SYMMETRY DESCENT NEURAL OPERATORS: FROM LOCAL EQUIVARIANCE TO GLOBAL CONSISTENCY

Anonymous authors
Paper under double-blind review

ABSTRACT

Equivariant neural networks commonly assume that a group action and its representations are known and consistent over the entire domain. In heterogeneous materials and around local defects, useful symmetries and coordinate frames can vary across regions. Even where transport between frames is locally flat, a loop around a hole can have nontrivial holonomy. Local equivariance alone does not ensure that learned operator layers agree across interfaces or remain compatible after transport around a loop. We propose Symmetry Descent Neural Operators (SDNO), which treat local symmetries and frame alignment as episode-conditioned structures determined from operator observations. Given multiple input–output observations from the same operator, SDNO learns the feature transformations that can be shared within each local region and the feature alignment between neighboring regions. These constraints define a cross-region weight-sharing space, while curvature and holonomy represent local defects and global topology, respectively. A differentiable soft projection maps trainable weights toward this constraint space. Our theory characterizes exact global compatibility and approximate descent, and analyzes local-symmetry recovery under a spectral gap. The experiments cover local transport with the unit complex phase group $\mathrm{U}(1)$ and the noncommutative special unitary group $\mathrm{SU}(2)$, operator prediction and topology transfer under hidden local transport, polycrystalline local frames, and a public Structure-from-Motion rotation averaging benchmark. SDNO maintains accurate structure recovery and operator prediction under unseen changes in holonomy, curvature, topology, and local frames, and improves global orientation estimation on public noisy relative-rotation data.

1 INTRODUCTION

Neural operators such as DeepONet (Lu et al., 2021) and FNO (Li et al., 2021) learn mappings between function spaces for families of problems with varying coefficients, boundary conditions, or geometries (Kovachki et al., 2023). ICON (Yang et al., 2023b) uses input–output examples from one system to condition predictions at test time. In our setting, an episode fixes the geometry, medium, and hidden connection but provides responses to several distinct inputs. These repeated probes help distinguish symmetries of the operator from incidental invariances of one response, and provide evidence for aligning neighboring local coordinate frames.

Group-equivariant CNNs (Cohen & Welling, 2016) and EMLPs (Finzi et al., 2021) share weights under specified group actions and representations. Gauge-equivariant CNNs (Cohen et al., 2019) prescribe how features transform between local frames; GINO (Cheng, 2026) enforces frame consistency for intrinsic elliptic PDE maps. AtlasD (Bhat et al., 2025) estimates local symmetry generators from data. Recovering local generators alone does not establish globally shared layers: on an overlap, the maps on neighboring patches must agree after transporting both input and output features.

Polycrystalline materials illustrate this compatibility condition: each grain has its own frame, and a local map must agree with its neighbor after feature transport across a boundary. Defects can produce face-level curvature, while a domain with holes can have nontrivial loop holonomy even where

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.

transport is locally flat. The available globally shared weights therefore depend on the episode's connection, even if each patch separately admits equivariant layers.

SDNO estimates local Lie subspaces and their input and output representations from Jacobian–vector products of the observed operator, then infers group-valued transport between patches. Local commutation and cross-patch compatibility give linear constraints on network weights. A differentiable soft projection maps network weights toward this constraint space during training, allowing gradients to update both the network and the estimated structure. An equivariant neural corrector models the episode-conditioned residual.

Locally admissible weights form parameter spaces whose edge restrictions encode transport; their global sections are the compatible shared layers. This cellular-sheaf view and categorical equivariance lead to the same sparse constraint system. Loop holonomy can reduce the dimension of the global parameter space. For $\mathrm{U}(1)$ transport, discrete Hodge decomposition separates frame changes, curvature, and loop holonomy; curvature and remaining symmetry breaking enter the prediction branches.

Our contributions are as follows.

- We jointly estimate local Lie algebras, input and output representations, and group-valued transport from operator observations, then enforce their constraints with a differentiable projection and an episode-conditioned corrector.
- We characterize exact descent, holonomy obstruction, gauge-covariant soft projection, approximate-descent stability, and identifiability up to gauge; proofs are in the appendix.
- We evaluate controlled connection recovery, hidden-connection operator transfer, and polycrystalline local frames, as well as public Structure-from-Motion $\mathrm{SO}(3)$ rotation averaging.

Implementation, protocols, cached results, and reproduction scripts are available in an anonymous OSF repository; Appendix G describes the artifact.

2 Related Work

Neural operators and episode conditioning. DeepONet (Lu et al., 2021) and FNO (Li et al., 2021) approximate maps between function spaces through branch–trunk networks and Fourier-domain kernels. OFormer (Li et al., 2023) and GNOT (Hao et al., 2023) use attention for operator learning, while ICON (Yang et al., 2023b) conditions predictions on input–output examples from the target operator. SDNO uses such examples to infer the symmetry representations and transport that constrain an episode's operator layers.

Geometric and topological operator models. Discrete exterior calculus and Hodge theory supply incidence and Laplacian operators for fields on cell complexes (Desbrun et al., 2005; Lim, 2020). HSD (Zheng et al., 2026) uses Hodge decomposition to separate topological and geometric components; TNO (Bastian et al., 2026) couples features on cells of different dimensions; and CSNO (Shikhman & Gilbertie, 2026) learns cellular-sheaf restrictions for constrained PDE surrogates. GINO (Cheng, 2026) builds gauge equivariance into intrinsic elliptic PDE maps, while HilbNet (Tandon et al., 2026) relates continuous Hilbert-bundle connections to sampled cellular sheaves. These models provide geometric structure for operator learning; SDNO additionally estimates the local symmetry representations that constrain its network layers.

Equivariant architectures and symmetry discovery. Group-equivariant CNNs (Cohen & Welling, 2016), EMLPs (Finzi et al., 2021), and gauge-equivariant CNNs (Cohen et al., 2019) construct layers for specified actions or local frame changes. G-FNO (Helwig et al., 2023), InfGCN (Cheng & Peng, 2023), and EGNO (Xu et al., 2024) encode equivariance in Fourier PDE layers, continuous 3D maps, and trajectory operators, respectively. LieGAN (Yang et al., 2023a) discovers linear group actions from data; LaLiGAN (Yang et al., 2024) learns nonlinear actions through a latent space; and AtlasD (Bhat et al., 2025) identifies symmetries of local predictors. SDNO jointly estimates local input and output representations and overlap transport from observations of one operator, then constrains the network layers with both local and cross-patch relations.

Sheaves, connections, and synchronization. Cellular sheaves encode local consistency through stalks, restriction maps, and global sections (Hansen & Ghrist, 2019). Sheaf Neural Networks (Hansen & Gebhart, 2020) use sheaf Laplacians for message passing, and Neural Sheaf Diffusion (Bodnar et al., 2022) learns restriction maps. Vector diffusion maps (Singer & Wu, 2012) and group synchronization (Gao et al., 2021) analyze alignment from group-valued edge relations; Torsor CNNs (Li et al., 2026) use edge potentials in locally equivariant graph convolutions. Robust IRLS (Chatterjee & Govindu, 2018), Shonan (Dellaert et al., 2020), and NeuRoRA (Purkait et al., 2020) estimate global orientations from noisy relative rotations. CENN (Maruyama, 2025) formalizes equivariance as naturality across categories. In SDNO, stalks contain locally admissible network layers and learned input and output transport induces cross-patch constraints; curvature and loop holonomy can reduce the compatible parameter space.

3 Problem Setup

In geometric deep learning, a local frame determines the coordinate representation of feature channels. Before neighboring regions exchange information, their feature channels must be aligned by group-valued transport. The composition of transport along a face boundary gives a face-level cycle-consistency residual, represented here as curvature; composition along a noncontractible loop gives holonomy. Gauge-equivariant networks usually prespecify the group representations and transport. We treat them as hidden structures that vary by episode, estimate them from input–output observations, and then use them to constrain local network layers and cross-region parameter sharing.

Let $\mathcal{K}$ be a finite cellular complex, and let $\{U_i\}_{i=1}^m$ be m local patches covering its underlying space. Episode e fixes an operator $\mathcal{S}_e$, geometry, medium, and hidden connection. It provides N_e input–output observations, collectively denoted by $\mathcal{D}_e$. Each observation (a, u) satisfies $u = \mathcal{S}_e(a)$, where a is an input field and u is the output field produced by the operator for that input. In the PDE and material tasks, a corresponds to an external excitation or source term, and u corresponds to the resulting physical field; we call u the response field in these tasks. The model uses $\mathcal{D}_e$ to predict u for a new input and to identify the local structure of that episode.

The input and output fibers on patch U_i are denoted by V_i^{in} and V_i^{out}, respectively. The local Lie algebra is $\mathfrak{g}_i$. Its representations on the input and output fibers are denoted by $\rho_i^s : \mathfrak{g}_i \to \mathrm{End}(V_i^s)$, where $s \in \{\mathrm{in}, \mathrm{out}\}$ and $\mathrm{End}(V)$ denotes the space of linear endomorphisms of V. Neighboring patches have invertible transport maps

$$T_{ij}^{\mathrm{in}} : V_i^{\mathrm{in}} \to V_j^{\mathrm{in}}, \qquad T_{ij}^{\mathrm{out}} : V_i^{\mathrm{out}} \to V_j^{\mathrm{out}}. \tag{1}$$

These objects are identifiable only up to local coordinate transformations. We therefore measure structure error using gauge invariants or metrics after gauge alignment.

Locally admissible operators. Let $W_i : V_i^{\mathrm{in}} \to V_i^{\mathrm{out}}$ be the local network layer on patch i. The space of locally admissible operators is defined as

$$\mathcal{A}_i := \{W_i : V_i^{\mathrm{in}} \to V_i^{\mathrm{out}} : W_i \rho_i^{\mathrm{in}}(X) = \rho_i^{\mathrm{out}}(X) W_i,\ \forall X \in \mathfrak{g}_i\}. \tag{2}$$

Neighboring local layers must also be compatible with the learned transport:

$$T_{ij}^{\mathrm{out}} W_i = W_j T_{ij}^{\mathrm{in}}. \tag{3}$$

Equation (3) requires that applying the local layer and then changing coordinates gives the same result as changing coordinates first and then applying the neighboring local layer.

Compatible parameters across patches. Let $\mathcal{A} := \{\mathcal{A}_i\}_{i=1}^m$ denote all locally admissible operator spaces. The parameter space satisfying local equivariance and all edge-compatibility conditions is

$$\Gamma(\mathcal{A}) = \left\{(W_i)_{i=1}^m :\ W_i \in \mathcal{A}_i,\ T_{ij}^{\mathrm{out}} W_i = W_j T_{ij}^{\mathrm{in}},\ \forall (i,j)\right\}. \tag{4}$$

In the implementation, we stack the local commutation and edge-compatibility relations into sparse linear constraints, and then parameterize the network weights using the differentiable projection in Section 4. When the transport is flat, Equation (4) is the space of global sections of the operator local system. Equation (3) also gives the naturality condition between the input and output local systems. Appendix A gives the corresponding sheaf-theoretic and categorical formulations.

4 SYMMETRY DESCENT NEURAL OPERATORS

4.1 DISCOVERING LOCAL SYMMETRIES FROM OPERATOR DATA

An infinitesimal generator gives the tangent direction to a group orbit at the current sample. If an operator preserves this symmetry, the tangent vector of the input orbit, after mapping by the operator Jacobian, should agree with the tangent vector of the output orbit. This condition converts local symmetry discovery into a linear measurement that can be computed with a Jacobian–vector product.

For an input a, let $D\mathcal{S}_a$ be the Fréchet derivative of the nonlinear operator $\mathcal{S}$ at a. Given $X \in \mathfrak{g}_i$, its input and output generators are $A_i^{\mathrm{in}} := \rho_i^{\mathrm{in}}(X)$ and $A_i^{\mathrm{out}} := \rho_i^{\mathrm{out}}(X)$, respectively. The infinitesimal equivariance condition is

$$D\mathcal{S}_a[A_i^{\mathrm{in}} a] = A_i^{\mathrm{out}} \mathcal{S}(a), \tag{5}$$

Coordinate actions can be included in the same residual through a Lie derivative. We first fit a surrogate $\widehat{\mathcal{S}}$, and then use Jacobian–vector products (JVPs) to evaluate Equation (5). Stacking the residuals from multiple observations gives a local linear measurement operator $\widehat{\mathcal{M}}_i$; its small-singular-value subspace gives a candidate $\mathfrak{g}_i$. The candidate dimension is selected jointly by the spectral gap, metric preservation, Lie-bracket closure, and orbit residuals on independent observations. Optimization is performed on the Grassmannian and therefore does not depend on the choice of generator basis.

4.2 TRANSPORT, CURVATURE, AND HOLONOMY

Different patches can recover the same Lie subspace while using different generator bases and local frames. Transport aligns these local descriptions. Composing transport along a face boundary gives a local cycle-consistency residual; composing transport along a noncontractible loop retains a global degree of freedom that cannot be removed by a local frame change.

Transport between neighboring patches is estimated within a candidate structure group. The objective requires the transported local Lie subspaces to align under a Grassmannian distance and acts separately on the input and output representations. The candidate group can be U(1), SO(2), or a corresponding unitary representation; Appendix F gives the objective. For each oriented 2-cell f, let ∂f be its oriented boundary and let T_e be the transport on edge e, oriented along the boundary. The face holonomy is

$$H_f = \prod_{e \in \partial f} T_e. \tag{6}$$

The deviation of H_f from the identity represents both the face-level cycle-consistency residual and the local curvature. The product of transport along a noncontractible loop gives the global holonomy. Both are evaluated through conjugacy invariants.

4.3 DIFFERENTIABLE CONSTRAINT PROJECTION

The learned structure ultimately acts on the network parameter space. Local equivariance, cross-patch compatibility, and physical constraints can all be written as linear relations on the weights. The null space of the corresponding constraint matrix contains all weights satisfying these relations. The model first optimizes unconstrained weights and then projects them onto this null space. The soft projection is differentiable with respect to both the weights and the constraint matrix, so structure estimation can be trained end to end with the neural operator.

Let $\mathrm{vec}(W)$ denote the column-wise vectorization of a matrix W. Equation (2) can be written as $C_i^{\mathrm{sym}} \mathrm{vec}(W_i) = 0$, where C_i^{sym} is the local symmetry constraint matrix. We stack it with transport compatibility, conservation, boundary, and constitutive constraints into a total matrix C_i. Let $\widetilde{W}_i$ be the unconstrained weight, let $\tau > 0$ be a regularization parameter, let C^* be the Hermitian adjoint of C, and let I be the identity matrix. The constrained network layer is

$$\mathrm{vec}(W_i) = P_\tau(C_i)\, \mathrm{vec}(\widetilde{W}_i), \qquad P_\tau(C) = I - C^*(CC^* + \tau I)^{-1} C. \tag{7}$$

$P_\tau(C)$ is a differentiable soft null-space projection. The parameter τ controls the constraint residual and numerical stability. This layer can be trained jointly with a conventional neural operator; rank continuation is used to stabilize changes in the candidate algebra dimension.

4.4 Descent, Defect, and Ambient Branches

SDNO separates the estimated connection into a compatible flat component and a curvature residual. The flat component defines a local system, while the curvature residual retains face-level cycle-consistency residuals. Denote the descent, defect, and ambient branches by $\mathcal{S}_{\mathrm{descent}}$, $\mathcal{S}_{\mathrm{defect}}$, and $\mathcal{S}_{\mathrm{ambient}}$, respectively. Let $\widehat{\mathcal{S}}$ be the predicted operator. Then

$$\widehat{\mathcal{S}} = \mathcal{S}_{\mathrm{descent}} + \mathcal{S}_{\mathrm{defect}} + \mathcal{S}_{\mathrm{ambient}}. \tag{8}$$

The descent branch uses a flat local system satisfying the cocycle condition. The defect branch receives face curvature, naturality residuals, and symmetry-breaking statistics. The ambient branch is an equivariant neural corrector. It aggregates the input–output observations from the same episode into a gauge-invariant context covariance, and a hypernetwork then generates channel-mixing matrices. Complex-linear propagation preserves global U(1) phase equivariance.

The neural parameters of the ambient branch are shared across episodes. The connection, medium, and scalar coefficients are estimated within each test episode by the descent and defect branches.

The two-dimensional U(1) case gives a direct parameterization of the three branches. Let $\mathbb{R}$ denote the real numbers, and let $C^k(\mathcal{K};\mathbb{R})$ denote the real-valued k-cochains. Let $\delta_0 : C^0(\mathcal{K};\mathbb{R}) \to C^1(\mathcal{K};\mathbb{R})$ and $\delta_1 : C^1(\mathcal{K};\mathbb{R}) \to C^2(\mathcal{K};\mathbb{R})$ be consecutive coboundary operators, and let δ_1^* be the Hermitian adjoint of δ_1. Let ϕ be the vertex phase, ψ the face potential, and h a harmonic 1-cochain. After fixing a lift sector, the edge angle $a \in C^1(\mathcal{K};\mathbb{R})$ decomposes as

$$a = \delta_0\phi + \delta_1^*\psi + h. \tag{9}$$

The three terms correspond to local frames, face curvature, and holonomy on noncontractible loops, respectively. Spanning-tree gauge fixing fixes the lift sector and avoids discrete 2π jumps induced by branch cuts. General twisted complexes and the noncommutative case are given in Appendix A.

4.5 Training Procedure

Training has four steps. First, we fit a surrogate for computing Jacobians. Second, we recover local Lie subspaces from JVP measurements and estimate group-valued transport. We then compute curvature, cycle holonomy, and naturality residuals, and construct $P_\tau(C_i)$ and the three-branch features. Finally, we train the equivariant neural corrector and jointly fine-tune the predictor under the fixed structural parameterization. The structure dimension is updated on a slower time scale and selected jointly by residuals on independent observations and the spectral gap.

5 Theoretical Properties

The theory considers finite connected cellular complexes, finite-dimensional unitary fibers, and transport satisfying representation compatibility. Flat transport is used in the descent branch, while the curvature residual of the original connection enters the defect branch. Appendix A gives the assumptions, auxiliary lemmas, and proofs.

Below, $\|A\|_{\mathrm{op}}$ denotes the operator norm of a linear map A.

Theorem 5.1 (Exact descent and holonomy obstruction)**.** *Under flat input and output transport and overlap naturality, the locally admissible operators form a local system and induce a linear map from global input sections to global output sections. Fix a base patch i_0, let $\pi_1(\mathcal{K}, i_0)$ be the fundamental group based at i_0, and let $\mathrm{Hol}^s(\gamma)$ be the ordered product of input or output transport along a closed path $\gamma \in \pi_1(\mathcal{K}, i_0)$, where $s \in \{\mathrm{in}, \mathrm{out}\}$. Let $W_0 := W_{i_0}$. Then the global architecture space satisfies*

$$\Gamma(\mathcal{A}) \cong \{W_0 \in \mathcal{A}_{i_0} : \mathrm{Hol}^{\mathrm{out}}(\gamma) W_0 = W_0 \,\mathrm{Hol}^{\mathrm{in}}(\gamma),\ \forall \gamma \in \pi_1(\mathcal{K}, i_0)\}. \tag{10}$$

Equation (10) characterizes the weight directions that remain consistent after transport along every loop.

Theorem 5.2 (Gauge-covariant soft projection)**.** *For a constraint operator C covariant under unitary fiber-frame changes, the projection in Equation* (7) *transforms by conjugation. If $P_{\ker C}$ projects onto the null space of C and γ is the smallest nonzero singular value of C, then*

$$\|P_\tau(C) - P_{\ker C}\|_{\mathrm{op}} \le \frac{\tau}{\gamma^2 + \tau}, \qquad \|C P_\tau(C)\|_{\mathrm{op}} \le \frac{\sqrt{\tau}}{2}. \tag{11}$$

Thus the projected layer transforms covariantly under a change of local feature frame, with approximation error controlled by τ and the spectral gap γ.

These results separate three compatibility checks on a candidate layer. Local commutation restricts the maps available in each patch, edge naturality asks whether those maps agree after transport between patches, and loop holonomy checks whether the agreement survives a full circuit. For one-dimensional complex fibers with input and output loop phases $e^{\mathrm{i}\theta_{\mathrm{in}}}$ and $e^{\mathrm{i}\theta_{\mathrm{out}}}$, Equation (10) reduces to $(e^{\mathrm{i}\theta_{\mathrm{out}}} - e^{\mathrm{i}\theta_{\mathrm{in}}})W_0 = 0$. A nonzero scalar map therefore requires matching phases on every loop, even when the local tests pass and all faces are flat. The soft projection approximates the local and edge constraints; the next theorem bounds the resulting global output deviation using the edge-naturality residual.

Theorem 5.3 (Approximate descent)**.** *Let $\Gamma_T(V^{\mathrm{out}})$ denote the space of global sections compatible with the output transport T, and let $W := \mathrm{diag}(W_i)$ denote the local operator acting patch-wise. Let λ_+^{out} be the smallest positive eigenvalue of the output twisted Laplacian, let $d_{\max}$ be the maximum out-degree of the overlap graph, and define the maximum edge naturality residual by $\nu := \max_{e:i\to j} \left\|W_j T_e^{\mathrm{in}} - T_e^{\mathrm{out}} W_i\right\|_{\mathrm{op}}$. Let $\mathrm{dist}(y, \Gamma_T(V^{\mathrm{out}}))$ denote the distance from y to this section space. For any compatible input section $x = (x_i)$ satisfying $x_j = T_e^{\mathrm{in}} x_i$,*

$$\mathrm{dist}(Wx, \Gamma_T(V^{\mathrm{out}})) \le \sqrt{\frac{d_{\max}}{\lambda_+^{\mathrm{out}}}}\, \nu \, \|x\| \,. \tag{12}$$

where $Wx = (W_i x_i)_i$, and the distance is induced by the direct-sum fiber norm. In the U(1) case, if a is an edge-angle 1-cochain, then $\delta_1 a$ is its face curvature, and the curvature correction is controlled by the $\|\delta_1 a\|$ terms of the input and output connections.

The bound controls global deviation by the edge-compatibility residual and the overlap-graph spectral gap; face curvature enters the defect branch.

Identifiability up to gauge. Under persistent excitation, let $G_i = \ker \mathcal{M}_i$ have gap $\gamma_i > 0$, and let $\delta_i := \left\|\widehat{\mathcal{M}}_i - \mathcal{M}_i\right\|_{\mathrm{op}} < \gamma_i/2$. The recovered $\dim G_i$-dimensional singular subspace $\widehat{G}_i$ obeys $\left\|P_{\widehat{G}_i} - P_{G_i}\right\|_{\mathrm{op}} \le 2\delta_i/\gamma_i$. For connected structure group G, flat connections up to gauge are classified by $\mathrm{Hom}(\pi_1(\mathcal{K}), G)/G$. Full statements and proofs appear in Appendix A.

6 Experiments

6.1 Evaluation Setup

The experiments include four classes of problems: controlled U(1) and SU(2) connection recovery, operator prediction and topology transfer under hidden connections, polycrystalline local-frame and response prediction, and $\mathrm{SO}(3)$ rotation averaging on public Structure-from-Motion data (Wilson & Snavely, 2014; Chatterjee & Govindu, 2013; Dellaert et al., 2020). The first three task classes provide each test episode with several input–output observation pairs from the same operator and evaluate predictions on unseen inputs. Given a test input a, let $u = \mathcal{S}_e(a)$ be the ground-truth output field and $\widehat{u}$ be the model prediction. In the corresponding physical tasks, these are called the ground-truth and predicted response fields, respectively. The prediction metric is the relative ℓ_2 error $\|u - \widehat{u}\|_2 \,/\, \|u\|_2$; structure recovery uses curvature, the conjugacy class of holonomy, or aligned local-frame error. We refer to test conditions in which system variables or domain geometry fall outside the training distribution as out-of-distribution (OOD) evaluation. All context-conditioned methods receive the same observations on a given task, while query-only rows retain the native input interface of the corresponding architecture. Data generation, splits, training configurations, and supplementary results are provided in Appendix F.

Prediction baselines include unconstrained local blocks, generic and equivariant GNNs, FNO-2d (Li et al., 2021), DeepONet (Lu et al., 2021), a Dilated ResNet, and an Operator Transformer. The last two implementations are specified in Appendix F. Rotation averaging uses spanning-tree initialization, Weiszfeld (Hartley et al., 2011), robust IRLS (Chatterjee & Govindu, 2018), Shonan (Dellaert et al., 2020), and NeuRoRA (Purkait et al., 2020). The two response-prediction tasks are summarized in Table 1.

Table 1: Median relative ℓ_2 prediction error on unseen inputs (↓). Standard is in-distribution in both panels. In (a), holonomy and curvature change the hidden connection, while disk and two holes change domain topology. In (b), the new grain-boundary density is out of distribution.

(a) Nonlinear U(1) operator transfer

Condition	Unconstrained complex blocks	Generic GNN	Fixed-equivariant GNN	**SDNO**
Standard	0.113	0.102	0.131	**0.019**
Holonomy	0.115	0.105	0.133	**0.020**
Curvature	0.114	0.101	0.131	**0.019**
Disk	4.94	0.405	0.588	**0.021**
Two holes	4.67	0.384	0.441	**0.028**

(b) Polycrystalline response prediction

Method	Standard (ID)	OOD density
FNO-2d	0.267	0.261
Dilated ResNet	0.223	0.228
DeepONet	0.629	0.648
Operator Transformer	0.300	0.291
SDNO	**0.188**	**0.184**

6.2 Hidden-Connection Recovery

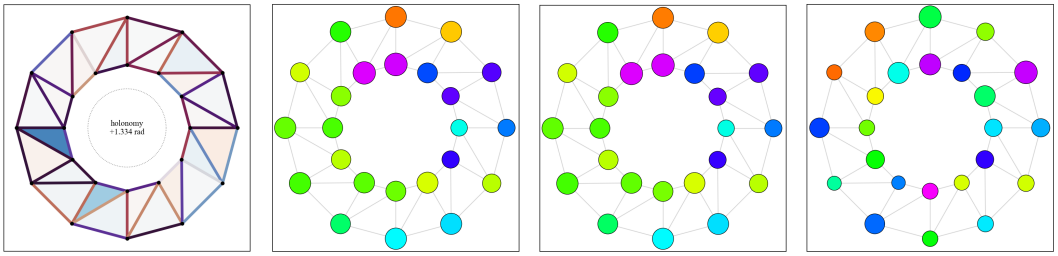
The controlled U(1) task varies the local gauge, noncontractible-loop holonomy, and sparse face curvature independently on a triangulated annulus. SDNO recovers flat transport and two curvature defects from 16 observation pairs. On unseen inputs, its median relative prediction error is 0.024, compared with 3.76 for unconstrained complex blocks. The latter fit a separate complex matrix on each edge and cannot identify shared edge transport under the same observation budget.

Under a hidden gauge, the median holonomy and curvature errors are 0.034 and 0.028 radians, respectively. In the noncommutative experiment, a central Wilson loop, a generic Wilson loop, and nonzero curvature restrict the compatible symmetry algebra to dimensions $4/2/1$, respectively; the estimated transport recovers this dimensional change, with a median Wilson-loop trace-angle error of 0.020 radians. Spectral recovery, perturbation-bound, and symmetry-misspecification experiments are given in Appendix F.

6.3 Operator Prediction and Topology Transfer under Hidden Connections

Table 1(a) compares transfer of the nonlinear U(1) operator to new holonomy, curvature, and topology. SDNO maintains relative errors of 0.019–0.028 over the five evaluation axes. The unconstrained complex blocks fit edge couplings independently, so a new topology changes both the number of parameters and the edge correspondence. The generic GNN aggregates only neighboring features, while the fixed-equivariant GNN uses a single prespecified representation; neither contains episode-conditioned edge transport and loop states.

On the two-hole complex, SDNO adapts to a topology with first Betti number $b_1 = 2$ without retraining. The holonomies of the two generating loops are recovered from $1.59/-0.80$ to $1.61/-$



(a) Recovered connection (b) True response (c) SDNO prediction (d) Unconstrained blocks

Figure 1: U(1) connection recovery and prediction on an unseen input under a hidden local gauge. (a) Edge transport and face curvature estimated by SDNO, with edge color encoding transport phase and face color encoding curvature. (b)–(d) Ground-truth response field, SDNO prediction, and unconstrained-complex-block prediction, respectively; node color and size encode the phase and magnitude of the response field.

Table 2: Median relative response error and global phase-equivariance error for a hidden connection in the magnetic Schrödinger task. Shared neural parameters are trained across episodes; episode parameters are fitted from 16 test-time observation pairs. Lower is better.

Method	Shared neural params.	Episode params.	In-distribution	New holonomy	New curvature defects	Equivariance error
FNO-2d (query)	596,450	0	0.443	0.607	0.454	0.265
FNO-2d (context)	598,498	0	0.486	0.599	0.508	0.207
DeepONet (query)	305,090	0	0.871	0.946	0.877	0.745
DeepONet (context)	4,237,250	0	0.994	0.996	0.995	1.151
Dilated ResNet (context)	211,154	0	0.616	0.622	0.622	0.157
Operator Transformer (context)	315,362	0	0.501	0.629	0.514	0.139
SDNO	**21,926**	881	**0.036**	**0.039**	**0.036**	**0.000**

0.84 radians, while the locations and signs of the local curvature are also preserved (Appendix Figure 13).

The magnetic Schrödinger task compares FNO, DeepONet, Dilated ResNet, and Operator Transformer under the same observation interface. Each test episode provides 16 input–output observation pairs and one new forcing; the new-holonomy and new-curvature tests change the annular flux range and the number of local flux tubes, respectively. Detailed network and training configurations are given in Appendix F.

In Table 2, SDNO uses $21{,}926$ shared neural parameters and 881 test-time structure parameters and attains $0.036/0.039/0.036$ on the three evaluation axes. FNO has an in-distribution error of 0.443, while Operator Transformer obtains 0.501–0.629 across the three axes.

The shared grid kernels of FNO and Dilated ResNet cannot distinguish episodes with identical coordinates but different connections, and therefore learn averaged phase couplings. Operator Transformer supports variable vertex sets but must recover edge transport and loop states implicitly within attention. SDNO estimates these quantities explicitly and models the residual operator component with shared neural parameters. Cross-task and cross-topology results are given in Appendix Table 6.

With 24 observation pairs, the held-out response error on the magnetic Schrödinger annulus is 0.037, and the correlation between the flux-scan curve and the ground truth is 0.94. Figure 2 shows recovery of response phase, magnitude, and face-wise magnetic flux; annular flux is represented by holonomy, while local flux tubes correspond to nonzero curvature.

6.4 Polycrystalline Local Frames and Response Prediction

The polycrystalline task estimates grain frames and weakened grain boundaries from anisotropic responses. In Table 1(b), SDNO obtains the lowest error both in distribution and under a new grain-boundary density. Figure 3 shows the corresponding displacement fields from four observation pairs.

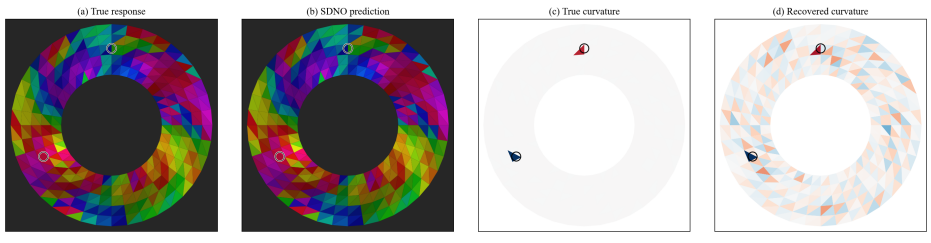


Figure 2: Magnetic Schrödinger response and curvature on an annulus from 24 observations. Ground truth (a,c) and SDNO recovery (b,d) show response phase and magnitude (a,b) and face-wise flux (c,d).

Table 3: Mean/median angular error after alignment on public Structure-from-Motion data, in degrees. V/E denotes the number of cameras/relative rotations. Lower is better.

Scene (V/E)	Spanning tree	Weiszfeld	Robust IRLS	Shonan	Shonan–Huber	NeuRoRA	SDNO-RA
Ellis (227/20k)	11.68/2.75	4.35/1.29	3.64/1.14	6.72/3.22	6.47/3.07	18.89/9.60	**3.38/1.09**
Madrid (341/24k)	26.56/12.23	17.16/8.77	**8.26/3.81**	12.31/8.08	11.94/7.70	32.28/21.42	8.47/4.03
Montreal (450/52k)	18.54/3.13	12.92/2.35	1.99/0.86	7.21/3.27	6.08/2.74	25.21/12.32	**1.60/0.74**
NYC (332/21k)	13.73/4.49	6.84/2.36	3.79/2.02	8.41/5.52	7.72/5.14	24.55/19.79	**3.75/1.79**
Tower (472/24k)	17.65/5.87	7.63/2.74	4.27/2.68	6.89/3.90	6.33/3.89	30.25/20.62	**4.16/2.67**
Alamo (577/97k)	9.35/2.25	6.36/1.97	4.84/1.97	8.18/4.03	7.51/3.72	21.16/15.36	**4.81/1.94**

6.5 Structure Constraints and Branch Ablations

Appendix Table 5 isolates the parameterization and prediction branches. With two observations, group-valued transport reduces error to 0.028, versus 0.557 for real and 0.457 for complex-linear matrices. Removing the defect or holonomy branch raises its targeted error to 0.025 or 0.089; further sweeps appear in Appendix Figure 7.

6.6 Structure-from-Motion Rotation Averaging

On public 1DSfM relative-rotation data (Wilson & Snavely, 2014), SDNO-RA combines spanning-tree initialization with loop-weighted Lie-algebra descent. Table 3 shows lower mean and median error than robust IRLS in five of six scenes. Baseline protocols and edge diagnostics appear in Appendix F.

7 Limitations and Conclusion

Recovery requires sufficient excitation of the local action, accurate surrogate Jacobians, and a spectral gap in symmetry measurement. The theory covers finite complexes with finite-dimensional unitary fibers; noncommutative curved-connection perturbations and scalable implicit or randomized projections remain open.

SDNO learns local symmetries and overlap transport from operator observations to constrain globally compatible layers. Curvature and holonomy inform defect prediction, while an equivariant corrector models the episode-conditioned residual. The experiments show gains when hidden connections, defects, and topology jointly affect responses. The controlled recovery, operator-transfer, and polycrystalline tasks provide fixed observation budgets and task-specific metrics for further evaluation.

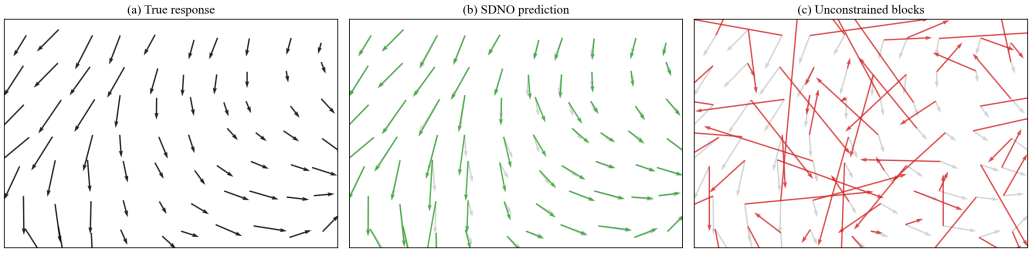


Figure 3: Polycrystalline displacement prediction from four observation pairs. (a) Ground-truth response, (b) SDNO prediction, and (c) unconstrained-local-block prediction; gray arrows show the ground-truth reference.

AI USE STATEMENT

Generative AI tools were used to aid and polish the writing, organize the manuscript for the conference format, and search for and identify relevant literature. The authors take responsibility for reviewing AI-assisted edits, checking the cited sources against the original publications, and verifying the final content of this paper.

REPRODUCIBILITY STATEMENT

Complete assumptions and proofs are in Appendix A. Appendix F documents data generation, splits, observation budgets, metrics, baselines, and training configurations. Appendix G describes the anonymous code artifact, evaluation scripts, fixed seeds, and instructions for obtaining the public datasets.

ETHICS STATEMENT

The controlled physical tasks are synthetic; public evaluations use 1DSfM pairwise rotations and WebKB graph connectivity, without processing scene images or webpage text. We did not recruit participants or make decisions about individuals. These benchmarks do not establish safety for deployment in consequential physical systems.

REFERENCES

Lennart Bastian, Samuel Leventhal, Mustafa Hajij, and Tolga Birdal. Topological neural operators. *arXiv preprint arXiv:2606.09806*, 2026. URL https://arxiv.org/abs/2606.09806.

Manu Bhat, Jonghyun Park, Jianke Yang, Nima Dehmamy, Robin Walters, and Rose Yu. AtlasD: Automatic local symmetry discovery. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 4153–4171. PMLR, 2025. URL https://proceedings.mlr.press/v267/bhat25a.html.

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*, volume 35, pp. 18527–18541, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/75c45fca2aa416ada062b26cc4fb7641-Abstract-Conference.html.

Avishek Chatterjee and Venu Madhav Govindu. Efficient and robust large-scale rotation averaging. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 521–528, 2013. doi: 10.1109/ICCV.2013.70. URL https://openaccess.thecvf.com/content_iccv_2013/html/Chatterjee_Efficient_and_Robust_2013_ICCV_paper.html.

Avishek Chatterjee and Venu Madhav Govindu. Robust relative rotation averaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):958–972, 2018. doi: 10.1109/TPAMI.2017.2693984. URL https://ee.iisc.ac.in/cvlab/papers/robustrelrotavg.pdf.

Chaoran Cheng and Jian Peng. Equivariant neural operator learning with graphon convolution. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/c362fbc0d182c6b4b8dadb90177239e4-Abstract-Conference.html.

Pengcheng Cheng. Gauge-equivariant intrinsic neural operators for geometry-consistent learning of elliptic PDE maps. *arXiv preprint arXiv:2603.14734*, 2026. URL https://arxiv.org/abs/2603.14734.

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

Taco S. Cohen, Maurice Weiler, Berkay Kicanaoglu, and Max Welling. Gauge equivariant convolutional networks and the icosahedral CNN. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 1321–1330. PMLR, 2019. URL https://proceedings.mlr.press/v97/cohen19d.html.

Frank Dellaert, David M. Rosen, Jing Wu, Robert Mahony, and Luca Carlone. Shonan rotation averaging: Global optimality by surfing $SO(p)^n$. In *Computer Vision – ECCV 2020*, volume 12351 of *Lecture Notes in Computer Science*, pp. 292–308. Springer, 2020. doi: 10.1007/978-3-030-58539-6_18. URL https://doi.org/10.1007/978-3-030-58539-6_18.

Mathieu Desbrun, Anil N. Hirani, Melvin Leok, and Jerrold E. Marsden. Discrete exterior calculus. *arXiv preprint arXiv:math/0508341*, 2005. URL https://arxiv.org/abs/math/0508341.

Marc Finzi, Max Welling, and Andrew Gordon Wilson. A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 3318–3328. PMLR, 2021. URL https://proceedings.mlr.press/v139/finzi21a.html.

Tingran Gao, Jacek Brodzki, and Sayan Mukherjee. The geometry of synchronization problems and learning group actions. *Discrete & Computational Geometry*, 65(1):150–211, 2021. doi: 10.1007/s00454-019-00100-2. URL https://doi.org/10.1007/s00454-019-00100-2.

Jakob Hansen and Thomas Gebhart. Sheaf neural networks. *arXiv preprint arXiv:2012.06333*, 2020. URL https://arxiv.org/abs/2012.06333.

Jakob Hansen and Robert Ghrist. Toward a spectral theory of cellular sheaves. *Journal of Applied and Computational Topology*, 3(4):315–358, 2019. doi: 10.1007/s41468-019-00038-7.

Zhongkai Hao, Zhengyi Wang, Hang Su, Chengyang Ying, Yinpeng Dong, Songming Liu, Ze Cheng, Jian Song, and Jun Zhu. GNOT: A general neural operator transformer for operator learning. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 12556–12569. PMLR, 2023. URL https://proceedings.mlr.press/v202/hao23c.html.

Richard Hartley, Khurrum Aftab, and Jochen Trumpf. L_1 rotation averaging using the weiszfeld algorithm. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3041–3048, 2011. doi: 10.1109/CVPR.2011.5995745.

Richard Hartley, Jochen Trumpf, Yuchao Dai, and Hongdong Li. Rotation averaging. *International Journal of Computer Vision*, 103(3):267–305, 2013. doi: 10.1007/s11263-012-0601-0.

Jacob Helwig, Xuan Zhang, Cong Fu, Jerry Kurtin, Stephan Wojtowytsch, and Shuiwang Ji. Group equivariant Fourier neural operators for partial differential equations. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 12907–12930, 2023. URL https://proceedings.mlr.press/v202/helwig23a.html.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2023. URL https://jmlr.org/papers/v24/21-1524.html.

Daiyuan Li, Shreya Arya, and Robert Ghrist. Learning from frustration: Torsor CNNs on graphs. In *Proceedings of the 4th (2025) and 3rd (2024) NeurIPS Workshops on Symmetry and Geometry in Neural Representations*, volume 282 of *Proceedings of Machine Learning Research*, pp. 298–315. PMLR, 2026. URL https://proceedings.mlr.press/v282/li26a.html.

Zijie Li, Kazem Meidani, and Amir Barati Farimani. Transformer for partial differential equations' operator learning. *Transactions on Machine Learning Research*, 2023. URL https://openreview.net/forum?id=EPPqt3uERT.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=c8P9NQVtmnO.

Lek-Heng Lim. Hodge laplacians on graphs. *SIAM Review*, 62(3):685–715, 2020. doi: 10.1137/18M1223101.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021. doi: 10.1038/s42256-021-00302-5.

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

Pulak Purkait, Tat-Jun Chin, and Ian Reid. NeuRoRA: Neural robust rotation averaging. In *Computer Vision – ECCV 2020*, volume 12369 of *Lecture Notes in Computer Science*, pp. 137–154. Springer, 2020. doi: 10.1007/978-3-030-58586-0_9. URL https://doi.org/10.1007/978-3-030-58586-0_9.

Lennon J. Shikhman and Shane Gilbertie. Cellular sheaf neural operators for structure-preserving surrogate modeling of constrained PDEs. *arXiv preprint arXiv:2606.00937*, 2026. URL https://arxiv.org/abs/2606.00937.

Amit Singer and Hsiao-Tung Wu. Vector diffusion maps and the connection laplacian. *Communications on Pure and Applied Mathematics*, 65(8):1067–1144, 2012. doi: 10.1002/cpa.21395.

Kartik Tandon, Julian Gould, Tanishq Bhatia, Francesca Dominici, Alejandro Ribeiro, and Claudio Battiloro. Consistent geometric deep learning via hilbert bundles and cellular sheaves. *arXiv preprint arXiv:2605.06395*, 2026. URL https://arxiv.org/abs/2605.06395.

Kyle Wilson and Noah Snavely. Robust global translations with 1DSfM. In *Computer Vision – ECCV 2014*, volume 8691 of *Lecture Notes in Computer Science*, pp. 61–75. Springer, 2014. doi: 10.1007/978-3-319-10578-9_5. URL https://doi.org/10.1007/978-3-319-10578-9_5.

Minkai Xu, Jiaqi Han, Aaron Lou, Jean Kossaifi, Arvind Ramanathan, Kamyar Azizzadenesheli, Jure Leskovec, Stefano Ermon, and Anima Anandkumar. Equivariant graph neural operator for modeling 3D dynamics. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 55015–55032, 2024. URL https://proceedings.mlr.press/v235/xu24j.html.

Jianke Yang, Robin Walters, Nima Dehmamy, and Rose Yu. Generative adversarial symmetry discovery. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 39488–39508, 2023a. URL https://proceedings.mlr.press/v202/yang23n.html.

Jianke Yang, Nima Dehmamy, Robin Walters, and Rose Yu. Latent space symmetry discovery. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 56047–56070, 2024. URL https://proceedings.mlr.press/v235/yang24g.html.

Liu Yang, Siting Liu, Tingwei Meng, and Stanley J. Osher. In-context operator learning with data prompts for differential equation problems. *Proceedings of the National Academy of Sciences*, 120 (39):e2310142120, 2023b. doi: 10.1073/pnas.2310142120.

Dongzhe Zheng, Tao Zhong, and Christine Allen-Blanchette. Topology-preserving neural operator learning via hodge decomposition. *arXiv preprint arXiv:2605.13834*, 2026. URL https://arxiv.org/abs/2605.13834.

A FINITE-DIMENSIONAL THEORETICAL SETTING

This appendix gives the assumptions and proofs for the results in Section 5. Let N be a finite, connected, oriented two-dimensional cellular complex. In the language of covers, N can be taken as the Čech nerve of a finite cover $\{U_i\}_{i\in\mathcal{I}}$, where $\mathcal{I}$ is a finite patch index set. If N is required to represent the homotopy type of the original space, the cover can additionally be assumed to be a good cover; the algebraic descent results below use only the cellular structure of the nerve. Background on cellular sheaves, connection synchronization, and discrete Hodge theory can be found in Hansen & Ghrist (2019); Gao et al. (2021); Lim (2020).

The theory uses two related types of objects. Gluing data consist of invertible transition maps satisfying a cocycle and define a flat local system. Connection data consist of parallel transports on oriented edges and may carry curvature. The descent branch of SDNO uses the flat projection of the connection, while the curvature residual enters the defect branch. This separation treats the cochain-complex identity of the twisted differential and the curvature signal separately.

Unless stated otherwise, $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{Z}$ denote the real numbers, complex numbers, and integers, respectively; I denotes the identity map; A^* denotes the Hermitian adjoint of A; $\ker A$ and $\operatorname{im} A$ denote the null space and image; P_S denotes the orthogonal projection onto a subspace S; $\|A\|_{\mathrm{op}}$ denotes the operator norm; $\operatorname{dist}(x,S) := \|x - P_S x\|$ denotes the distance from a point to a subspace; and $\sigma_+(A)$ denotes the smallest nonzero singular value of A. All direct-sum spaces use the Hilbert norm induced by the fiber inner products.

For $s \in \{\mathrm{in}, \mathrm{out}\}$, each vertex or patch i is equipped with a finite-dimensional complex Hilbert space V_i^s. In a PDE discretization, V_i^s represents a patch-local function or feature space. Each oriented edge $e : i \to j$ is equipped with unitary transport

$$T_e^s : V_i^s \longrightarrow V_j^s, \qquad T_{\bar{e}}^s = (T_e^s)^{-1}. \tag{13}$$

where $\bar{e} : j \to i$ is the reverse of e. For each oriented 2-cell f, let ∂f be the sequence of oriented edges ordered along its boundary, and define

$$H_f^s := \prod_{e\in\partial f} T_e^s.$$

Here H_f^s is the holonomy of face f, and I is the identity map on the corresponding fiber. If $H_f^s = I$ for every f, the collection of transports $T^s := \{T_e^s\}_e$ is called flat. Face flatness on a triangular nerve corresponds to the triple-overlap cocycle.

Each i is also equipped with a Lie algebra $\mathfrak{g}_i$. Each oriented edge is equipped with a Lie-algebra isomorphism

$$\phi_e : \mathfrak{g}_i \longrightarrow \mathfrak{g}_j, \qquad \phi_{\bar{e}} = \phi_e^{-1},$$

and these isomorphisms satisfy path composition along incident edges. The input and output representations are

$$\rho_i^s : \mathfrak{g}_i \longrightarrow \mathrm{End}(V_i^s)$$

where $\mathrm{End}(V_i^s)$ denotes the space of linear endomorphisms of V_i^s. The representations satisfy transport compatibility

$$T_e^s \rho_i^s(X) = \rho_j^s(\phi_e X) T_e^s, \qquad X \in \mathfrak{g}_i. \tag{14}$$

Definition A.1 (Local admissible-operator stalk)**.** Define

$$\mathcal{A}_i := \mathrm{Hom}_{\mathfrak{g}_i}\left(V_i^{\mathrm{in}}, V_i^{\mathrm{out}}\right) = \left\{W :\ W\rho_i^{\mathrm{in}}(X) = \rho_i^{\mathrm{out}}(X)W,\ \forall X \in \mathfrak{g}_i\right\}. \tag{15}$$

For an edge $e : i \to j$, define the linear map $R_e : \mathcal{A}_i \to \mathcal{A}_j$ by

$$R_e(W) := T_e^{\mathrm{out}} W (T_e^{\mathrm{in}})^{-1}. \tag{16}$$

Proposition A.2 (Flat transitions and locally constant cellular sheaves)**.** *Let $\Pi_1(N)$ be the path groupoid of N. A collection of invertible edge transitions satisfies the reverse-edge relation and every 2-cell cocycle if and only if it defines a functor from $\Pi_1(N)$ to the category of finite-dimensional vector spaces. This functor defines a flat local system. When N is viewed as a cellular complex, it is equivalent to a locally constant cellular sheaf whose restriction maps are invertible.*

Proof. Compose the transitions successively along an edge path. The reverse-edge relation cancels adjacent $e\bar{e}$ or $\bar{e}e$ pairs, and the face cocycle keeps the composition unchanged when a 2-cell boundary is inserted or removed. The composition therefore depends only on the morphism class in the path groupoid and defines a functor. Conversely, preservation of identities and composition by the functor gives the reverse-edge relation and the face cocycle.

For a cellular sheaf $\mathcal{F}$, let the endpoints of e be i, j, and let $\rho_{i\trianglelefteq e} : \mathcal{F}(i) \to \mathcal{F}(e)$ denote the restriction map from the vertex stalk to the edge stalk. If $\rho_{i\trianglelefteq e}$ and $\rho_{j\trianglelefteq e}$ are invertible, define

$$T_{ji} = \rho_{j\trianglelefteq e}^{-1}\rho_{i\trianglelefteq e}.$$

Restriction functoriality gives the face cocycle. Conversely, choose a local trivialization in each cell and define restrictions using path transport; face flatness ensures that different paths produce the same restriction. The two constructions are naturally isomorphic. □

B Exact Descent and Holonomy Obstruction

Theorem B.1 (Operator-constraint local system and exact descent)**.** *Assume that Equations* (13)–(14) *hold and that the input and output transports are flat. Then:*

(a) $R_e(\mathcal{A}_i) = \mathcal{A}_j$*; for any composable oriented edges* e_1, e_2*,* $R_{e_2}R_{e_1} = R_{e_2e_1}$*, where* $R_{e_2e_1}$ *denotes the operator transport induced along the composite path* e_2e_1*. Hence,* $\{\mathcal{A}_i, R_e\}$ *forms an operator-constraint local system;*

(b) a family of local operators $W = (W_i)$ *is a global section of this local system if and only if every edge satisfies*

$$T_e^{\mathrm{out}}W_i = W_jT_e^{\mathrm{in}}; \tag{17}$$

(c) let

$$\Gamma_T(V^s) := \{x = (x_i) :\ x_j = T_e^s x_i,\ \forall e : i \to j\}.$$

any global operator section induces a unique linear map

$$\mathcal{W} : \Gamma_T(V^{\mathrm{in}}) \longrightarrow \Gamma_T(V^{\mathrm{out}}), \qquad (\mathcal{W}x)_i = W_ix_i;$$

(d) if $X = (X_i)$ *is a compatible Lie-algebra section, i.e.,* $X_j = \phi_eX_i$*, and* $\rho^s(X) := \mathrm{diag}(\rho_i^s(X_i))$*, then*

$$\mathcal{W}\rho^{\mathrm{in}}(X) = \rho^{\mathrm{out}}(X)\mathcal{W}.$$

Proof. Take $W \in \mathcal{A}_i$. For any $Y \in \mathfrak{g}_j$, invertibility of ϕ_e gives an $X \in \mathfrak{g}_i$ such that $Y = \phi_eX$. Equation (14) gives

$$\begin{aligned} R_e(W)\rho_j^{\mathrm{in}}(Y) &= T_e^{\mathrm{out}}W(T_e^{\mathrm{in}})^{-1}T_e^{\mathrm{in}}\rho_i^{\mathrm{in}}(X)(T_e^{\mathrm{in}})^{-1} \\ &= T_e^{\mathrm{out}}W\rho_i^{\mathrm{in}}(X)(T_e^{\mathrm{in}})^{-1} \\ &= T_e^{\mathrm{out}}\rho_i^{\mathrm{out}}(X)W(T_e^{\mathrm{in}})^{-1} \\ &= \rho_j^{\mathrm{out}}(Y)R_e(W). \end{aligned}$$

Thus, $R_e(W) \in \mathcal{A}_j$. Applying the same result to the reverse edge gives $R_e(\mathcal{A}_i) = \mathcal{A}_j$. Path composition of the transports directly gives $R_{e_2}R_{e_1} = R_{e_2e_1}$, proving (a).

By Equation (16), the condition $W_j = R_e(W_i)$ is equivalent to Equation (17), proving (b). If $x \in \Gamma_T(V^{\mathrm{in}})$, then

$$(\mathcal{W}x)_j = W_jx_j = W_jT_e^{\mathrm{in}}x_i = T_e^{\mathrm{out}}W_ix_i.$$

Therefore, $\mathcal{W}x \in \Gamma_T(V^{\mathrm{out}})$. All local components are specified by W_i, so the induced map is unique. Finally, Equation (15) holds for every i; applying it component-wise to a compatible X gives (d). □

Remark B.2 (Natural transformations and network layers). Flat transports define functors $\mathcal{V}^{\mathrm{in}}, \mathcal{V}^{\mathrm{out}} : \Pi_1(N) \to \mathbf{FdVect}$, where $\Pi_1(N)$ is the path groupoid of N and $\mathbf{FdVect}$ is the category of finite-dimensional complex vector spaces and linear maps. Equation (17) shows that a family of local operators $W = (W_i)$ is a natural transformation from $\mathcal{V}^{\mathrm{in}}$ to $\mathcal{V}^{\mathrm{out}}$ if and only if it is a global section of the operator-constraint local system. This equivalence converts categorical compatibility into linear parameter constraints that can be computed directly on each edge.

Remark B.3 (Descent error and equivariance error). Exact descent uses the cocycle and naturality conditions. Local equivariance additionally gives the descended operator global equivariance. The theoretical estimates and experimental metrics therefore record the globalization defect and equivariance defect separately.

Theorem B.4 (Holonomy obstruction)**.** *Fix a base point i_0, let $\pi_1(N, i_0)$ be the fundamental group based at i_0, let $\Gamma(\mathcal{A})$ denote the global-section space of the operator-constraint local system, and let*

$$\mathrm{Hol}^s : \pi_1(N, i_0) \longrightarrow U(V^s_{i_0})$$

be the holonomy representations of the flat input and output local systems, where $U(V)$ denotes the unitary group of a Hilbert space V. Taking the base-point component $W_0 := W_{i_0}$ of a global operator section gives the natural isomorphism

$$\Gamma(\mathcal{A}) \cong \left\{ W_0 \in \mathcal{A}_{i_0} : \ \mathrm{Hol}^{\mathrm{out}}(\gamma) W_0 = W_0 \, \mathrm{Hol}^{\mathrm{in}}(\gamma), \ \forall \gamma \in \pi_1(N, i_0) \right\}. \tag{18}$$

Proof. Let $W = (W_i)$ be a global section. Repeatedly applying Equation (17) along a closed path $\gamma = e_m \cdots e_1$ that begins and ends at i_0 gives

$$T^{\mathrm{out}}_{e_m} \cdots T^{\mathrm{out}}_{e_1} W_0 = W_0 T^{\mathrm{in}}_{e_m} \cdots T^{\mathrm{in}}_{e_1}.$$

This is precisely the intertwining relation in Equation (18).

Conversely, take a W_0 satisfying Equation (18). For each vertex i, choose a path $p_i : i_0 \to i$, let $T^s_{p_i}$ denote the ordered transport product along p_i, and define

$$W_i := T^{\mathrm{out}}_{p_i} W_0 (T^{\mathrm{in}}_{p_i})^{-1}. \tag{19}$$

If another path p_i' is used, the holonomy intertwining relation for the loop $p_i'^{-1} p_i$ makes the two paths give the same W_i in Equation (19). Thus, W_i is independent of the path choice. For any edge $e : i \to j$, choosing $p_j = e p_i$ verifies Equation (17). Theorem B.1(a) also ensures that $W_i \in \mathcal{A}_i$. The two constructions are inverse to one another and give the stated natural isomorphism. □

Corollary B.5 (Dimension of the global architecture)**.** *Temporarily omit the additional $\mathfrak{g}$ constraints. Let α range over pairwise inequivalent irreducible unitary representations, let U_α denote the corresponding representation space, and let m^s_α denote its multiplicity in $V^s_{i_0}$. Suppose that the complex unitary holonomy representations decompose as*

$$V^s_{i_0} \cong \bigoplus_\alpha \mathbb{C}^{m^s_\alpha} \otimes U_\alpha,$$

Then

$$\dim \Gamma\big(\mathrm{Hom}(V^{\mathrm{in}}, V^{\mathrm{out}})\big) = \sum_\alpha m^{\mathrm{in}}_\alpha m^{\mathrm{out}}_\alpha. \tag{20}$$

Proof. Theorem B.4 identifies the global architecture space with the intertwiner space of the holonomy representations. Apply Schur's lemma block-wise to the irreducible decompositions. Only isomorphic irreducible blocks have nonzero intertwiners; the dimension of the linear maps between the corresponding multiplicity spaces is $m^{\mathrm{in}}_\alpha m^{\mathrm{out}}_\alpha$. Summing over α gives Equation (20). □

Corollary B.6 (Differentiable holonomy-obstruction score)**.** *Take b loops $\gamma_1, \ldots, \gamma_b$ that generate $\pi_1(N, i_0)$. For a base-point linear operator $W : V^{\mathrm{in}}_{i_0} \to V^{\mathrm{out}}_{i_0}$, define the holonomy-compatibility operator*

$$\mathcal{C}_{\mathrm{hol}}(W) := \Big(\mathrm{Hol}^{\mathrm{out}}(\gamma_\ell) W - W \, \mathrm{Hol}^{\mathrm{in}}(\gamma_\ell) \Big)_{\ell=1}^{b}.$$

The globalizable base-point operators form $\ker \mathcal{C}_{\mathrm{hol}}$. If $\sigma_+(\mathcal{C}_{\mathrm{hol}}) > 0$ is the smallest nonzero singular value, then

$$\mathrm{dist}(W, \ker \mathcal{C}_{\mathrm{hol}}) \le \frac{\|\mathcal{C}_{\mathrm{hol}}(W)\|}{\sigma_+(\mathcal{C}_{\mathrm{hol}})}. \tag{21}$$

Proof. Orthogonally decompose W as $W_\parallel + W_\perp$, where $W_\parallel \in \ker \mathcal{C}_{\mathrm{hol}}$ and $W_\perp \perp \ker \mathcal{C}_{\mathrm{hol}}$. The definition of the smallest nonzero singular value gives

$$\|\mathcal{C}_{\mathrm{hol}}(W)\| = \|\mathcal{C}_{\mathrm{hol}}(W_\perp)\| \ge \sigma_+(\mathcal{C}_{\mathrm{hol}}) \, \|W_\perp\|.$$

Since $\|W_\perp\| = \mathrm{dist}(W, \ker \mathcal{C}_{\mathrm{hol}})$, Equation (21) follows. □

C Gauge Covariance and Differentiable Null-Space Projection

Theorem C.1 (Gauge covariance). *For each patch, take unitary frame changes $Q_i^s : V_i^s \to V_i^s$, let $x = (x_i)$ be an input assignment, and define*

$$\begin{aligned}\widetilde{T}_e^s &= Q_j^s T_e^s (Q_i^s)^{-1},\\ \widetilde{\rho}_i^s(X) &= Q_i^s \rho_i^s(X)(Q_i^s)^{-1},\\ \widetilde{W}_i &= Q_i^{\mathrm{out}} W_i (Q_i^{\mathrm{in}})^{-1},\\ \widetilde{x}_i &= Q_i^{\mathrm{in}} x_i.\end{aligned}$$

Let $\widetilde{x} := (\widetilde{x}_i)$, let $Q^s := \mathrm{diag}(Q_i^s)$ be the block-diagonal action on the corresponding direct-sum space, and let $\widetilde{\mathcal{W}}$ be the global map induced by the local operator family $\{\widetilde{W}_i\}_i$. Then:

(a) flatness, representation compatibility, local equivariance, and naturality are preserved;

(b) the descended prediction satisfies $\widetilde{\mathcal{W}}\widetilde{x} = Q^{\mathrm{out}}\mathcal{W}x$;

(c) the holonomy of a loop based at the base point satisfies

$$\widetilde{\mathrm{Hol}}^s(\gamma) = Q_{i_0}^s \,\mathrm{Hol}^s(\gamma)(Q_{i_0}^s)^{-1};$$

(d) if δ_T^k is the kth coboundary of the flat local system, Q^k denotes the block-diagonal gauge action on the k-cochain space, and the twisted Laplacian is defined by $\Delta_T^k := \delta_T^{k-1}(\delta_T^{k-1})^* + (\delta_T^k)^*\delta_T^k$, then

$$\delta_{\widetilde{T}}^k = Q^{k+1}\delta_T^k (Q^k)^{-1}, \qquad \Delta_{\widetilde{T}}^k = Q^k \Delta_T^k (Q^k)^{-1}. \tag{22}$$

Proof. Multiply the transformed transports $\widetilde{T}_e^s$ along a face boundary, and let $\widetilde{H}_f^s$ denote the face holonomy of the transformed transport. The factors $(Q_i^s)^{-1}Q_i^s$ between adjacent edges cancel term by term, leaving only conjugation at the base point. Thus, $H_f^s = I$ implies $\widetilde{H}_f^s = I$. Substituting $\widetilde{T}_e^s$ and $\widetilde{\rho}_i^s$ into Equation (14), the intermediate gauge factors again cancel and give representation compatibility. Local equivariance and naturality follow by the same calculation.

For every component,

$$\widetilde{W}_i\widetilde{x}_i = Q_i^{\mathrm{out}} W_i (Q_i^{\mathrm{in}})^{-1} Q_i^{\mathrm{in}} x_i = Q_i^{\mathrm{out}} W_i x_i,$$

Hence, the descended prediction is covariant under the output gauge. Multiplying the transports along a loop cancels all gauge factors at intermediate vertices and leaves conjugation at the base point, proving (c).

Each incidence block of the coboundary transforms under the gauge actions on the source and target fibers, giving the first identity in Equation (22). Unitarity gives

$$(\delta_{\widetilde{T}}^k)^* = Q^k (\delta_T^k)^* (Q^{k+1})^{-1}.$$

Substituting into

$$\Delta_T^k = \delta_T^{k-1}(\delta_T^{k-1})^* + (\delta_T^k)^*\delta_T^k$$

which yields the conjugation relation for the Laplacian. □

Remark C.2 (General invertible frame changes). When Q_i^s is a general invertible map, the fiber metrics must be pushed forward simultaneously. Defining adjoints and orthogonal projections with the transformed metrics recovers the corresponding covariance relations. We use the unitary or orthogonal setting so that the adjoint, spectrum, and projector have the expressions above under fixed inner products.

Lemma C.3 (Gauge covariance of hard and soft projectors). *Let $\mathcal{C}_i$ be a linear constraint operator acting on $\mathrm{Hom}(V_i^{\mathrm{in}}, V_i^{\mathrm{out}})$, and let $\widetilde{\mathcal{C}}_i$ be the constraint operator after a frame change. Assume that $\mathcal{C}_i$ is tensorial: under a gauge change, there is a residual-space unitary Q_i^{res} such that*

$$\widetilde{\mathcal{C}}_i \circ S_i = Q_i^{\mathrm{res}} \circ \mathcal{C}_i, \qquad S_i(W) = Q_i^{\mathrm{out}} W (Q_i^{\mathrm{in}})^{-1}. \tag{23}$$

Then

$$P_{\ker \widetilde{\mathcal{C}}_i} = S_i P_{\ker \mathcal{C}_i} S_i^{-1}. \tag{24}$$

For any constraint operator $\mathcal{C}$ and $\tau > 0$, define

$$P_\tau(\mathcal{C}) := I - \mathcal{C}^*(\mathcal{C}\mathcal{C}^* + \tau I)^{-1}\mathcal{C}. \tag{25}$$

Then

$$P_\tau(\widetilde{\mathcal{C}}_i) = S_i P_\tau(\mathcal{C}_i) S_i^{-1}. \tag{26}$$

If γ is the smallest nonzero singular value of $\mathcal{C}$, then

$$\|P_\tau(\mathcal{C}) - P_{\ker \mathcal{C}}\|_{\mathrm{op}} \le \frac{\tau}{\gamma^2 + \tau}, \qquad \|\mathcal{C}P_\tau(\mathcal{C})\|_{\mathrm{op}} \le \frac{\sqrt{\tau}}{2}. \tag{27}$$

Proof. Under the Frobenius inner product, S_i is unitary. Equation (23) gives

$$\ker \widetilde{\mathcal{C}}_i = S_i(\ker \mathcal{C}_i).$$

Unitary maps also map orthogonal complements to orthogonal complements, so the orthogonal projector satisfies Equation (24). From

$$\widetilde{\mathcal{C}}_i = Q_i^{\mathrm{res}} \mathcal{C}_i S_i^{-1}$$

and the unitarity of S_i, Q_i^{res}, direct substitution into Equation (25) gives Equation (26).

Take a singular value decomposition $\mathcal{C} = U\Sigma V^*$, where U, V are unitary factors and Σ is a nonnegative diagonal matrix of singular values. On $\ker \mathcal{C}$, the eigenvalue of $P_\tau(\mathcal{C})$ is 1. On a right singular direction corresponding to a singular value $\sigma > 0$, this eigenvalue is

$$1 - \frac{\sigma^2}{\sigma^2 + \tau} = \frac{\tau}{\sigma^2 + \tau}.$$

Thus, the operator norm of $P_\tau - P_{\ker \mathcal{C}}$ is at most $\tau/(\gamma^2 + \tau)$. On the same singular direction, the singular value of $\mathcal{C}P_\tau$ is

$$\frac{\sigma\tau}{\sigma^2 + \tau}.$$

This function attains its maximum $\sqrt{\tau}/2$ at $\sigma = \sqrt{\tau}$, giving the second bound. □

D Approximate Descent and Curvature Correction

Fix one orientation for each undirected edge. Let $\bigoplus_i V_i^s$ be the direct sum of the vertex fibers, and let $\bigoplus_e V_{\mathrm{head}(e)}^s$ be the edge space corresponding to the selected oriented edges, where $\mathrm{head}(e) := j$ for $e : i \to j$. For a vertex assignment $y = (y_i) \in \bigoplus_i V_i^s$, define the twisted incidence operator $D_T^s : \bigoplus_i V_i^s \to \bigoplus_e V_{\mathrm{head}(e)}^s$ by

$$(D_T^s y)_e := y_j - T_e^s y_i, \qquad e : i \to j. \tag{28}$$

Let L_T^s be the corresponding twisted graph Laplacian, let $\Gamma_T(V^s)$ be the compatible section space, and let P_T^s be the orthogonal projection onto this space:

$$L_T^s = (D_T^s)^* D_T^s, \qquad \Gamma_T(V^s) = \ker D_T^s, \qquad P_T^s = P_{\ker D_T^s},$$

Denote its smallest positive eigenvalue by

$$\lambda_+^s := \lambda_{\min}^+(L_T^s) = \sigma_+^2(D_T^s).$$

Here $\lambda_{\min}^+$ denotes the smallest positive eigenvalue.

Theorem D.1 (Quantitative descent under fixed flat transports)**.** *Suppose that the input and output transports are unitary. Let $W = \mathrm{diag}(W_i)$, and define the edge naturality residual*

$$\mathcal{R}_e := W_j T_e^{\mathrm{in}} - T_e^{\mathrm{out}} W_i, \qquad \nu := \max_e \|\mathcal{R}_e\|_{\mathrm{op}}. \tag{29}$$

Let $d_{\max}$ be the maximum out-degree under the selected edge orientation, and assume that $\lambda_+^{\mathrm{out}} > 0$. Then, for any $x \in \Gamma_{T^{\mathrm{in}}}(V^{\mathrm{in}})$,

$$\operatorname{dist}\big(Wx, \Gamma_{T^{\mathrm{out}}}(V^{\mathrm{out}})\big) \le \sqrt{\frac{d_{\max}}{\lambda_+^{\mathrm{out}}}}\, \nu\, \|x\|. \tag{30}$$

Moreover,

$$\widehat{\mathcal{W}} := P_T^{\mathrm{out}} W\big|_{\Gamma_{T^{\mathrm{in}}}(V^{\mathrm{in}})}$$

gives the compatible output closest to the raw local prediction.

Proof. For any output assignment y, write

$$z = (I - P_T^{\mathrm{out}})y \in (\ker D_T^{\mathrm{out}})^{\perp}.$$

Because $D_T^{\mathrm{out}} P_T^{\mathrm{out}} = 0$, the definition of the spectral gap gives

$$\begin{aligned} \big\|D_T^{\mathrm{out}} y\big\|^2 &= \big\|D_T^{\mathrm{out}} z\big\|^2 \\ &\ge \lambda_+^{\mathrm{out}} \|z\|^2. \end{aligned}$$

Therefore,

$$\operatorname{dist}\big(y, \Gamma_{T^{\mathrm{out}}}(V^{\mathrm{out}})\big) \le (\lambda_+^{\mathrm{out}})^{-1/2} \big\|D_T^{\mathrm{out}} y\big\|. \tag{31}$$

Take $y = Wx$. A compatible input satisfies $x_j = T_e^{\mathrm{in}} x_i$, so

$$(D_T^{\mathrm{out}} W x)_e = W_j x_j - T_e^{\mathrm{out}} W_i x_i = \mathcal{R}_e x_i.$$

If d_i^+ is the number of edges with tail i, then

$$\begin{aligned} \big\|D_T^{\mathrm{out}} W x\big\|^2 &= \sum_{e: i \to j} \|\mathcal{R}_e x_i\|^2 \\ &\le \nu^2 \sum_i d_i^+ \|x_i\|^2 \\ &\le d_{\max} \nu^2 \|x\|^2. \end{aligned}$$

Substitution into Equation (31) gives Equation (30). The closest-point property follows from the definition of P_T^{out} as an orthogonal projection. □

Remark D.2 (Cover constant). In Equation (30),

$$C_{N,T} = \sqrt{d_{\max}/\lambda_+^{\mathrm{out}}}$$

records both the overlap multiplicity and the twisted-nerve spectral gap. High overlap multiplicity increases $d_{\max}$; a connection close to gaining additional compatible modes makes λ_+^{out} small. General invertible transports require a generalized spectral gap adapted to the fiber metrics, and the corresponding metric condition numbers enter the constant.

Lemma D.3 (Flat projection of a two-dimensional lifted U(1) connection)**.** *Let*

$$C^0(N;\mathbb{R}) \xrightarrow{d_0} C^1(N;\mathbb{R}) \xrightarrow{d_1} C^2(N;\mathbb{R})$$

be a finite cochain complex with positive-definite cochain inner products, and let d_0^* and d_1^* denote the corresponding adjoints. For any lifted edge-angle cochain $a \in C^1(N;\mathbb{R})$, let

$$\bar{a} = P_{\ker d_1} a.$$

Then

$$\|a - \bar{a}\| \le \frac{\|d_1 a\|}{\sigma_+(d_1)}. \tag{32}$$

and a has a unique orthogonal component decomposition

$$a = a_{\mathrm{ex}} + a_{\mathrm{coex}} + h, \quad a_{\mathrm{ex}} \in \operatorname{im} d_0, \quad a_{\mathrm{coex}} \in \operatorname{im} d_1^*, \quad h \in \mathcal{H}^1 := \ker d_1 \cap \ker d_0^*. \tag{33}$$

where

$$\bar{a} = a_{\mathrm{ex}} + h, \qquad a - \bar{a} = a_{\mathrm{coex}}.$$

We can write $a_{\mathrm{ex}} = d_0\phi$ and $a_{\mathrm{coex}} = d_1^*\psi$, where $\phi \in C^0(N;\mathbb{R})$ and $\psi \in C^2(N;\mathbb{R})$ may contain the corresponding kernel freedom. For any vertex phase $\lambda \in C^0(N;\mathbb{R})$, the gauge transformation $a \mapsto a + d_0\lambda$ changes only the exact component.

Proof. The finite-dimensional Hodge decomposition gives the orthogonal direct sum

$$C^1(N;\mathbb{R}) = \operatorname{im} d_0 \oplus \mathcal{H}^1 \oplus \operatorname{im} d_1^*.$$

Therefore, the three components in Equation (33) are unique. From

$$\ker d_1 = \operatorname{im} d_0 \oplus \mathcal{H}^1, \qquad (\ker d_1)^\perp = \operatorname{im} d_1^*,$$

we have $\bar{a} = a_{\mathrm{ex}} + h$ and $b := a - \bar{a} = a_{\mathrm{coex}} \in \operatorname{im} d_1^*$. The smallest singular value of d_1 on $(\ker d_1)^\perp$ is $\sigma_+(d_1)$, so

$$\|d_1 a\| = \|d_1 b\| \geq \sigma_+(d_1)\, \|b\| .$$

Rearranging gives Equation (32). The gauge statement follows from $d_1 d_0 = 0$ and the orthogonal direct sum. □

Remark D.4 (Phase lift and integer flux sectors). Write U(1) edge transport as $U_e = e^{ia_e}$, where a_e is the phase of edge e and is defined modulo $2\pi\mathbb{Z}$. Lemma D.3 is used within a fixed lift or topological sector: after fixing an integer 2-cochain $n \in C^2(N;\mathbb{Z})$, the lifted curvature is $d_1 a - 2\pi n$. An experiment can restrict the face flux to an interval away from the branch cut or explicitly record finitely many integer sectors. The experiments in Section 6 use enforced sector selection. Estimated edge phases naturally lie in $(-\pi, \pi]$; when the hidden vertex gauge has large magnitude, the lift of the true connection crosses the branch cut. In this case, directly applying the orthogonal decomposition in Equation (33) to wrapped phases spreads the energy of the 2π jumps across all edges and contaminates the three components. A valid sector is obtained by spanning-tree gauge fixing. After the phases on tree edges are gauged to zero along the spanning tree, the remaining edge phases carry only holonomy and curvature contributions. When $|\text{holonomy}| \leq \pi$ and the curvature magnitude is much smaller than π, these phases remain away from the branch cut and the decomposition is well defined in that sector; predictions depend only on the class of a modulo 2π, and the gauge phase can be added back directly.

Corollary D.5 (Approximate descent with a curvature residual)**.** *Let $a^{\mathrm{in}}, a^{\mathrm{out}}$ be U(1) edge angles in a fixed lift sector. For $s \in \{\mathrm{in}, \mathrm{out}\}$, let $\bar{a}^s = P_{\ker d_1} a^s$, and let T^s and $\bar{T}^s$ denote the raw and flat transports generated by a^s and $\bar{a}^s$, respectively. For any edge angles a, b in this sector, assume that*

$$\|W_i\|_{\mathrm{op}} \leq M, \qquad \|T(a) - T(b)\|_{\mathrm{op}} \leq L_T |a - b|$$

where M is an upper bound on the local operator norm, $T(a)$ denotes the transport generated by the edge angle a, and L_T is its Lipschitz constant. Take an edge-evaluation constant c_E such that

$$\max_e |b_e| \leq c_E\, \|b\|_{C^1}$$

for all edge cochains $b \in C^1(N;\mathbb{R})$, where $\|\cdot\|_{C^1}$ is the selected cochain norm; under the standard unweighted ℓ^2 norm, one can take $c_E = 1$. Let the raw naturality residual be $\nu := \max_{e:i\to j} \left\| W_j T_e^{\mathrm{in}} - T_e^{\mathrm{out}} W_i \right\|_{\mathrm{op}}$, and define the flat-corrected residual

$$\bar{\nu} := \max_{e:i\to j} \left\| W_j \bar{T}_e^{\mathrm{in}} - \bar{T}_e^{\mathrm{out}} W_i \right\|_{\mathrm{op}},$$

Then

$$\bar{\nu} \leq \nu + \frac{M c_E L_T}{\sigma_+(d_1)} \left(\left\| d_1 a^{\mathrm{in}} \right\| + \left\| d_1 a^{\mathrm{out}} \right\| \right). \tag{34}$$

For $x \in \Gamma_{\bar{T}^{\mathrm{in}}}(V^{\mathrm{in}})$,

$$\begin{aligned} &\operatorname{dist}\big(Wx, \Gamma_{\bar{T}^{\mathrm{out}}}(V^{\mathrm{out}})\big) \\ &\quad \leq \sqrt{\frac{d_{\max}}{\lambda_{\min}^+(L_{\bar{T}}^{\mathrm{out}})}} \left[\nu + \frac{M c_E L_T}{\sigma_+(d_1)} \left(\left\| d_1 a^{\mathrm{in}} \right\| + \left\| d_1 a^{\mathrm{out}} \right\| \right) \right] \|x\|. \end{aligned} \tag{35}$$

Proof. For each edge,

$$\begin{aligned} W_j \bar{T}_e^{\mathrm{in}} - \bar{T}_e^{\mathrm{out}} W_i = {} & W_j T_e^{\mathrm{in}} - T_e^{\mathrm{out}} W_i \\ & + W_j (\bar{T}_e^{\mathrm{in}} - T_e^{\mathrm{in}}) - (\bar{T}_e^{\mathrm{out}} - T_e^{\mathrm{out}}) W_i. \end{aligned}$$

Taking operator norms and using $\|W_i\|_{\mathrm{op}} \le M$ gives

$$\bar{\nu} \le \nu + M(\eta_{\mathrm{in}} + \eta_{\mathrm{out}}), \qquad \eta_s := \max_e \left\|\bar{T}_e^s - T_e^s\right\|_{\mathrm{op}}.$$

The Lipschitz condition on transport, the edge-evaluation bound, and Lemma D.3 give

$$\eta_s \le L_T \max_e |\bar{a}_e^s - a_e^s| \le \frac{c_E L_T}{\sigma_+(d_1)} \|d_1 a^s\|.$$

This proves Equation (34). Applying Theorem D.1 to the flat transports $\bar{T}^s$ then gives Equation (35). □

Corollary D.6 (Globalization and global equivariance)**.** *Let $X = (X_i)$ be a compatible symmetry section, i.e., $X_j = \phi_e X_i$. Let $\rho^s(X) := \mathrm{diag}(\rho_i^s(X_i))$ be the direct-sum representation, and let $\varepsilon_{\mathrm{eq}} \ge 0$ be the local equivariance-residual constant in*

$$E_i(X_i) := W_i \rho_i^{\mathrm{in}}(X_i) - \rho_i^{\mathrm{out}}(X_i) W_i, \qquad \max_i \|E_i(X_i)\|_{\mathrm{op}} \le \varepsilon_{\mathrm{eq}} \|X\|.$$

Here $E_i(X_i)$ is the local equivariance residual on patch i. If the symmetry-group action is unitary and preserves the compatible subspaces, then the descended map

$$\widehat{\mathcal{W}} = P_{\bar{T}}^{\mathrm{out}} W\big|_{\Gamma_{\bar{T}^{\mathrm{in}}}}$$

satisfies

$$\left\|\widehat{\mathcal{W}} \rho^{\mathrm{in}}(X) - \rho^{\mathrm{out}}(X) \widehat{\mathcal{W}}\right\|_{\mathrm{op}} \le \varepsilon_{\mathrm{eq}} \|X\|. \tag{36}$$

Proof. The unitary group action preserves $\Gamma_{\bar{T}}^{\mathrm{out}}$ and therefore also preserves its orthogonal complement, and $P_{\bar{T}}^{\mathrm{out}}$ commutes with the action. For a compatible input x,

$$\begin{aligned} &\widehat{\mathcal{W}} \rho^{\mathrm{in}}(X) x - \rho^{\mathrm{out}}(X) \widehat{\mathcal{W}} x \\ &\quad = P_{\bar{T}}^{\mathrm{out}} \left(W \rho^{\mathrm{in}}(X) - \rho^{\mathrm{out}}(X) W\right) x. \end{aligned}$$

The operator norm of the orthogonal projector is 1, so taking norms gives Equation (36). The globalization error is controlled separately by Corollary D.5. □

E Identifiability up to Gauge

E.1 Local Symmetry-Subspace Recovery

Let Θ_i be the finite-dimensional Hilbert space containing candidate generator pairs

$$\theta = (A^{\mathrm{in}}, A^{\mathrm{out}})$$

where A^{in} and A^{out} act on the local input and output spaces, respectively. Linear priors such as metric preservation can be encoded in Θ_i. Let N_{obs} be the number of observed inputs. For the nth input $a^{(n)}$, let $D\mathcal{S}_{a^{(n)}}$ denote the Fréchet derivative of $\mathcal{S}$ at that input. For $a^{(1)}, \ldots, a^{(N_{\mathrm{obs}})}$, define the stacked infinitesimal-equivariance measurement operator

$$\mathcal{M}_i \theta := \left(D\mathcal{S}_{a^{(n)}}[A^{\mathrm{in}} a^{(n)}] - A^{\mathrm{out}} \mathcal{S}(a^{(n)})\right)_{n=1}^{N_{\mathrm{obs}}}. \tag{37}$$

Coordinate actions, Lie derivatives, and patch restrictions can be included in Equation (37) as known linear terms.

Assumption E.1 (Persistent excitation)**.** The true candidate symmetry subspace $G_i \subset \Theta_i$ satisfies

$$G_i = \ker \mathcal{M}_i.$$

and

$$\gamma_i := \inf_{\substack{\theta \perp G_i \\ \|\theta\|=1}} \|\mathcal{M}_i \theta\| > 0. \tag{38}$$

Theorem E.2 (Local Lie-subspace recovery)**.** *Suppose that* $\dim G_i = r_i$*, let* $\widehat{\mathcal{M}}_i$ *be the estimated measurement operator obtained from finite samples and a surrogate Jacobian, and let* $\delta_i := \left\|\widehat{\mathcal{M}}_i - \mathcal{M}_i\right\|_{\mathrm{op}}$ *be the measurement-operator perturbation. Assume that*

$$\delta_i < \gamma_i/2.$$

Let $\widehat{G}_i$ *be the subspace spanned by the* r_i *smallest right singular vectors of* $\widehat{\mathcal{M}}_i$*. Then*

$$\left\|P_{\widehat{G}_i} - P_{G_i}\right\|_{\mathrm{op}} \le \frac{2\delta_i}{\gamma_i}. \tag{39}$$

If r_i *is unknown, choose any*

$$\tau_i \in (\delta_i, \gamma_i - \delta_i),$$

The number of singular values of $\widehat{\mathcal{M}}_i$ *below* τ_i *is equal to* r_i*.*

Proof. Since $\mathcal{M}_i|_{G_i} = 0$,

$$\left\|\widehat{\mathcal{M}}_i|_{G_i}\right\| \le \left\|\widehat{\mathcal{M}}_i - \mathcal{M}_i\right\|_{\mathrm{op}} \le \delta_i.$$

the min–max principle for singular values shows that the r_ith smallest singular value of $\widehat{\mathcal{M}}_i$ is at most δ_i. Therefore, for any unit vector $\widehat{g} \in \widehat{G}_i$,

$$\left\|\widehat{\mathcal{M}}_i\widehat{g}\right\| \le \delta_i.$$

Consequently,

$$\begin{aligned}\|\mathcal{M}_i\widehat{g}\| &\le \left\|\widehat{\mathcal{M}}_i\widehat{g}\right\| + \left\|(\widehat{\mathcal{M}}_i - \mathcal{M}_i)\widehat{g}\right\| \\ &\le 2\delta_i.\end{aligned}$$

Write

$$\widehat{g} = P_{G_i}\widehat{g} + (I - P_{G_i})\widehat{g}.$$

From $\mathcal{M}_i P_{G_i} = 0$ and Equation (38),

$$\gamma_i \,\|(I - P_{G_i})\widehat{g}\| \le \|\mathcal{M}_i\widehat{g}\| \le 2\delta_i.$$

Taking the supremum over unit vectors in $\widehat{G}_i$ shows that the sine of the largest principal angle is at most $2\delta_i/\gamma_i$. Since G_i and $\widehat{G}_i$ have the same dimension, the orthogonal-projector difference equals this largest sine, proving Equation (39).

Finally, Weyl's singular-value perturbation bound shows that the r_i zero singular values of $\mathcal{M}_i$ are at most δ_i after perturbation, while all remaining singular values are at least $\gamma_i - \delta_i$. A threshold in the stated interval therefore recovers r_i. □

Remark E.3 (Observable error and gap). Surrogate-Jacobian error, observation noise, and finite-difference error jointly enter the perturbation bound δ_i. The persistent-excitation gap γ_i measures the strength with which the observed inputs excite nonsymmetry directions. The experiments report the empirical spectral gap, the subspace-projector distance across random seeds, and the prediction error.

Proposition E.4 (Basis-aligned stability of the structure tensor)**.** *Let* $\mathfrak{u}(d)$ *be the Lie algebra of* $d \times d$ *skew-Hermitian matrices, let* $G, \widehat{G} \subset \mathfrak{u}(d)$ *be* r*-dimensional Lie subalgebras, and define the subspace perturbation* $\eta := \left\|P_G - P_{\widehat{G}}\right\|_{\mathrm{op}}$*. Assume that*

$$\eta < 1.$$

Let tr *denote the matrix trace, and let* $\langle A, B\rangle_F := \mathrm{tr}(A^*B)$ *and* $\|A\|_F := \sqrt{\langle A, A\rangle_F}$ *denote the Frobenius inner product and norm; define the commutator by* $[A, B] := AB - BA$*. One can choose Frobenius-orthonormal principal-vector bases* $A_1, \ldots, A_r$ *and* $\widehat{A}_1, \ldots, \widehat{A}_r$ *such that*

$$\left\|\widehat{A}_a - A_a\right\|_F \le \sqrt{2}\eta.$$

For $a, b, c \in \{1, \ldots, r\}$*, define the basis-aligned structure tensors*

$$c_{ab}^{\ c} = \langle [A_a, A_b], A_c\rangle_F, \qquad \widehat{c}_{ab}^{\ c} = \langle [\widehat{A}_a, \widehat{A}_b], \widehat{A}_c\rangle_F,$$

Then

$$\|\widehat{c} - c\|_F \le 6\sqrt{2}\, r^{3/2}\eta. \tag{40}$$

Proof. Let θ_a be the principal angles. For the corresponding principal-vector bases,

$$\left\|\widehat{A}_a - A_a\right\|_F = 2\sin(\theta_a/2) \le \sqrt{2}\sin\theta_a \le \sqrt{2}\eta.$$

The operator norm of a unit-Frobenius-norm matrix is at most 1, and

$$\|[X,Y]\|_F \le 2\|X\|_{\mathrm{op}}\|Y\|_F.$$

Let $\varepsilon = \sqrt{2}\eta$. Adding and subtracting intermediate terms gives

$$\begin{aligned}|\widehat{c}_{ab}^{\,c} - c_{ab}^{\,c}| \le &\left|\langle[\widehat{A}_a - A_a, \widehat{A}_b], \widehat{A}_c\rangle_F\right| \\ &+ \left|\langle[A_a, \widehat{A}_b - A_b], \widehat{A}_c\rangle_F\right| \\ &+ \left|\langle[A_a, A_b], \widehat{A}_c - A_c\rangle_F\right| \\ &\le 6\varepsilon.\end{aligned}$$

There are r^3 tensor entries; taking their Frobenius norm gives Equation (40). □

Remark E.5 (Lie-algebra family). Equation (40) controls the structure tensor after orthogonal basis alignment. Recovery of a discrete isomorphism class additionally requires a separation margin for the candidate Lie-algebra family or a restriction to a locally rigid family. In the general case, the recovery objects are the subspace and the structure tensor up to a basis change.

E.2 Transport Ambiguity and Gauge Classification

Definition E.6 (Automorphism group of local data)**.** Given L local anchor operators $\{W_i^{(\ell)}\}_{\ell=1}^{L}$, where $W_i^{(\ell)} : V_i^{\mathrm{in}} \to V_i^{\mathrm{out}}$, let $\mathcal{D}_i$ denote the local data consisting of these local representations and anchor operators, and define

$$\begin{aligned}\mathrm{Aut}(\mathcal{D}_i) := \big\{(C^{\mathrm{in}}, C^{\mathrm{out}}) : &C^s \in U(V_i^s), \quad s \in \{\mathrm{in},\mathrm{out}\}, \\ &C^s\rho_i^s(X) = \rho_i^s(X)C^s,\ \forall X \in \mathfrak{g}_i,\ s \in \{\mathrm{in},\mathrm{out}\}, \\ &C^{\mathrm{out}}W_i^{(\ell)} = W_i^{(\ell)}C^{\mathrm{in}},\ \forall \ell\big\}.\end{aligned}$$

Theorem E.7 (Residual ambiguity of the transport solution set)**.** *Fix the source and target representations, the algebra isomorphism ϕ_{ji}, and the anchor pairs $(W_i^{(\ell)}, W_j^{(\ell)})$. For $s \in \{\mathrm{in},\mathrm{out}\}$ and $k \in \{0,1\}$, let $T_k^s : V_i^s \to V_j^s$. If two pairs of unitary transports*

$$(T_0^{\mathrm{in}}, T_0^{\mathrm{out}}), \qquad (T_1^{\mathrm{in}}, T_1^{\mathrm{out}})$$

both align the representations and satisfy naturality for every ℓ, then

$$C^s := (T_0^s)^{-1}T_1^s$$

form an element of $\mathrm{Aut}(\mathcal{D}_i)$. Conversely, for any $(C^{\mathrm{in}}, C^{\mathrm{out}}) \in \mathrm{Aut}(\mathcal{D}_i)$,

$$(T_0^{\mathrm{in}}C^{\mathrm{in}}, T_0^{\mathrm{out}}C^{\mathrm{out}})$$

still satisfies representation alignment and naturality. Thus, a nonempty transport solution set is a torsor for $\mathrm{Aut}(\mathcal{D}_i)$.

Proof. The two sets of transports align the same representations, so

$$T_k^s\rho_i^s(X) = \rho_j^s(\phi_{ji}X)T_k^s, \qquad k \in \{0,1\}.$$

Substitute $T_1^s = T_0^sC^s$ into the identity for $k = 1$, compare it with the identity for $k = 0$, and multiply on the left by $(T_0^s)^{-1}$ to obtain

$$C^s\rho_i^s(X) = \rho_i^s(X)C^s.$$

The two naturality relations are

$$T_k^{\mathrm{out}}W_i^{(\ell)} = W_j^{(\ell)}T_k^{\mathrm{in}}.$$

Again substitute $T_1^s = T_0^sC^s$, and use the relation for $k = 0$ to eliminate the target quantities, giving

$$C^{\mathrm{out}}W_i^{(\ell)} = W_i^{(\ell)}C^{\mathrm{in}}.$$

Thus, $C \in \mathrm{Aut}(\mathcal{D}_i)$. The converse follows by direct substitution. This right action is free and transitive, so the solution set forms a torsor. □

Theorem E.8 (Gauge classification of flat connections)**.** *Let N be connected, let G be a matrix group, and let $\mathrm{Hom}(\pi_1(N, i_0), G)$ denote the set of group homomorphisms from the fundamental group to G. The vertex-gauge equivalence classes of flat G-connections are in natural bijection with*

$$\mathrm{Hom}(\pi_1(N, i_0), G)/G \tag{41}$$

where G acts on representations by global conjugation.

Proof. Given a flat connection, compose transports along loops based at the base point to obtain a representation of $\pi_1(N, i_0)$. A vertex gauge cancels at all intermediate vertices and conjugates the holonomy representation only at the base point, giving a map from gauge classes to Equation (41).

Conversely, choose a spanning tree of the overlap graph and fix the base point i_0. Recursively choosing vertex gauges from the base point along the tree makes every tree-edge transport the identity. Each non-tree edge and the unique connecting path in the tree form a fundamental cycle; in the tree gauge, the transport on that non-tree edge equals the corresponding cycle holonomy. A holonomy representation therefore determines all edge transports in the tree gauge. If two holonomy representations differ by conjugation at the base point, applying the corresponding constant gauge gives the same tree-gauge connection. The two directions are inverse. □

Corollary E.9 (Gauge classification of a lifted U(1) connection)**.** *Within a fixed lift sector, let the gauge relation be $a \sim a + d_0\lambda$, and let $[a]$ denote the gauge-equivalence class of a. Let $P_{\mathcal{H}^1}$ be the orthogonal projection onto the harmonic subspace $\mathcal{H}^1$. Then the map*

$$[a] \longmapsto (d_1 a, P_{\mathcal{H}^1} a) \tag{42}$$

gives the linear isomorphism

$$C^1(N; \mathbb{R})/\operatorname{im} d_0 \cong \operatorname{im} d_1 \oplus \mathcal{H}^1. \tag{43}$$

Thus, face curvature and harmonic periods determine the gauge orbit of a lifted U(1) connection.

Proof. The Hodge decomposition gives

$$C^1(N; \mathbb{R}) = \operatorname{im} d_0 \oplus \operatorname{im} d_1^* \oplus \mathcal{H}^1.$$

After taking the quotient by $\operatorname{im} d_0$, each gauge class has a unique representative $b+h$, where $b \in \operatorname{im} d_1^*$ and $h \in \mathcal{H}^1$. The restricted map

$$d_1 : \operatorname{im} d_1^* \longrightarrow \operatorname{im} d_1$$

is a bijection. Surjectivity follows from the definition of $\operatorname{im} d_1$. If $b = d_1^*\psi$ and $d_1 b = 0$ for some $\psi \in C^2(N; \mathbb{R})$, then

$$0 = \langle \psi, d_1 d_1^* \psi \rangle = \|d_1^* \psi\|^2 = \|b\|^2,$$

so the kernel is zero. Thus, $d_1 a$ uniquely determines b, while $P_{\mathcal{H}^1} a$ uniquely determines h. This proves Equations (42)–(43). The coexact representative can also be written as

$$b = d_1^* (d_1 d_1^*)^\dagger d_1 a,$$

where † denotes the Moore–Penrose pseudoinverse. □

Remark E.10 (Scope of identifiable objects). The twisted spectrum and cohomology dimensions are invariant under a unitary gauge change. Gauge-orbit classification additionally requires the conjugacy class of the holonomy representation. A global flat-approximation bound for a general noncommutative curved connection requires a compact group, a small-curvature regime, a fixed topological sector, and gauge fixing. Our curvature-correction theorem uses the lifted U(1)/SO(2) setting. Recovery of a discrete isomorphism class for a general Lie algebra requires rigidity or separation of the candidate family.

F Experimental Protocols and Supplementary Results

F.1 Tasks and Evaluation Protocols

We construct three types of tasks. **Controlled connection recovery** varies the pure-gauge, curvature, and harmonic components independently and measures local symmetry, face curvature, cycle holonomy, and Wilson loops under U(1) and SU(2) transport. **Hidden-connection operator prediction and transfer** varies holonomy, local defects, medium, topology, and mesh for a nonlinear U(1) operator, a magnetic Schrödinger operator, and WebKB magnetic graphs. **Polycrystalline local-frame and response prediction** recovers grain coordinates and weakened grain boundaries from anisotropic responses, and evaluates response error in distribution and under changes in grain-boundary density. Public **1DSfM rotation averaging** uses relative rotations, scene splits, and globally aligned angular error to evaluate an $\mathrm{SO}(3)$ connection in public data.

The first three task classes specify random seeds, data splits, observation budgets, and metrics. Gauge-dependent errors are computed on conjugacy invariants or after global alignment according to the task protocol; prediction error is computed on held-out forcings.

F.2 Controlled U(1) Generative Model

The base complex is a triangulated annulus with $V = 24$ vertices, $E = 48$ edges, $F = 24$ faces, and first Betti number $b_1 = 1$. The fiber is $\mathbb{C}^4$. Let f be a forcing, let u be the corresponding solution field, let a be an edge-angle 1-cochain, let D_a be the U(1)-twisted incidence defined by a, and let D_a^* be its Hermitian adjoint. Let B_i be a Hermitian positive-definite local medium block, let $\mathrm{diag}(B_i)$ be the corresponding block-diagonal matrix, let I be the identity matrix, and let $\mu = 0.5$ be a regularizing mass term. The linear solution operator is defined by

$$L_a = D_a^* D_a + \mathrm{diag}(B_i) + \mu I, \qquad u = L_a^{-1} f \tag{44}$$

Connection sampling uses the cellular coboundary δ_0 from vertices to edges and the adjoint coboundary δ_1^* from faces to edges. Let ϕ be the vertex phase, ψ the face potential, and h a harmonic 1-cochain, sampled independently according to

$$a = \delta_0 \phi + \delta_1^* \psi + h \tag{45}$$

The term ϕ generates a hidden local gauge, ψ generates sparse curvature on 2–3 faces, and h controls holonomy on the noncontractible loop. Let $\varepsilon \geq 0$ be the strength of the pointwise cubic nonlinearity, and let $\odot$ denote component-wise multiplication; the nonlinear task uses

$$L_a u + \varepsilon |u|^2 \odot u = f, \tag{46}$$

where ε, the local medium blocks, and the connection are all unknown. Observed inputs contain equal numbers of low-pass random fields and white-noise fields to avoid near-collinearity in the local design matrices at high-degree vertices.

Table 4 supplements Table 1(a) with all evaluation axes and branch ablations. The first three axes retain the training medium, while the last three change both the medium and topology. SDNO estimates the connection and medium parameters separately for each episode under a new medium.

Table 4: Median relative operator error on each evaluation axis of the nonlinear protocol. Values are means over five random seeds.

Method	In-distribution	New holonomy	New defects	Disk	Two holes	New mesh
Descent and defect branches	**0.009**	**0.009**	**0.009**	**0.021**	**0.022**	0.021
Without defect branch	0.011	0.011	0.025	0.022	0.023	0.022
Compressed holonomy	0.033	0.089	0.041	0.022	0.047	0.031
Unconstrained complex blocks	0.113	0.115	0.114	4.94	4.67	4.18
generic GNN	0.102	0.105	0.101	0.405	0.384	0.343
Equivariant GNN	0.131	0.133	0.131	0.588	0.441	0.309
SDNO	0.019	0.020	0.019	**0.021**	0.028	**0.020**

The equivariant corrector in SDNO contains $47{,}036$ trainable parameters shared across episodes. The descent-and-defect-branches row isolates the contribution of structure estimation.

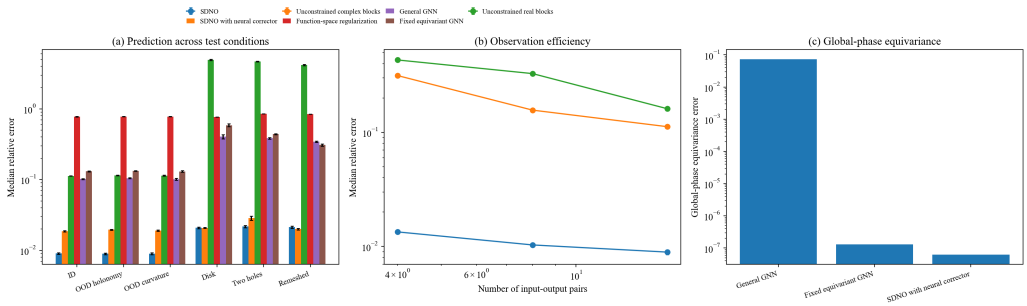


Figure 4: Transfer, observation efficiency, and global gauge-equivariance error for the nonlinear U(1) operator. (a) Prediction errors under different test conditions. (b) Sweep over the number of observations. (c) Equivariance error under an unseen global phase transformation. Error bars show the standard deviation across random seeds.

Table 5: Structure constraints and branch ablations on the controlled U(1) task. N_{obs} is the number of observation pairs per episode; values are median relative operator errors over 48 test episodes, with observation-noise standard deviation 0.02.

Method	Real params. per edge	$N_{\mathrm{obs}} = 2$	$N_{\mathrm{obs}} = 8$	$N_{\mathrm{obs}} = 24$
General real matrix	64	0.557	0.302	0.103
Complex-linear matrix	32	0.457	0.085	0.087
SDNO	1	0.033	0.014	0.012
SDNO, direct transport estimation	1	**0.028**	**0.013**	**0.011**
Without defect branch	1	0.033	0.021	0.020
Without loop holonomy	1	0.067	0.062	0.061

F.3 Symmetry Recovery and Misspecification Detection

Local symmetry recovery constructs commutator measurements over a skew-symmetric candidate space. Let δ be the measurement-operator perturbation norm, and let γ be the smallest singular value of the true measurement operator on the orthogonal complement of the symmetry subspace. Controlled perturbations span three orders of magnitude and are restricted to the theoretical range $\delta < \gamma/2$. All 60 experiments satisfy Equation (39), and the median ratio between empirical error and the upper bound is 0.13. Data-driven recovery exhibits a stable spectral gap after the number of observation pairs reaches the identifiability threshold for local row fitting; with no noise and $N_{\mathrm{obs}} = 128$, the median projection error is 0.011.

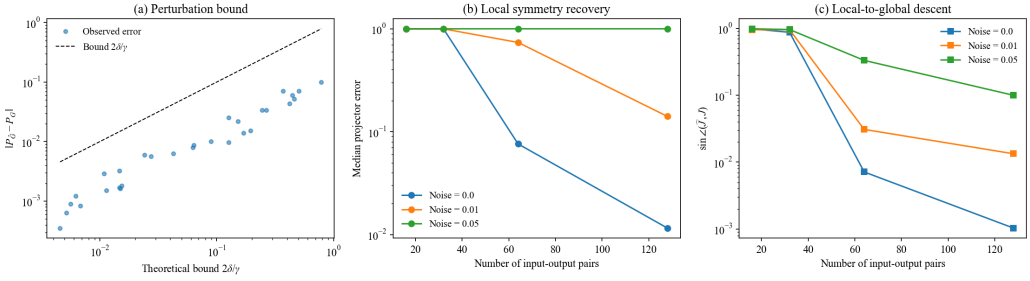


Figure 5: From local symmetry recovery to global descent. (a) Verification of the subspace perturbation bound. (b) Local symmetry recovery as the number of observation pairs and noise vary. (c) Cross-patch consistency reduces a local four-dimensional candidate space to a global one-dimensional $\mathfrak{u}(1)$.

Evaluation of transport structure uses gauge invariants and does not compare single-edge phases in an arbitrary gauge. Staged estimation first fits edge transport within the U(1) structure and then reads out cycle holonomy and face curvature. Figure 6 shows that this procedure clearly outperforms joint fitting of all edge blocks with few observation pairs and recovers holonomy over the entire fundamental interval.

Transport estimation aligns the local Lie subspaces of neighboring patches within the candidate structure group. Let $\mathcal{E}_{\mathrm{ov}}$ be the set of oriented edges formed by overlapping patch pairs, let d_{Gr} be the projection distance on the Grassmannian, and define the adjoint action $\mathrm{Ad}_T(A) := TAT^{-1}$. For an input or output representation ρ_i, the corresponding objective is

$$\mathcal{L}_{\mathrm{align}} = \sum_{(i,j)\in\mathcal{E}_{\mathrm{ov}}} d^2_{\mathrm{Gr}}\big(\rho_j(\mathfrak{g}_j), \mathrm{Ad}_{T_{ij}}\,\rho_i(\mathfrak{g}_i)\big)\,, \tag{47}$$

where T_{ij} is the candidate transport from patch i to patch j. The same objective is applied separately to the input and output representations.

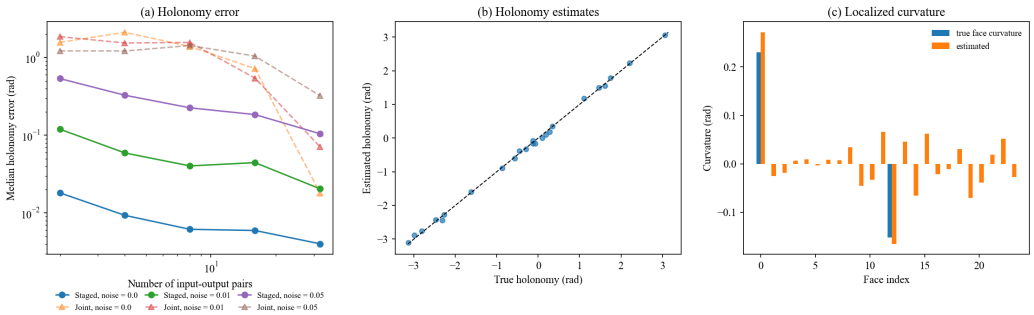


Figure 6: Recovery of gauge invariants under a hidden local gauge. (a) Cycle-holonomy error for staged and joint estimation. (b) Ground-truth and recovered holonomy. (c) Recovery of the locations and magnitudes of sparse face curvature.

Figure 7 supplements Table 5 with sample-efficiency curves. Constraining the edge matrix to one U(1) parameter per edge reduces the error to the order of 10^{-2} with few observation pairs. Removing loop holonomy produces a persistent error floor in both structure recovery and prediction; when the harmonic component is retained, the equivariance error under an unseen global gauge remains at numerical precision.

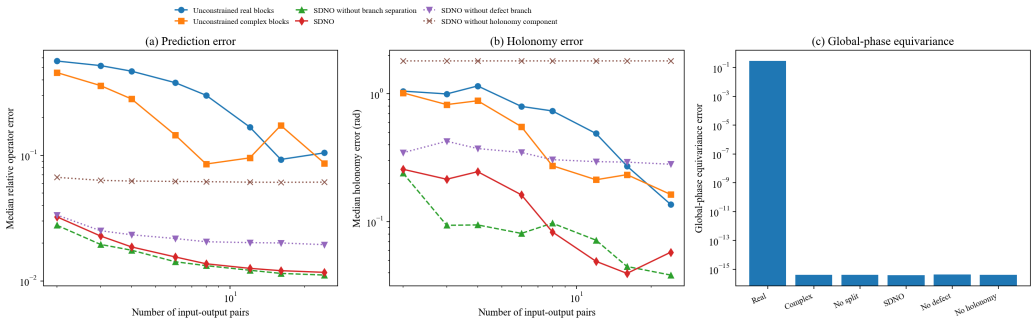


Figure 7: Structure constraints and branch ablations. (a) Held-out relative operator error. (b) Holonomy-recovery error. (c) Equivariance defect under an unseen global gauge. The observation-noise standard deviation is 0.02.

Symmetry misspecification is constructed with the antilinear term $u \mapsto Lu + \lambda_A M\bar{u}$, where L is a complex-linear reference operator, M is a block-diagonal mixing operator formed by random complex matrices on each patch, $\bar{u}$ is the complex conjugate of u, and $\lambda_A \geq 0$ controls the strength of antilinear contamination. Let $\lambda_{\min}$ and λ_2 be the smallest and second-smallest eigenvalues of

the symmetry-measurement quadratic form, respectively; as λ_A increases from 0 to 0.4, $\lambda_{\min}/\lambda_2$ increases from 8×10^{-4} to 0.14. For an edge e, let a_e be the edge phase, let K_e be a traceless skew-Hermitian matrix with unit operator norm, and let $\lambda_B \geq 0$ measure departure from the U(1) transport family; transport misspecification is constructed as $T_e = e^{ia_e} \exp(\lambda_B K_e)$. The U(1) fitting residual increases from 0.04 to 0.29 together with the excess prediction error of the structure branch. Under both abundant and limited observations, the ambient branch approaches the smaller error of the structured and unconstrained models.

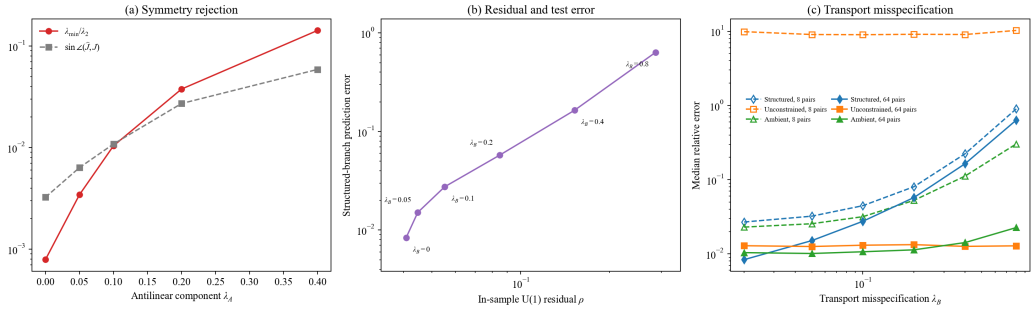


Figure 8: Rejection, error prediction, and degradation under structure misspecification. (a) When antilinear contamination removes U(1) symmetry, the discovery measurement no longer supports a one-dimensional null space. (b) The observable U(1) fitting residual predicts the error of the structure branch. (c) Prediction errors of the structure branch, unconstrained blocks, and ambient branch under two observation budgets.

F.4 Symmetry Discovery on a Neural Surrogate and Scaling

The JVP procedure uses an independent neural surrogate with approximately 2.0×10^5 trainable parameters to estimate local Jacobians. The shared-parameter column in Table 2 counts the equivariant corrector in the deployed predictor.

For a nonlinear surrogate $\widehat{\mathcal{S}}$, let f_n be the nth forcing, let A be a candidate infinitesimal generator, and let $D\widehat{\mathcal{S}}_{f_n}$ denote the Fréchet derivative of $\widehat{\mathcal{S}}$ at f_n. We compute the equivariance residual with a JVP:

$$r_n(A) = D\widehat{\mathcal{S}}_{f_n}[Af_n] - A\widehat{\mathcal{S}}(f_n), \tag{48}$$

where $r_n(A)$ is the infinitesimal-equivariance residual on the nth observation. We recover the generator from the smallest right singular vector of the stacked measurements. Figure 9 shows that as the held-out error of the surrogate decreases from 0.21 to 0.029, the sine of the angle between the recovered direction and the true generator decreases from 0.080 to 0.0064. Discovery quality therefore depends on the local accuracy of the surrogate Jacobian in addition to output error.

To rule out structural bias from a single measurement route, Figure 10 compares two independent routes: commutant stacking and JVP measurements. For a complex matrix X, let $\Re X$ and $\Im X$ denote its real and imaginary parts, and define the real block embedding $\mathrm{realify}(X) := \left(\begin{smallmatrix} \Re X & -\Im X \\ \Im X & \Re X \end{smallmatrix}\right)$; then $J := \mathrm{realify}(iI)$ is the real-representation generator of the global complex phase action. Both routes recover the same J in a skew-symmetric basis, with component-wise agreement in the locations, signs, and magnitudes of the nonzero coefficients.

The scaling sweep uses annulus cell counts $K \in \{12, 24, 32, 48, 64\}$, corresponding to complex dimensions from 96 to 512. Figure 11 shows that the structured cold-start error remains at 0.021 ± 0.001, while identification and solution time per episode increases from approximately 0.2 to 0.6 seconds; the error of unconstrained complex blocks under the same observation budget remains above 4.

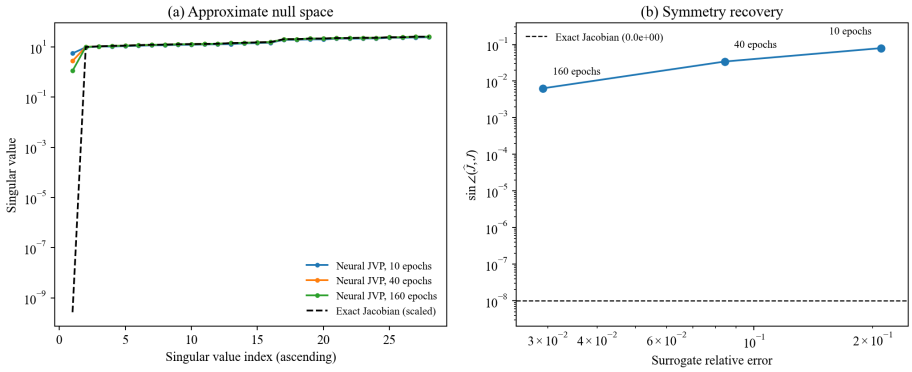


Figure 9: JVP symmetry discovery on a neural surrogate. (a) Formation of a one-dimensional near-zero singular space during surrogate training. (b) Generator angular error as the relative surrogate error decreases. The dashed line shows the result from the exact implicit Jacobian.

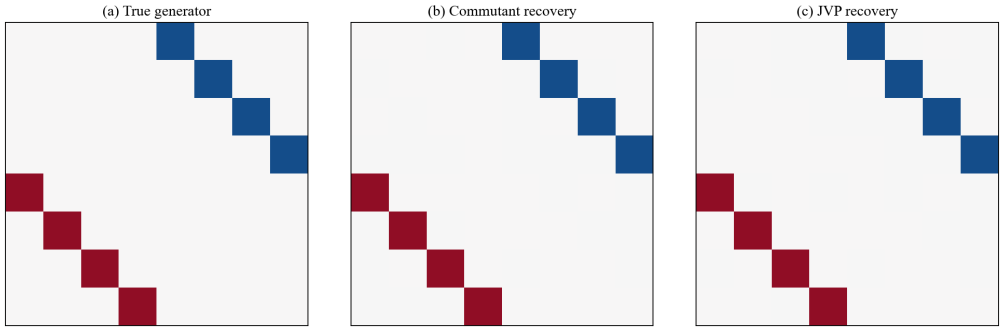


Figure 10: Agreement between independent generator recoveries. (a)–(c) Ground-truth generator, commutant-stacking recovery from 128 noisy observations, and JVP recovery using the implicit Jacobian, respectively. The nonzero structure, signs, and magnitudes of both recoveries agree with the ground-truth generator.

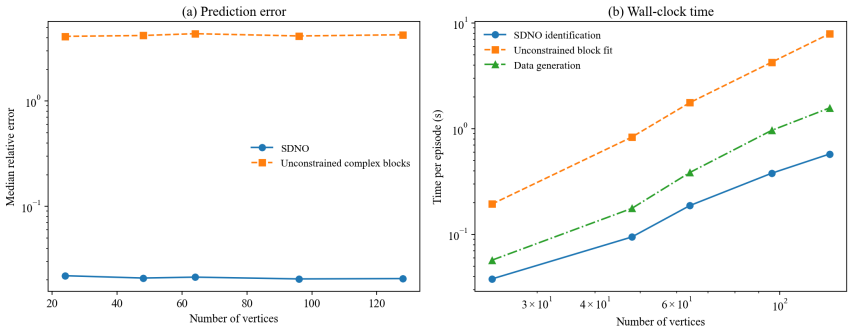


Figure 11: Scaling sweep of cold-start accuracy and runtime. (a) Relative errors of structured identification and unconstrained complex blocks. (b) Wall-clock time per episode for system identification and solution, and for data generation.

F.5 Noncommutative Holonomy Obstruction

The noncommutative experiment uses $V_i = \mathbb{C}^2 \otimes \mathbb{C}^3$, where $\mathbb{C}^2$ carries the fundamental representation of SU(2) and $\mathbb{C}^3$ is a three-dimensional multiplicity space. SU(2) transport constructs three states: a central Wilson loop, a generic Wilson loop, and nonzero curvature. Figure 12 compares the naturality spectra, Wilson-loop trace angles, and rotation-axis errors obtained from the ground-truth and data-estimated transports.

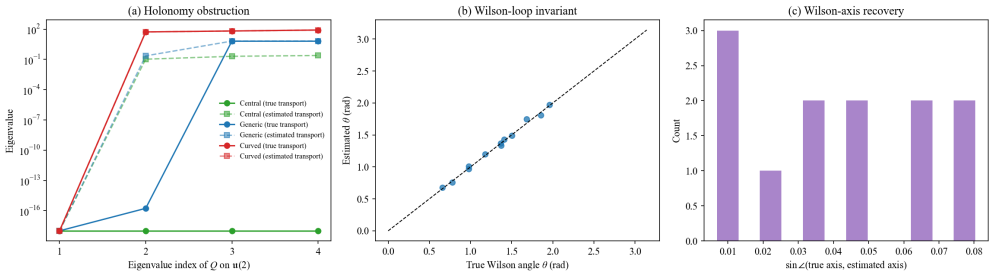


Figure 12: SU(2) holonomy obstruction. (a) Spectrum of the $\mathfrak{u}(2)$ naturality quadratic form and its $4/2/1$ dimensional cascade. (b) Wilson-loop trace-angle recovery. (c) Rotation-axis error. Solid lines use ground-truth transport, and dashed lines use data-estimated transport.

F.6 Physical and Public Graph Structures

The magnetic Schrödinger experiment uses a Peierls discretization on a 6×40 annular grid. Let H be the Hamiltonian, let A be the magnetic vector potential, let $V(x)$ be the scalar disorder potential at position x, let $i = \sqrt{-1}$, and let ∇ be the gradient operator with respect to spatial coordinates. Then

$$H = (-i\nabla - A)^2 + V(x). \tag{49}$$

The annular flux has zero curvature in the sample domain and represents the Aharonov–Bohm phase through cycle holonomy. Two flux tubes penetrating the sample produce local curvature. Figure 14 shows the recovery of the flux sweep, holonomy, and disorder potential by the identified operator.

Magnetic Schrödinger comparison protocol. The comparison uses a fixed $V(x)$ and resamples the Aharonov–Bohm flux, local flux tubes, and hidden per-vertex phase convention in each episode. The training set contains 256 episodes; each episode provides 16 pairs of context forcings and responses and four query forcings. Each evaluation axis contains 32 episodes, and the observation-noise standard deviation is 0.02. Let Φ denote the total flux through the annulus. The training range is $\Phi \in [0, \pi)$ with two flux tubes; OOD-holonomy uses $\Phi \in [\pi, 2\pi)$, and OOD-defect increases the number of flux tubes to four. All data are reconstructed in the same order from the same random seed.

FNO-2d uses four spectral-convolution layers and a frequency-domain periodic parameterization; DeepONet uses a branch–trunk decomposition. Query-only and context-conditioned results are reported for both. Let (r, θ) denote the radial and angular coordinates of the annular grid. Dilated ResNet contains two residual stacks with dilation rates $(1, 2, 4, 2, 1)$, circular padding in the θ direction, and zero padding in the r direction. Operator Transformer represents each vertex by a token containing its position, query forcing, and all context values, and outputs a complex response using a four-layer, six-head self-attention encoder of width 96. All neural baselines are trained for 32 epochs with Adam, an initial learning rate of 10^{-3}, and a cosine schedule.

The equivariant corrector in SDNO takes a gauge-invariant context covariance as the hypernetwork input and outputs a response correction through complex-linear channel mixing; this configuration contains $21{,}926$ shared trainable parameters. At test time, the model fits 881 episode parameters from 16 context observation pairs, including 640 edge phases, 240 per-vertex potential values, and one scalar operator coefficient. Equivariance error is the relative prediction deviation after applying an unseen global phase simultaneously to the forcing, context, and response; the complex-linear corrector remains equivariant under this action. Table 2 reports the median relative ℓ_2 response error.

Cross-task evaluation. Table 6 compares the nonlinear hidden-connection operator and polycrystalline response under the same context information. The former is evaluated in distribution, under new holonomy, and under new curvature defects on the training annular complex, with additional zero-shot evaluations on a remeshed annulus, a disk, and a two-hole complex; the latter is evaluated in distribution and under changes in weakened-grain-boundary density on a 7×10 polycrystalline grid. The two tasks use 16 and four context observation pairs, respectively, and all values are median relative ℓ_2 response errors.

Table 6: Cross-task errors of the prediction methods. The upper table reports the nonlinear hidden-connection operator and the lower table reports the polycrystalline response. Lower is better. N/A indicates that the model input parameterization does not support the corresponding discretization or topology.

Nonlinear hidden-connection operator, 16 context observation pairs

Method	In-distribution	New holonomy	New curvature	New mesh	Disk	Two holes
FNO-2d (context)	0.540	0.525	0.497	0.534	N/A	N/A
Dilated ResNet (context)	0.536	0.526	0.469	0.510	N/A	N/A
DeepONet (context)	0.969	0.974	0.964	N/A	N/A	N/A
Operator Transformer (context)	0.600	0.600	0.573	0.608	0.653	0.638
SDNO	**0.022**	**0.021**	**0.020**	**0.020**	**0.020**	**0.021**

Polycrystalline response, four context observation pairs

Method	In-distribution	New grain-boundary density
FNO-2d (context)	0.267	0.261
Dilated ResNet (context)	0.223	0.228
DeepONet (context)	0.629	0.648
Operator Transformer (context)	0.300	0.291
SDNO	**0.188**	**0.184**

The tensor-grid parameterizations of FNO and Dilated ResNet support remeshing of the annulus with the same topology, but do not support the irregular disk and two-hole complexes used here. The branch-input dimension of DeepONet is fixed to the training sensor set, so it is N/A on the three axes where the vertex set changes. Operator Transformer uses coordinate tokens and can operate on all three new discretizations; its errors on the remeshed annulus, disk, and two-hole complex are $0.608/0.653/0.638$, respectively.

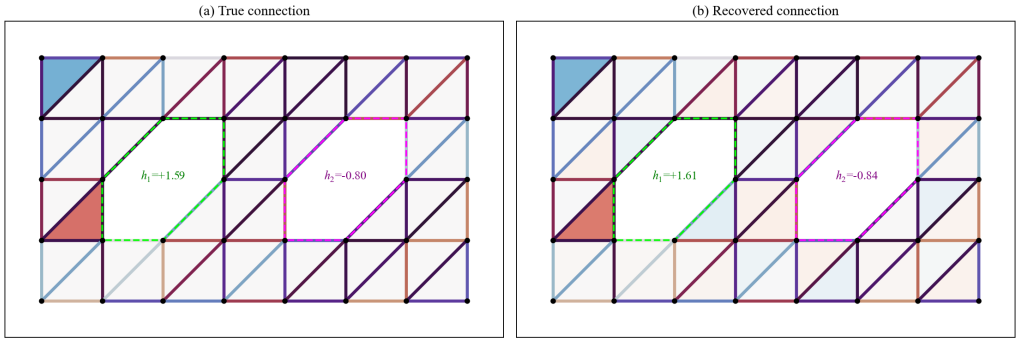


Figure 13: Zero-shot topology transfer on a two-hole complex. (a) Ground-truth connection. (b) Recovery from 16 input–output observations under a new medium and hidden local gauge. Dashed lines mark the two generating loops and their holonomies.

For an oriented edge $e = (u, v)$ in a directed WebKB graph, let A_{uv} be the adjacency-matrix entry from vertex u to vertex v, and set the direction parameter to $q = 1/4$. The standard parameterization of the magnetic Laplacian defines the edge phase by $a_e = 2\pi q(A_{uv} - A_{vu})$. A unidirectional edge carries phase $\pm\pi/2$, while a reciprocal edge has zero phase. Table 7 reports prediction and structure-recovery errors from 32 observation pairs.

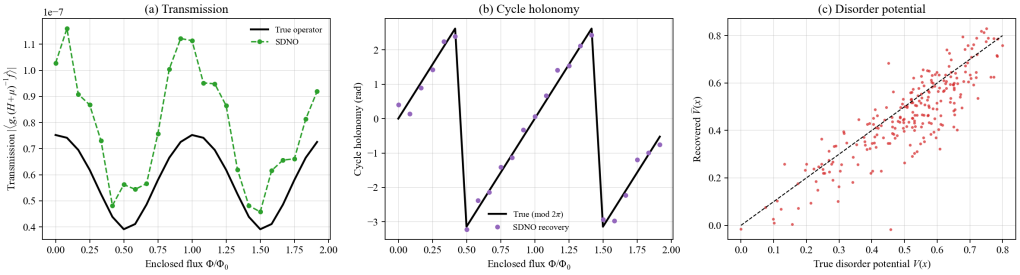


Figure 14: Physical readouts on the magnetic Schrödinger annulus. (a) Aharonov–Bohm oscillation under a flux sweep, with a period of one flux quantum. (b) Cycle holonomy. (c) Disorder-potential recovery, with point-wise correlation coefficient 0.862.

Table 7: Magnetic-connection recovery on directed WebKB graphs. Structure errors are in radians, and the prediction column reports median relative error.

Dataset	Prediction error	Holonomy error	Curvature error
Texas	0.007	0.120	0.122
Cornell	0.008	0.096	0.095
Wisconsin	0.008	0.213	0.102

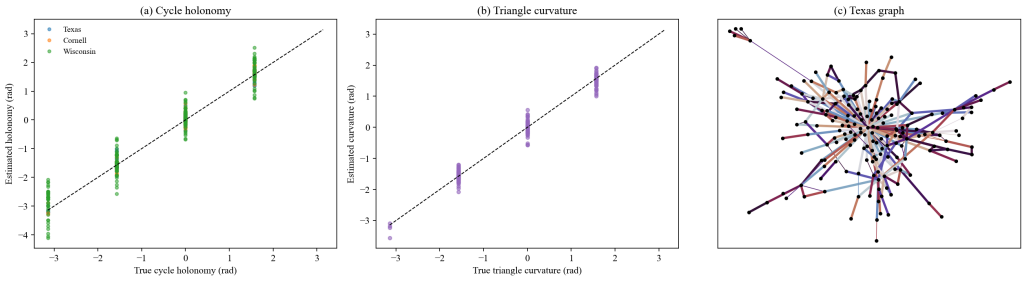


Figure 15: Magnetic connections on public directed graphs. (a) and (b) Ground-truth and recovered cycle holonomy and face curvature, respectively. (c) Estimated edge phases on the Texas graph at $q = 1/4$, with thick edges denoting unidirectional edges.

F.7 Polycrystalline Local Frames

The two-dimensional polycrystalline task contains 7×10 grains. Let φ_i be the material-axis orientation of the ith grain, defined modulo π; let $R(\varphi_i) := \begin{pmatrix} \cos\varphi_i & -\sin\varphi_i \\ \sin\varphi_i & \cos\varphi_i \end{pmatrix}$ be the planar rotation matrix; and let $(\cdot)^\top$ denote matrix transpose. The local stiffness A_i of the grain is

$$A_i = R(\varphi_i)\,\mathrm{diag}(3,1)R(\varphi_i)^\top, \tag{50}$$

where $\mathrm{diag}(3,1)$ gives the stiffness ratio along the principal material axes. We randomly select 8% of the grain boundaries and reduce their coupling coefficient from 1 to 0.25. Figure 16 shows the local frames, defective grain boundaries, and per-grain errors. With eight observation pairs and 1% noise, the median frame error is 0.008 radians and the defect-detection AUC is 1.0. With four observation pairs and 5% noise, the frame error is 0.122 radians and the AUC is 0.87.

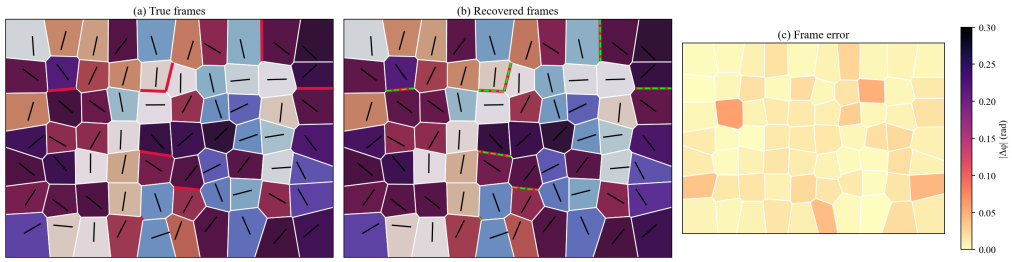


Figure 16: Polycrystalline anisotropy task. (a) Ground-truth grain frames and weakened grain boundaries. (b) Structure recovery. (c) Per-grain frame error.

Figure 17 further shows recovery as the observation budget varies. Grain-frame error decreases with the number of observation pairs, and weakened grain boundaries are already separable at a moderate observation budget. The gap between structured prediction and unconstrained blocks remains larger than an order of magnitude under high noise.

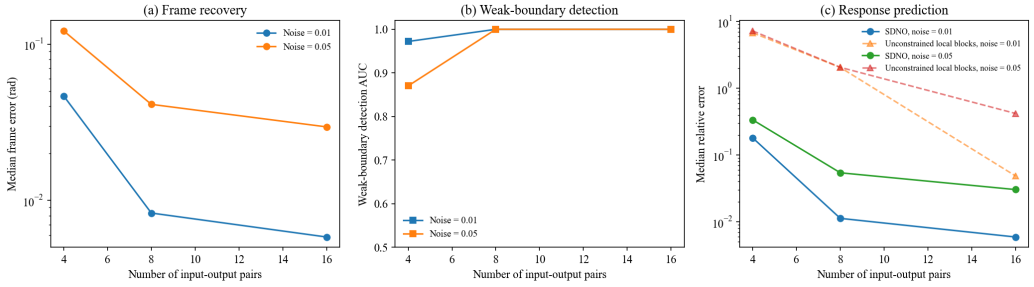


Figure 17: Observation-budget sweep on the polycrystalline task. (a) Median grain-frame error. (b) AUC for weakened-grain-boundary detection. (c) Prediction errors of structured identification and unconstrained blocks.

F.8 Rotation-Averaging Protocol and Edge Scoring

All 1DSfM experiments use the same largest connected component, relative-rotation convention, and global alignment. Let ℓ_1 denote the sum of absolute Lie-algebra rotation residuals. Weiszfeld and ℓ_1 IRLS use geometric-median updates and Lie-algebra reweighting, respectively. Shonan is run through GTSAM `ShonanAveraging3` with squared and Huber losses. NeuRoRA uses a one-step graph-network update trained on synthetic graphs.

SDNO-RA adds triangular-loop consistency-residual weights to the spanning-tree initialization and Lie-algebra updates of ℓ_1 IRLS.

The public 1DSfM data provide noisy pairwise relative rotations $\widetilde{R}_{ij} \in \mathrm{SO}(3)$ and Bundler absolute-rotation references (Wilson & Snavely, 2014). Rotation averaging estimates camera orientations $R_i \in \mathrm{SO}(3)$ such that $\widetilde{R}_{ij} \approx R_j R_i^{-1}$, and reports the mean and median angular errors after global alignment (Hartley et al., 2013). A triangular loop product that deviates from the identity is a local connection inconsistency. The comparison methods include spanning-tree initialization, the Weiszfeld geometric median (Hartley et al., 2011), robust IRLS based on absolute-rotation residuals (Chatterjee & Govindu, 2018), GTSAM Shonan (Dellaert et al., 2020), and a one-step NeuRoRA graph network (Purkait et al., 2020).

SDNO-RA reduces both the mean and median error of robust IRLS in five of the six scenes. Shonan and its Huber variant have higher errors than robust IRLS and SDNO-RA in all six scenes. Among the top 10% of edges selected by triangular-loop consistency score, the fraction of high-error edges is 0.77, compared with 0.13 among all edges; local loop inconsistency therefore provides direct information about outlier edges.

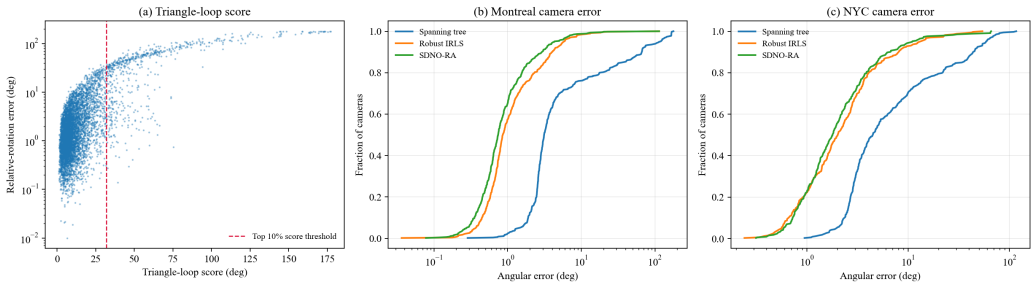


Figure 18: Edge scores and per-camera errors on 1DSfM. (a) Triangular-loop scores and ground-truth edge errors on Montreal Notre Dame; among the top 10% ranked edges, the fraction with ground-truth error above 10 degrees is 0.77, compared with 0.13 overall. (b) and (c) Per-camera angular-error distributions on Montreal Notre Dame and NYC Library.

G Reproducibility

The anonymous OSF artifact contains `SDNO-code-and-data.zip`, the SDNO benchmark release 0.1.0. The archive includes the installable `sdno` package, experiment entry points under `experiments`, the benchmark specification in `BENCHMARK.md`, and the standard evaluator `benchmark.py`. The environment and optional dependencies are specified in `pyproject.toml`. The eight smoke tests in `tests/test_smoke.py` check cellular-complex identities, connection recovery, nonlinear prediction, and noncommutative obstruction spectra.

Cached numerical results and publication figures are stored under `artifacts/results` and `artifacts/figures`. The archive also contains the neural rotation-averaging checkpoint used by the corresponding baseline. Controlled tasks generate synthetic observations from the fixed seeds and settings in their experiment files. Third-party data are not redistributed: `configs` records the selected 1DSfM scenes and WebKB graphs, while `scripts/download_1dsfm.sh` and `scripts/download_web_graphs.sh` download them from their public sources. File-level hashes are provided in `SHA256SUMS`.