through this quotient. Because H is compact, an averaged inner product gives an H-invariant complement of the kernel of a surjection. Therefore a copy of the target sits in the source representation, with the asserted inequality on each isotypic part.

For the converse, identify $T_xX/T_x(Kx)$ with the normal representation N_x for an invariant metric. The compact slice theorem gives an invariant neighborhood of Kx modeled on $K \times_H S$, where S is an invariant neighborhood of 0 in N_x [1, Ch. 25]. For an H-linear surjection $L : N_x \to V$, the assignment $[k, s] \mapsto kL(s)$ is well defined by equivariance of L. Its normal derivative at $[1, 0]$ is L, and translation by K transports this surjectivity along the orbit. □

The converse concerns arbitrary equivariant smooth germs. An architecture restricts which such germs can occur. In the total space the parameter derivative contributes an additional H-map $T_pP \to V$. Surjectivity at (p, x) tests the sum of its image and the image of the feature derivative. The two regularity tests correspond to different domains of the same evaluation.

3.1 Two sign readouts on a torus

Take $G = C_2$ and $X = (\mathbb{R}/2\pi\mathbb{Z})^2$, with the reflection $r(\theta, \phi) = (-\theta, \phi)$. Write $\mathbb{R}_+$ and $\mathbb{R}_-$ for the trivial and sign representations. Use two real sign readouts, so that $V = \mathbb{R}^2_-$. The fixed locus consists of the two circles $\theta = 0, \pi$. A G-equivariant readout vanishes there because $V^G = 0$. On each fixed circle the source tangent representation is $\mathbb{R}_- \oplus \mathbb{R}_+$: the normal angle changes sign and the angle along the circle is fixed. Its derivative can have rank at most one by Proposition 3.1. Every smooth equivariant map of this feature and output type therefore has a critical zero at every point of the fixed locus. Increasing the number of trainable coefficients within this equivariance constraint leaves the representation bound in force. A single sign readout can be regular along the circles, as is demonstrated by the first output below by itself. The obstruction concerns the simultaneous pair of constraints. Increasing output multiplicity makes their joint zero set a different regularity problem.

Consider the specific realization

$$f(\theta, \phi) = (\sin\theta,\ \sin\theta\cos\phi).$$

The coordinate features $u = \sin\theta$ and $v = \cos\phi$ are respectively sign and trivial features. A square activation realizes their product using $uv = ((u+v)^2 - (u-v)^2)/4$. Under reflection the two inputs $u+v, u-v$ are swapped and simultaneously negated, so their squares are swapped and the difference changes sign. Both readouts therefore have the required transformation law. The zero set is exactly the pair of fixed circles.

3.2 Functions and infinitesimal equations

Near a fixed circle, $t = \sin\theta$ is a transverse coordinate, since its derivative in the angular chart is 1 at $\theta = 0$ and -1 at $\theta = \pi$. Write $c = \cos\phi$ and A for the smooth germ ring on the ambient chart at a point of the circle. The equations are $t = 0$ and $tc = 0$. The ordinary ideal is (t), so the ordinary fiber has a smooth one-dimensional function ring on this chart.

For the derived fiber, use the smooth Koszul model [1, Chs. 21–22]. Its cochain algebra is $A \otimes \bigwedge(\eta_1, \eta_2)$ with $|\eta_i| = -1$, $d\eta_1 = t$, and $d\eta_2 = tc$. The invertible change $\eta_2' = \eta_2 - c\eta_1$ gives $d\eta_2' = 0$. Multiplication by t on smooth germs is injective because $t \neq 0$ on a dense open subset of the chart. The first generator thus resolves $A/(t)$ and the second gives cohomology

$$H^0 = A/(t), \qquad H^{-1} = (A/(t))\eta_2'.$$

The remaining negative cohomology vanishes. The generator η_2' transforms by sign. The tangent calculation gives the same excess on the dual side:

$$T_x^{\mathrm{der}} = \left[\, \mathbb{R}_- \oplus \mathbb{R}_+ \xrightarrow{(a,b)\mapsto(a,ca)} \mathbb{R}^2_- \,\right].$$

The terms have cohomological degrees zero and one. The kernel is the trivial tangent line of the fixed circle. The quotient of the target by the image is the sign line, explicitly identified by $(u_1, u_2) \mapsto u_2 - cu_1$. The virtual dimension is zero. The ordinary dimension is one, and the derived equation record specifies the excess direction and its stabilizer action.

The squared readout loss provides another comparison. With the usual invariant Euclidean metric on the outputs, its local expression is

$$\ell(t, \phi) = \tfrac{1}{2}\|f(t, \phi)\|^2 = \tfrac{1}{2}t^2(1 + \cos^2\phi).$$

At $t = 0$ the Hessian has normal eigenvalue $1 + \cos^2\phi$ and tangential eigenvalue zero in the chosen coordinate basis. Thus each fixed circle is a smooth critical manifold with positive-definite Hessian on its normal line. This loss detects movement away from the circle. The excess target line survives even though the loss controls every normal displacement. At a zero, the general identity $D^2\ell = (Df)^* Df$ identifies the null directions of the Hessian with $\ker Df$, whereas the derived tangent complex encodes the cokernel in a separate degree.

3.3 A moving bias and the symmetric fiber

Take a bias $b \in P = V$ transforming by the same sign action, and set $F(b, x) = f(x) - b$. The total zero equation is regular because $D_b F = -I_V$. Its ordinary total incidence space $Z = \{(b, x) : b = f(x)\}$ is the smooth graph of f. Projection $Z \to P$ identifies with $f : X \to V$ and is singular along its fiber over zero. The parameter $b = 0$ has stabilizer G, and the readout on this slice is exactly the original rank-deficient equivariant map. In contrast, if $0 < |b_1| < 1$ and $|b_2/b_1| < 1$, the fiber has four points. The equations prescribe $\sin\theta = b_1$ and $\cos\phi = b_2/b_1$, with two choices for each angle. The derivative determinant is $-\cos\theta\sin\theta\sin\phi$, which is nonzero at all four points. At such a bias the parameter stabilizer is trivial.

The quotient also retains the type of the parameter. For this finite group, the smooth action stacks give a map $[Z/G] \to [P/G]$ as in [1, Ch. 17]. The zero-parameter orbit in the base is $BG = [\{0\}/G]$. In the classical smooth setting the pullback over this orbit is $[C/G]$, where C is the pair of fixed circles. It has the product description $C \times BG$. In derived smooth stacks the homotopy pullback over that orbit is $[Z_0^{\mathrm{der}}/G]$, with Z_0^{der} the derived zero fiber computed above. Indeed, pulling back along the atlas of BG trivializes the torsor and gives the derived incidence square on the manifolds. Gluing this calculation under the torsor's changes of frame recovers its G-action. Thus the total evaluation can be made regular by the bias, while its symmetric fiber keeps both the stabilizer and the excess equation of the fixed-parameter model.

4 Decision boundaries with parameter monodromy

The regular fibers of a score can carry a permutation representation of the parameter space's fundamental group. An invariant computed separately at each parameter can miss this information. We apply the signed-fiber, relative mapping, discriminant, and persistence constructions of [1, Chs. 9, 18, 20, and 28] to one exact network whose decision boundary has this property.

4.1 An exact rotation-covariant score

Fix an integer $k \geq 2$. Let the input and parameter spaces both be the unit circle in $\mathbb{C}$, and consider

$$s(p, z) = \operatorname{Im}(\overline{p}\, z^k). \tag{2}$$

The real and imaginary coordinates of z^k can be computed by a finite network with affine maps and the activation $x \mapsto x^2$. Indeed,

$$ab = \tfrac{1}{4}\big((a+b)^2 - (a-b)^2\big)$$

implements a real multiplication, hence the real and imaginary parts of a complex multiplication. Iterating this operation computes the indicated power. If a scalar needs to pass through an activation, the identity is $x = ((x+1)^2 - (x-1)^2)/4$. With $p = p_1 + ip_2$ the final affine layer reads out $p_1 \operatorname{Im}(z^k) - p_2 \operatorname{Re}(z^k)$. Thus p_1, p_2 may be treated as the only trainable weights, constrained to the parameter circle. The classifier thresholds the score at zero.

For $g \in S^1$, act on inputs by $z \mapsto gz$ and on parameters by $p \mapsto g^k p$. The feature z^k transforms by the character $g \mapsto g^k$, with its real two-dimensional form in the network. The final expression pairs that feature with a parameter of the same character, giving $s(g^k p, gz) = s(p, z)$. At a fixed p the rotations stabilizing the parameter are the kth roots of unity, denoted μ_k. They act on the entire zero fiber of the classifier.

4.2 Oriented sheets and their transport

Orient the input circle by increasing angle. In the coordinates $p = e^{i\theta}$ and $z = e^{i\varphi}$, the score is $\sin(k\varphi - \theta)$, with input derivative $k\cos(k\varphi - \theta)$. A boundary point is therefore regular and has positive or negative derivative according as $z^k = p$ or $z^k = -p$. Write

$$B^{\pm} = \{(p, z) : z^k = \pm p\}, \qquad B = B^+ \amalg B^-,$$

with projection $\pi : B \to P = S^1$.

Proposition 4.1 (Permutation of boundary points)**.** *Each of B^+ and B^- is a connected k-sheeted covering of P. The positive generator of $\pi_1(P)$ acts on each signed fiber by a k-cycle. Neither covering admits a continuous section.*

Proof. Projection onto z identifies each component with a circle. The projection onto p is then $z \mapsto z^k$ on the positive component and $z \mapsto -z^k$ on the negative one. Over an angular interval the roots can be written as

$$z_j^{\pm}(\theta) = \exp\left(i\,\frac{\theta + c_{\pm} + 2\pi j}{k}\right), \qquad c_+ = 0, \quad c_- = \pi,$$

for j reduced modulo k. Increasing θ by 2π sends label j to $j+1$. A continuous section would select a point fixed by this permutation. Equivalently, its map into the input circle would have a degree which, when multiplied by k, gives 1, contradicting integral degree. □

Pullback along the parameter cover $r(\alpha) = \alpha^k$ trivializes the positive component.

$$\begin{array}{ccc} S^1 \times \mu_k & \xrightarrow{\;h\;} & B^+ \\ {\scriptstyle \mathrm{pr}_1}\downarrow & & \downarrow{\scriptstyle \pi} \\ S^1 & \xrightarrow[\;r\;]{} & P \end{array}$$

Here $h(\alpha, \zeta) = (\alpha^k, \alpha\zeta)$. Choosing a fixed kth root of -1 gives the corresponding trivialization for B^-. More generally, pullback along $\alpha \mapsto \alpha^r$ for a positive integer r replaces the permutation by $j \mapsto j + r$. Its cycles have length $k/\gcd(k, r)$, and there are $\gcd(k, r)$ of them. Thus the cover with $r = k$ is the least positive winding cover that permits a global labeling of all the sheets. In each fiber the signed count is $k - k = 0$. The transport still gives two copies of the cyclic permutation representation. This persists on a range of perturbed networks. Suppose a smooth perturbed score $\widetilde{s}$ satisfies, for some $0 < \eta < 1$,

$$\|\widetilde{s} - s\|_\infty < \eta, \qquad \|\partial_\varphi(\widetilde{s} - s)\|_\infty < k\sqrt{1 - \eta^2}.$$

At a zero of any linear interpolation between the scores we have $|s| < \eta$, hence $|\partial_\varphi s| \geq k\sqrt{1-\eta^2}$. Thus zero remains regular in the input direction and its derivative sign is preserved at the interpolated zeros. The relative zero set is a proper local diffeomorphism over the cylinder of interpolation times and parameters. Path lifting along the interpolation identifies the two signed covers at the ends of the cylinder.

4.3 Persistence over the parameter space

Use the closed sublevels $F_p(t) = \{z : s(p, z) \leq t\}$. At $t < -1$ the fiber is empty. At $t = -1$ it consists of the k minima. For $-1 < t < 1$ it is a disjoint union of k closed arcs, each containing one minimum. At $t = 1$ it becomes the whole circle. Over a field $\mathbb{K}$, the degree-zero barcode therefore has one copy of $[-1, \infty)$ and $k-1$ copies of $[-1, 1)$. Degree one has one copy of $[1, \infty)$. These statements follow because inclusions before 1 preserve the components, while the inclusion into the circle sends all component classes to the same class. Each p has the same barcodes.

The total sublevels $\mathcal{F}(t) = \{(p, z) : s(p, z) \le t\} \to P$, with their inclusions for increasing t, define a diagram of spaces over P. A local lift $\alpha(p)^k = p$ simultaneously trivializes the diagram by $z = \alpha(p)u$, since the score becomes $\mathrm{Im}(u^k)$. The transitions rotate u by elements of μ_k, and transport around the base permutes the k components for every $-1 \le t < 1$. Fiberwise H_0 is therefore a local system with permutation module $\mathbb{K}[\mu_k]$ in that range. At $t = 1$ the structure map is the augmentation $\mathbb{K}[\mu_k] \to \mathbb{K}$, where $\mathbb{K}$ carries trivial transport. Its kernel gives the local system of reduced degree-zero classes that die. If the characteristic divides k, this augmentation has no equivariant section. An invariant element is a multiple of the sum of the k basis vectors, whose augmentation is zero. This extension is invisible in the barcode of each fiber, where the augmentation is a surjective map of vector spaces. Over $\mathbb{C}$, let $\omega = e^{2\pi i/k}$, with e_j the component basis of H_0. The character vector $\sum_j \omega^{-\ell j} e_j$ acquires the factor ω^ℓ under monodromy. The trivial character survives the merge at 1 and each nontrivial character dies there. The $k - 1$ numerically equal finite bars consequently carry distinct transport characters.

For parameters p, q, the chord distance $\delta = |p - q|$ bounds the uniform score change, since

$$|s(p, z) - s(q, z)| \le |\overline{p} - \overline{q}|\,|z^k| = \delta.$$

Thus $F_p(t) \subseteq F_q(t + \delta)$ and $F_q(t) \subseteq F_p(t + \delta)$. The successive inclusion maps are exactly the shift by 2δ, giving a δ-interleaving both before and after taking homology. The same proof applies to any uniform score perturbation. For q-tame persistent homology the stability theorems bound the bottleneck distance by that uniform change [4, 5]. In this rotation family the distance between fiber barcodes is zero. The sublevel diagram over P additionally records which components are identified by going around a loop. For example, the constant family of scores $s_0(p, z) = \mathrm{Im}(z^k)$ has the same barcode at every parameter. Its component local system is trivial, whereas the generator of the parameter loop acts on the components for s by the nonidentity permutation. The two filtered diagrams therefore cannot be equivalent over P. For two general scores on the same parameter and input spaces, a uniform bound on their difference also gives the shifted inclusions of the *total* sublevels over P. This makes the sublevel diagram over the parameter space a stable comparison object while still retaining the representation.

5 Reconstruction through a Euclidean latent space

Let X be a nonempty compact smooth manifold without boundary, and fix an embedding $i : X \hookrightarrow \mathbb{R}^D$. An encoder $e : X \to \mathbb{R}^m$ and ambient decoder $d : \mathbb{R}^m \to \mathbb{R}^D$ realize the reconstruction $x \mapsto d(e(x))$. Here the latent width m is the dimension of the intermediate Euclidean space. The encoder and decoder may be evaluation maps of neural architectures at selected parameters. We first allow arbitrary smooth maps to isolate the topological constraints on any such architecture. This applies the tube and factorization constructions of [1, Chs. 8, 24, 27]. For related effects of data topology on autoencoding, see [3].

5.1 Exact recovery and smooth families

Theorem 5.1 (Smooth ambient reconstruction)**.** *A smooth encoder $e : X \to \mathbb{R}^m$ admits a smooth ambient decoder on all of $\mathbb{R}^m$ satisfying $d \circ e = i$ if and only if e is a smooth embedding. Moreover e extends to a smooth encoder on all of $\mathbb{R}^D$ agreeing with it along $i(X)$.*

Proof. If $d \circ e = i$, injectivity of i forces injectivity of e. Differentiating gives $D_x i = D_{e(x)} d\, D_x e$, so $D_x e$ is injective. The compact-to-Hausdorff criterion makes e a homeomorphism onto its image. That and immersion give an embedding.

Conversely let $U \subset \mathbb{R}^m$ be a tubular neighborhood of $e(X)$ with smooth retraction $\pi : U \to e(X)$. On U the map $d_0 = i \circ e^{-1} \circ \pi$ satisfies $d_0 \circ e = i$. Choose a smooth bump χ equal to one near $e(X)$ and with compact support in U. Set $d = \chi d_0$ on U and $d = 0$ off U. The support condition makes this extension smooth, and the equality holds along the image. To extend e from $i(X)$ to $\mathbb{R}^D$, apply the same construction to $e \circ i^{-1}$ using a tube about $i(X)$. □

Consequently the least possible exact latent width in unrestricted smooth ambient reconstruction is the smooth Euclidean embedding dimension of X. Derivative rank alone does not determine that width. For example, the double traversal $z \mapsto z^2$ from S^1 to the unit circle in $\mathbb{R}^2$ is a smooth immersion, but a decoder cannot separate the two points at any encoded value. The tube construction supplies a smooth map in the ambient model class. Exact realization by a prescribed finite neural architecture requires a separate representability argument.

Corollary 5.2 (A decoder varying smoothly with parameters)**.** *Let P be a smooth manifold and $e : P \times X \to \mathbb{R}^m$ a smooth family. There exists a jointly smooth $d : P \times \mathbb{R}^m \to \mathbb{R}^D$ with $d(p, e(p, x)) = i(x)$ for all p, x if and only if each e_p is a smooth embedding.*

Proof. Necessity follows by restricting to any p. For sufficiency, the map $j(p, x) = (p, e(p, x))$ is an injective immersion, since a vector in its derivative kernel has zero parameter component and then zero X-component. The preimage of a compact set K is closed in the compact set $\mathrm{pr}_P(K) \times X$. Thus j is proper and is a closed embedding. Take a smooth tubular retraction π to its image in $P \times \mathbb{R}^m$. On this tube use $i \circ \mathrm{pr}_X \circ j^{-1} \circ \pi$. A smooth cutoff equal to one near the closed image and supported in the tube

lets us extend this map by zero. The cutoff can be chosen by a partition of unity subordinate to the tube and the complement of the image, with locally finite supports when P is noncompact. Its value on $j(p,x)$ is $i(x)$, as required. □

5.2 Uniform approximation and a sharp spherical bound

The existence of a uniform error gap extends beyond exact recovery. Suppose $\dim X = n > 0$. Choose a tubular neighborhood $U \subset \mathbb{R}^D$ of $i(X)$ and a number $\delta > 0$ such that all points at distance less than δ from $i(X)$ belong to U. Such a number exists by compactness.

Proposition 5.3 (A closed-manifold error gap)**.** *If $m \leq n$, every continuous encoder and ambient decoder satisfy $\sup_{x\in X} \|d(e(x)) - i(x)\| \geq \delta$.*

Proof. If the supremum were less than δ, the straight segment from $i(x)$ to $d(e(x))$ would lie in U for every x. Put $V = d^{-1}(U)$, an open subset of $\mathbb{R}^m$ containing $e(X)$, and let $r : U \to X$ be the tubular retraction with $r \circ i = \mathrm{id}_X$. Applying r to those segments gives a homotopy to the identity for the composite

$$X \xrightarrow{e} V \xrightarrow{d|_V} U \xrightarrow{r} X.$$

With coefficients in $\mathbb{F}_2$, the identity on $H_n(X)$ would then factor through $H_n(V)$. For $m > 0$, every component of V is a noncompact m-manifold, whose ordinary homology vanishes in degrees at least m [7, Prop. 3.29]. For $m = 0$ its positive-degree homology also vanishes. The mod-two fundamental class of any component of X is nonzero, yielding a contradiction. □

The use of the open latent subset V is essential because the decoder may leave the tubular neighborhood elsewhere in latent space. The argument applies to continuous layers, including piecewise smooth realizations. It bounds uniform error for an entire data manifold and therefore also bounds each eligible member of a parameter family.

Proposition 5.4 (Precise bounds on the unit sphere)**.** *Take $X = S^n$ and let $i : S^n \hookrightarrow \mathbb{R}^{n+1}$ be the standard unit sphere, with $n \geq 1$. For $m \leq n$, the optimal uniform error among continuous encoders and ambient decoders is 1. For $m \geq n+1$, exact ambient recovery is possible. If the decoder is constrained to take values in S^n on all of $\mathbb{R}^m$, the uniform error is 2 for every width m.*

Proof. For $m \leq n$, apply Borsuk–Ulam to the composite of the encoder and the coordinate inclusion of $\mathbb{R}^m$ in $\mathbb{R}^n$. It gives an x with $e(x) = e(-x)$ [9]. If u is their common decoded value, $2 \leq \|x-u\| + \|-x-u\|$. The zero decoder attains error 1. For $m \geq n+1$, a coordinate inclusion and projection give exact recovery. Finally a sphere-valued reconstruction through all of $\mathbb{R}^m$ is nullhomotopic. If it avoided antipodal coincidences, its straight segments to the identity, normalized to unit length, would give a homotopy to the identity. This would contradict its zero degree. Hence it reaches error 2, the largest distance between points of the sphere. □

The optimal errors for the unit sphere are summarized below.

Decoder range	$m \leq n$	$m \geq n+1$
Ambient $\mathbb{R}^{n+1}$	1	0
Sphere S^n	2	2

For a sphere of radius $R > 0$, rescaling replaces the positive entries by R and $2R$, respectively. The bound holds for every parameter of an appropriate family. Its proof chooses an antipodal pair separately at each parameter and needs no continuous choice of witnesses.

Uniform and average losses have different implications. For example, at width n choose stereographic coordinates σ from the complement of the north pole of S^n to $\mathbb{R}^n$. Multiply σ by a smooth cutoff equal to one outside a small polar cap C and equal to zero near the pole, extending the encoder by zero there. The smooth decoder $d = \sigma^{-1}$ gives exact recovery off C. For normalized spherical measure μ, the squared error averaged over the sphere is at most $4\mu(C)$, which can be made arbitrarily small while the uniform error is 2. Thus the uniform obstruction does not force a positive lower bound on this average across all smooth maps of width n.

A uniform Lipschitz bound prevents this concentration. If the reconstruction has Lipschitz constant at most L with respect to chord distance, its error function has Lipschitz constant at most $1 + L$. For ambient width $m \leq n$, some point has error at least 1, so every point within chord distance $1/(2(1+L))$ of it has error at least $1/2$. Its mean squared error is at least one quarter of the normalized measure of such a cap, a positive bound at fixed L. This estimate states the additional analytic hypothesis under which the pointwise obstruction controls a population loss.

6 Conclusion

The analysis separates three conditions on a parametrized network. The available weight directions determine where the total jet map is transverse. The symmetry representation determines which normal derivatives a fixed realization can have at a stabilizer. The topology of the domain and latent factorization determines which global reconstructions are possible. A model can satisfy one of these conditions while exhibiting an obstruction governed by another.

Examples on the circle and torus make these distinctions explicit. Ordinary point counts jump at the fold

discriminant. They can also remain constant while boundary sheets are permuted along a regular parameter loop. The resulting persistence module over the parameter retains that permutation representation. In the equivariant torus model the squared residual controls normal movement, while the tangent complex records an excess target direction. The autoencoder estimates bound uniform reconstruction error and give a separate mean-square bound under a Lipschitz constraint.

Relative evaluation and descent explain how these calculations pass between local presentations of a model. The assumptions used for the calculations are geometric: smoothness of the relevant charts, specified perturbation directions, properness for topology of fibers, and an explicit choice of decoder range. They apply to a realized architecture once those hypotheses have been checked.

References

[1] Anonymous. *Differential Topology in Higher Categories: From Regular Values and Cobordism to Derived Smooth Families*. Submitted manuscript, 2026.

[2] Anonymous. Representation theory of parametrized neural models on spaces. *Higher Categorical Semantics*, 2026. Published article.

[3] J. Batson, C. G. Haaf, Y. Kahn, and D. A. Roberts. Topological obstructions to autoencoding. *Journal of High Energy Physics* 04 (2021), 280. DOI.

[4] F. Chazal, V. de Silva, M. Glisse, and S. Oudot. *The Structure and Stability of Persistence Modules*. Springer, 2016. DOI.

[5] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete & Computational Geometry* 37 (2007), 103–120. DOI.

[6] B. Fong, D. I. Spivak, and R. Tuyéras. Backprop as functor: A compositional perspective on supervised learning. In *Proceedings of LICS*, 2019. Author manuscript.

[7] A. Hatcher. *Algebraic Topology*. Cambridge University Press, 2002. Author's copy.

[8] M. W. Hirsch. *Differential Topology*. Graduate Texts in Mathematics 33. Springer, 1976. Publisher's edition.

[9] J. Matoušek. *Using the Borsuk–Ulam Theorem*. Springer, 2003. DOI.

[10] G. Naitzat, A. Zhitnikov, and L.-H. Lim. Topology of deep neural networks. *Journal of Machine Learning Research* 21(184) (2020), 1–40. Journal article.