# A STABILITY THEORY FOR GENERATOR-BASED CONTINUOUS SYMMETRY LEARNING

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Symmetry discovery learns transformations that preserve functions, data distributions, or the solution sets of differential equations, revealing structure that can guide model design and scientific understanding. Existing methods address these target objects through different formulations, making it difficult to identify common conditions for success. We develop a stability theory for generator-based symmetry discovery by formulating it as a subspace recovery problem through a task-specific symmetry-measuring operator. This provides a common framework for analyzing how surrogate models and numerical approximations affect recovered symmetries. We establish sufficient conditions for accurate recovery and derive an error bound governed by the surrogate operator error relative to the separation between symmetry and non-symmetry directions. A taxonomy clarifies how existing methods relate and which settings our theory covers. We validate the theoretical predictions in controlled environments across three settings, where the true symmetry subgroup, perturbation strength, and failure modes can be explicitly controlled.

## 1 INTRODUCTION

Symmetry plays an important role in modern machine learning (Bronstein et al., 2017). When the relevant symmetry group of a task is known in advance, it can be built into a model through invariant or equivariant architectures, such as group-equivariant convolution (Cohen & Welling, 2016), steerable convolution (Weiler & Cesa, 2019), or equivariant MLPs (Finzi et al., 2021), often improving sample efficiency and generalization (Wang et al., 2021). However, in many scientific and data-driven settings, the correct symmetry is not fully known before modeling. For example, it may be only approximately satisfied (Wang et al., 2022), dependent on the input region (Bhat et al., 2025), or expressed through a nonlinear action that is difficult to specify manually (Mitchel et al., 2024; Yang et al., 2024). These scenarios motivate the problem of *symmetry discovery*: learning the transformations under which a target mathematical object is invariant.

Recent years have seen a growing body of machine-learning methods for discovering continuous symmetries. Some methods learn Lie groups under which a target function is invariant from supervised datasets (Moskalev et al., 2022); others infer transformations that preserve a data distribution (Yang et al., 2023; Chen et al., 2025); still others discover symmetries of differential equations from solution trajectories (Ko et al., 2024; Hu et al., 2025a). These advances demonstrate the promise of learning symmetries from data, but differences in target objects, representations, and learning objectives can obscure the mathematical connections between methods. Their empirical successes motivate a common theoretical account: one that identifies shared principles and quantifies the conditions under which recovery is reliable.

A challenge shared across these settings is that symmetry must often be inferred from finite, potentially noisy data rather than evaluated directly on the ground-truth target object. Many methods therefore proceed in two steps: first constructing a surrogate model or numerical approximation from data, and then optimizing for transformations using this approximation to assess symmetry. For example, when the underlying function is unknown, a predictor is trained on input-output samples and used in place of the true function to evaluate candidate symmetries. Errors in these approximations can alter which transformations appear to preserve the target object. Understanding successful discovery therefore requires relating the accuracy of these approximations to the accuracy of the recovered symmetry.



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.

This raises our central question: *under what conditions does symmetry discovery from an imperfect surrogate reliably recover the underlying symmetry?*

In this paper, we address this question for *generator-based* discovery of *subgroup* symmetries through a common operator framework and a stability theory. Infinitesimal generators describe the directions of continuous transformations, and the true symmetry generators form a linear subspace. The key idea is to view generator learning as a *subspace recovery problem*: recovering the span of the true generators rather than a particular choice of basis. We characterize this subspace as the kernel of a task-specific symmetry-measuring operator, which maps each candidate generator to its infinitesimal symmetry violation. This perspective connects different discovery settings through a shared mathematical problem, allowing their sensitivity to approximation errors to be analyzed within the same framework.

Our main theoretical result gives explicit sufficient conditions for accurate recovery. Under sufficient regularization and global optimization, the recovered generator subspace is stable when the surrogate operator is accurate relative to the intrinsic separation between symmetry and non-symmetry directions. The resulting error bound quantifies how accurately the operator must be approximated to guarantee a desired recovery accuracy. We examine these predictions in controlled sandboxes for functional, distributional, and PDE symmetries, where the ground-truth symmetry and operator perturbations are explicitly available. To clarify the scope of the analysis, we also provide a taxonomy that distinguishes the target object, support structure, algorithmic paradigm, and action representation, situating generator-based subgroup discovery within the broader landscape of continuous symmetry learning. To summarize, our contributions are:

- **A supporting taxonomy.** We organize continuous symmetry discovery along four axes to clarify the assumptions of existing methods and the scope of our theoretical analysis.
- **A general operator framework.** We formulate generator-based symmetry learning as subspace recovery through a symmetry-measuring operator, providing a common framework for analyzing functional, distributional, and PDE symmetries.
- **A stability guarantee.** For a known symmetry dimension and a globally optimized objective, we establish sufficient conditions for accurate recovery and bound the generator-subspace error in terms of operator perturbation and spectral separation, under sufficient regularization.
- **Controlled empirical validation.** We construct sandboxes with known symmetry subgroups and controlled operator perturbations to evaluate the predicted recovery behavior and examine failures associated with loss of spectral separation and generator collapse.

## 2 A Taxonomy of Symmetry Discovery

Let $\mathcal{X}$ denote the input/data space, $\omega$ the target object whose symmetry is being studied, and $G$ an ambient connected Lie group of candidate transformations acting on $\mathcal{X}$. A task-specific symmetry relation $R_\omega(x, g) = 0$ induces a valid transformation set $S_x := \{g \in G : R_\omega(x, g) = 0\}$. The goal of symmetry discovery is to recover or approximate the family of valid transformation sets $\{S_x\}_{x\in\mathcal{X}}$. We separate this problem along four dimensions as follows.

**Target object.** This axis distinguishes the mathematical targets of symmetry discovery. For *functional symmetry* (Moskalev et al., 2022), the target is a function $f : \mathcal{X} \to \mathcal{Y}$, and the relation is $f(gx) = f(x)$, or $f(gx) = gf(x)$ for equivariance. For *distributional symmetry* (Yang et al., 2023), the target is a probability measure $q$, and the relation is $g_\# q = q$, where $g_\# q$ denotes the transformed measure. More recent work also considers *PDE symmetry* (Ko et al., 2024), where the target is the solution set of a PDE $\mathcal{F}[u] = 0$. Here, $\mathcal{F}$ is a differential operator and $u$ is a solution field. The symmetry relation $\mathcal{F}[u] = 0 \Rightarrow \mathcal{F}[gu] = 0$ requires that the transformed $gu$ is also a solution.

**Support structure.** We consider the *input domain locality* and *algebraic structure* of the valid transformation set. Global subgroup symmetry corresponds to a fixed subgroup $H \leq G$ that preserves the target object across the entire domain. More generally, the valid transformation set $S_x$ may depend on the input (Bhat et al., 2025) and need not satisfy closure under composition and inversion (Allingham et al., 2024). Such local, input-dependent subsets can express transformations that satisfy the symmetry relation only within a bounded range, beyond the scope of a fixed subgroup.

Table 1: Categorizing representative continuous symmetry discovery methods under our taxonomy.

| Method | Target Object | Support Structure | | Algorithm | Action Type |
|---|---|---|---|---|---|
| | | Locality | Subgroup/Subset | | |
| LieGAN (Yang et al., 2023) | Distribution | Global | Subgroup | Generator-based | Linear |
| LieGG (Moskalev et al., 2022) | Function | Global | Subgroup | Generator-based | Linear |
| LaLiGAN (Yang et al., 2024) | Distribution | Global | Subgroup | Generator-based | Nonlinear |
| Shaw et al. (2024) | Function | Global | Subgroup | Generator-based | Nonlinear |
| LieSD (Hu et al., 2025b) | Function | Global | Subgroup | Generator-based | Linear |
| Karjol et al. (2026) | Function | Global | Subgroup | Generator-based | Linear |
| Ko et al. (2024) | PDE | Global | Subgroup | Generator-based | Nonlinear |
| LieNLSD (Hu et al., 2025a) | PDE | Global | Subgroup | Generator-based | Nonlinear |
| LieDynNet (Liu et al., 2025) | PDE | Global | Subgroup | Generator-based | Nonlinear |
| MLSD (Hou et al., 2026) | Hamiltonian | Global | Subgroup | Generator-based[a] | Nonlinear |
| Otto et al. (2025) | Function | Global | Subgroup | Generator-based | Nonlinear |
| Desai et al. (2022) | Distribution | Global | - | Single element[b] | Linear |
| Augerino (Benton et al., 2020) | Function | Global | Subset | Distribution-based | Linear |
| AtlasD (Bhat et al., 2025) | Function | Local | Subgroup | Generator-based | Linear |
| Allingham et al. (2024) | Distribution | Local | Subset | Distribution-based | Linear |
| LieFlow (Chen et al., 2025) | Distribution | Local | Subset | Distribution-based | Linear |
| RECON (Urbano et al., 2025) | Distribution | Local | Subset | Distribution-based | Linear |

[a] The generators are represented indirectly via scalar conserved quantities in Hamiltonian dynamics.
[b] The method discovers one symmetry transformation at a time, without learning the entire group structure.

**Algorithmic paradigm.** Generator-based methods learn transformation directions, usually in the form of matrix generators (Moskalev et al., 2022; Hu et al., 2025b) or, more generally, vector fields (Ko et al., 2024; Hu et al., 2025a). Once such a direction is learned, the method implicitly assumes that the symmetry relation remains valid along the corresponding one-parameter subgroup. Distribution-based methods (Allingham et al., 2024; Chen et al., 2025) instead learn a probability distribution over non-infinitesimal group elements, whose support approximates the valid transformation set. The former represents subgroup symmetries through infinitesimal generators, while the latter can represent generic subsets.

**Action representation.** The same abstract group may act linearly or nonlinearly on the data. Matrix generators represent linear actions; nonlinear actions may be represented by transformations in a learned latent space (Yang et al., 2024), nonlinear basis functions (Shaw et al., 2024; Hu et al., 2025a), or neural vector fields (Ko et al., 2024). This distinction is orthogonal to support structure: a nonlinear action can still define a global subgroup symmetry.

Table 1 summarizes representative methods according to these dimensions. The full taxonomy is provided in Appendix E. Our analysis focuses on generator-based discovery of global subgroup symmetries, using a vector-field formulation that includes matrix generators as a special case.

## 3 Stability of Generator-Subspace Recovery

Our goal is to study when generator-based discovery of subgroup symmetries is stable. Informally, the symmetry loss should vanish on the true generator subspace and grow away from it. In practice, however, this loss is evaluated through approximations: learned predictors for functional symmetry, discriminators or score models for distributional symmetry, and numerical solutions or estimated coefficients for PDE symmetry. We therefore ask: how do errors in these approximations affect the recovered generator subspace?

We study this question through a *symmetry-measuring operator*, which maps an infinitesimal transformation direction to its symmetry violation. Exact symmetry generators are precisely the elements in its kernel. This viewpoint is closely related to the fundamental operator of Otto et al. (2025), where the Lie derivative defines a linear map whose nullspace recovers the Lie algebra of functional

symmetries. Our formulation adopts this nullspace perspective across the target objects in Section 2: each symmetry notion induces an operator, after which the same subspace stability analysis applies.

### 3.1 Symmetry-Measuring Operators

Let $\mathcal{X}$ be the data domain with probability measure $q$, and let $\mathcal{H} = L^2(q; \mathbb{R}^n)$ denote the Hilbert space in which infinitesimal generators are represented as vector fields $v : \mathcal{X} \to T\mathcal{X}$. Here, $T\mathcal{X}$ denotes the tangent bundle, and the inner product and induced norm are

$$\langle v, v' \rangle_{\mathcal{H}} = \mathbb{E}_{x\sim q}[v(x)^T v'(x)], \quad \|v\|^2_{\mathcal{H}} = \mathbb{E}_{x\sim q}[\|v(x)\|_2^2]. \tag{1}$$

This vector-field formulation includes linear actions represented by matrix generators. A target symmetry notion induces a linear operator

$$\mathcal{A} : \mathcal{H} \to \mathcal{H}', \tag{2}$$

where $\mathcal{H}'$ is a Hilbert space of infinitesimal symmetry residuals, equipped with an inner product analogous to (1).

The defining property of $\mathcal{A}$ is that $v \in \ker(\mathcal{A})$ if and only if $v$ is an infinitesimal generator of the target symmetry. Generator-based symmetry discovery thus seeks to recover the true symmetry subspace $\mathcal{V}_{\text{true}} = \ker(\mathcal{A})$, which we assume has finite dimension $k$.

Discovery methods learn a $k$-tuple[1] of vector fields $V = (v_1, \ldots, v_k) \in \mathcal{H}^k$. Since $\mathcal{A}$ is often inaccessible, they construct a *surrogate operator* $\widehat{\mathcal{A}}$ from data and penalize the residuals $\|\widehat{\mathcal{A}} v_i\|^2_{\mathcal{H}'}$. To encourage orthonormality, we introduce the Gram matrix $G_V = \langle V, V \rangle_{\mathcal{H}}$, with entries $[G_V]_{ij} = \mathbb{E}_{x\sim q}[v_i(x)^T v_j(x)]$. With regularization strength $\tau > 0$, the learning objective is

$$\mathcal{L}(V) = \sum_{i=1}^{k} \|\widehat{\mathcal{A}} v_i\|^2_{\mathcal{H}'} + \tau \|G_V - I_k\|^2_{\mathrm{F}}. \tag{3}$$

We next instantiate this framework for functional, distributional, and PDE symmetries.

**Functional Symmetry.** For invariance $f(gx) = f(x)$, linearizing the residual along $g_\epsilon = \exp(\epsilon v)$ gives the operator

$$(\mathcal{A}v)(x) = \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \big[f(g_\epsilon x) - f(x)\big] = \nabla f(x)^T v(x), \tag{4}$$

with symmetry error $\mathcal{E}(v) = \|\mathcal{A}v\|^2_{\mathcal{H}'}$. For equivariance, with infinitesimal input and output actions $v|_{\mathcal{X}}$ and $v|_{\mathcal{Y}}$, the operator becomes $(\mathcal{A}v)(x) = \nabla f(x)^T v|_{\mathcal{X}}(x) - v|_{\mathcal{Y}}(f(x))$. Practical methods train a surrogate predictor $\hat{f}$ and construct $\widehat{\mathcal{A}}$ by replacing $f$ with $\hat{f}$ in these expressions.

**Distributional Symmetry.** For a smooth positive density $q$, linearizing the invariance relation $g_{\#}q = q$ gives the infinitesimal condition $\nabla \cdot (qv) = 0$. Normalizing by $q$ yields

$$\mathcal{A}v = \frac{\nabla \cdot (qv)}{q} = \nabla \log q \cdot v + \nabla \cdot v. \tag{5}$$

Appendix B.1 gives the density expansion. Among existing methods, LieGAN (Yang et al., 2023) learns transformations that preserve the joint input-output distribution. Its connection to this framework uses a discriminator $\hat{D}_v$ that distinguishes original samples from samples transformed along $v$. The required adaptations and surrogate-operator derivation are given in Appendix B.2.

**PDE Symmetry.** A symmetry of a PDE $\mathcal{F}[u] = 0$ maps solutions to solutions. Linearizing the transformed residual gives

$$\mathcal{A}v = \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \mathcal{F}[g_\epsilon u] = \nabla_{\text{jet}} \mathcal{F} \cdot \operatorname{pr} v, \tag{6}$$

[1] In practice, the number of generators is a hyperparameter. For the subsequent analysis, we fix it to the dimension of the ground-truth symmetry.

evaluated on the solution manifold. Here, the jet coordinates include the independent variables, dependent variables, and derivatives appearing in $\mathcal{F}$; the prolongation $\mathrm{pr}\, v$ describes the action of $v$ on these coordinates (Olver, 1993). Exact generators make this residual vanish on all solutions. The same criterion is also used to constrain differential equations given known symmetries (Akhound-Sadegh et al., 2023; Yang et al., 2026). A surrogate $\widehat{\mathcal{A}}$ may use numerical solution data or estimated PDE coefficients, with operator error arising from discretization, measurement, or coefficient estimation.

Table 2 summarizes these operators and representative methods. The ground-truth operator $\mathcal{A}$ is determined by the target symmetry, whereas the surrogate $\widehat{\mathcal{A}}$ depends on how a method estimates and evaluates the residual. Different methods may therefore address the same symmetry problem through different surrogate operators.

Table 2: Target symmetries, methods, and their definitions of the symmetry-measuring operator.

| Target symmetry | Operator $\mathcal{A} : v \mapsto$ | Method | Operator $\widehat{\mathcal{A}} : v \mapsto$ |
|---|---|---|---|
| $f(x) = f(gx)$ | $\nabla f \cdot v$ | LieGG (Moskalev et al., 2022) | $\nabla \hat{f} \cdot v$ |
| $gf(x) = f(gx)$ | $\nabla f \cdot v\|_{\mathcal{X}} - v\|_{\mathcal{Y}} \circ f$ | Hu et al. (2025b) | $\nabla \hat{f} \cdot v\|_{\mathcal{X}} - v\|_{\mathcal{Y}} \circ \hat{f}$ |
| $q = g_{\#}q$ | $\nabla \log q \cdot v + \nabla \cdot v$ | LieGAN (Yang et al., 2023) | $\nabla \log(1 - \widehat{D}_v) \cdot v$ |
| $\mathcal{F}[u] = 0 \Rightarrow \mathcal{F}[gu] = 0$ | $\nabla_{\mathrm{jet}} \mathcal{F} \cdot \mathrm{pr}\, v$ | Ko et al. (2024) | $\nabla_{\mathrm{jet}} \mathcal{F}[u] \cdot \mathrm{pr}\, v[u]$ |

### 3.2 Bounding the Subspace Distance

We have reformulated generator-based symmetry discovery as the problem of recovering the kernel of a task-specific symmetry-measuring operator. When the learning objective (3) uses the surrogate $\widehat{\mathcal{A}}$ in place of the ground-truth operator $\mathcal{A}$, how close is the subspace spanned by the learned generators to the true symmetry subspace?

The following theorem answers this question in terms of two quantities. The operator perturbation strength $\|\widehat{\mathcal{A}} - \mathcal{A}\|_{\mathrm{op}}$ measures the discrepancy between surrogate and ground-truth symmetry residuals. The spectral gap $\lambda$ measures how strongly the true operator separates symmetry directions from non-symmetry directions. Intuitively, if symmetry violation is uniformly bounded away from zero over unit directions orthogonal to $\mathcal{V}_{\mathrm{true}}$, then the kernel of $\mathcal{A}$ is isolated and can be stably recovered under sufficiently small perturbations.

**Theorem 3.1.** *Let $\mathcal{H} = L^2(q; \mathbb{R}^n)$ be the Hilbert space of vector fields. Let $\mathcal{A}, \widehat{\mathcal{A}} : \mathcal{H} \to \mathcal{H}'$ be the ground-truth and surrogate symmetry-measuring operators defined in Section 3.1. Let $\mathcal{V}_{\mathrm{true}} = \ker(\mathcal{A})$ be the true $k$-dimensional subspace of infinitesimal symmetry generators, and $\widehat{\mathcal{V}} \subset \mathcal{H}$ be the $k$-dimensional subspace spanned by the global optimum of the surrogate objective $\mathcal{L}(V)$ (3). Let $P$ and $\widehat{P}$ be the orthogonal projectors onto $\mathcal{V}_{\mathrm{true}}$ and $\widehat{\mathcal{V}}$ in $\mathcal{H}$. Assume:*

- *(Surrogate operator error) $\|\widehat{\mathcal{A}} - \mathcal{A}\|_{\mathrm{op}} \leq \delta$.*
- *(Isolated spectral gap) For any vector field $v \in \mathcal{H}$ orthogonal to $\mathcal{V}_{true}$, the symmetry residual satisfies: $\|\mathcal{A}v\|_{\mathcal{H}'} \geq \lambda \|v\|_{\mathcal{H}}$ for some $\lambda > 0$.*

*If $\tau > \frac{\delta^2}{2}$ and $\delta < \frac{\lambda}{2}$, the distance between the true and discovered subspaces is bounded by*

$$\|\widehat{P} - P\|_{\mathrm{HS}} \leq \sqrt{k\left(1 - \sqrt{1 - 4(\delta/\lambda)^2}\right)} \approx \sqrt{2k}\frac{\delta}{\lambda} + \mathcal{O}\left(\frac{\delta^3}{\lambda^3}\right). \tag{7}$$

The proof is given in Appendix A.1. We measure subspace error using the Hilbert-Schmidt norm of the difference between the orthogonal projectors. Direct comparisons between individual generators depend on the choice of basis: an orthogonal change of basis represents the same symmetry subspace, and the objective (3) is invariant under orthogonal mixing of the learned generators. The projector distance $\|\widehat{P} - P\|_{\mathrm{HS}}$ depends only on the subspaces and is therefore unaffected by this ambiguity.

The theorem shows that the recovered generator subspace is stable when the surrogate operator is accurate relative to the intrinsic spectral gap of the symmetry problem. In the small-perturbation regime $\delta/\lambda \to 0$, the error bound scales linearly with $\delta/\lambda$. Thus, recovery depends on the surrogate error relative to the separation between symmetry and non-symmetry directions, rather than on the absolute surrogate error alone.

The conditions $\tau > \delta^2/2$ and $\delta < \lambda/2$ also have practical implications. The first is sufficient to keep a global minimizer of the objective (3) full rank, ensuring that its generators span the bottom-$k$ right singular subspace of $\widehat{\mathcal{A}}$. Once this threshold is exceeded, $\tau$ does not enter the subspace error bound, although it still affects the generator amplitudes and optimization. In particular, a finite Gram-matrix penalty need not produce exactly orthonormal generators. In practice, $\tau$ should be large enough to discourage collapse without making optimization excessively dominated by this penalty. The second condition, $\delta < \lambda/2$, specifies the operator accuracy required by the guarantee: the perturbation must remain below half of the ground-truth spectral gap.

The theorem provides a concrete criterion for surrogate quality by linking symmetry recovery to the accuracy of the induced symmetry-measuring operator. For scalar functional invariance in the $L^2$ setup, $\|\widehat{\mathcal{A}} - \mathcal{A}\|_{\mathrm{op}} = \operatorname{ess\,sup}_{x\sim q} \|\nabla \hat{f}(x) - \nabla f(x)\|$. This relation identifies derivative accuracy as the aspect of surrogate approximation that enters the recovery guarantee: accurately capturing the target function's local variation yields a smaller operator perturbation and a tighter subspace error bound. More generally, the operator perspective clarifies which properties of a surrogate should be preserved to support reliable symmetry discovery. The spectral gap $\lambda$ also characterizes the difficulty of the recovery problem for a given task and generator space. Enlarging $\mathcal{H}$ may introduce additional unit directions with small but nonzero symmetry residuals, potentially reducing $\lambda$ and making recovery more sensitive to surrogate error.

## 4 Experiments: Subspace Stability in Synthetic Environments

We study functional, distributional, and PDE symmetry discovery in controlled environments with known generator subspaces. By perturbing their symmetry-measuring operators, we can directly control the surrogate error and evaluate the recovery bound in Theorem 3.1. Section 4.1 introduces the three synthetic tasks, and Section 4.2 describes the shared perturbation protocol, recovery methods, and evaluation metric. Section 4.3 presents the results, comparing iterative generator learning with direct recovery by singular value decomposition (SVD), examining smooth error growth and abrupt spectral transitions, and evaluating the role of regularization in preventing generator collapse.

### 4.1 Synthetic Tasks

**Functional symmetry.** We use $f(X) = -\operatorname{tr}(W^T X)$, where $X \in \mathbb{R}^{n\times n}$ and $W = [0_{m\times n}; B]$, with $B \in \mathbb{R}^{(n-m)\times n}$ of full row rank. Under the left $\mathrm{SO}(n)$ action, the symmetry-measuring operator has the matrix representation $\mathcal{A}_f : L \mapsto LW$. Its kernel is $\mathfrak{so}(m)$ embedded in $\mathfrak{so}(n)$: because $W$ ignores the first $m$ rows, $f$ is invariant under rotations of those rows. Specifically, for $R \in \mathrm{SO}(m)$ and $\widetilde{R} = \operatorname{diag}(R, I_{n-m})$, we have $f(\widetilde{R}X) = f(X)$.

**Distributional symmetry.** We use the isotropic Gaussian distribution $q = \mathcal{N}(0, I_m)$ and linear generators $v_B(x) = Bx$, with $B \in \mathbb{R}^{m\times m}$. The symmetry-measuring operator is

$$(\mathcal{A}_q B)(x) = \operatorname{tr} B - x^T B x.$$

Its kernel is $\mathfrak{so}(m)$, corresponding to rotational invariance of the Gaussian. Under the Frobenius metric on generators and the $L^2(q)$ metric on residuals, all positive singular values equal $\sqrt{2}$.

**PDE symmetry.** We use $\mathcal{F}_D[u] = \operatorname{tr}(D\nabla^2 u) + \kappa u = 0$, where $D = \operatorname{diag}(aI_m, bI_{n-m})$, $a, b, \kappa > 0$, and $a \neq b$. For rotational generators $L \in \mathfrak{so}(n)$, the prolonged residual is

$$(\mathcal{A}_D L)[u](x) = \operatorname{tr}([D, L] H_u(x)), \qquad [D, L] = DL - LD,$$

where $H_u(x) = \nabla_x^2 u(x)$. The kernel is $\mathfrak{so}(m) \oplus \mathfrak{so}(n-m)$, representing independent rotations within the two eigenspaces of $D$.

Each sandbox admits a finite-dimensional operator representation with a known kernel. We set $k = \dim\ker \mathcal{A}$ and compute the spectral gap $\lambda$ as the smallest positive singular value of $\mathcal{A}$.

### 4.2 Experimental Protocol

For each fixed ground-truth problem, we vary the perturbation strength $\delta$ while keeping $\lambda$ fixed. We normalize each perturbation so that $\|\widehat{\mathcal{A}} - \mathcal{A}\|_{\mathrm{op}} = \delta$ and report recovery error against $\delta/\lambda$. We compare five perturbation modes, including random coefficient perturbations and four different types of adversarial perturbations.

**Random perturbations.** We add Gaussian noise to the underlying coefficient matrix: $W$ for the functional sandbox, the matrix specifying the Gaussian density for the distributional sandbox, and $D$ for the PDE sandbox. The distributional and PDE perturbations use symmetric noise matrices. We scale the coefficient perturbation according to the norm of the induced operator change, so that all three settings use the same operator-error parameter $\delta$.

**Adversarial perturbations.** We also construct adversarial perturbations to test whether the error bound in Theorem 3.1 can be realized in the worst case. We refer to different types of such perturbations by their effects on the singular structure of the operator $\mathcal{A}$, namely the rotation of singular directions and the scaling of singular values. The direct *rotational* adversary perturbs the operator to rotate its bottom singular subspace, testing the sensitivity of recovered symmetry directions. The direct *scaling* adversary instead changes singular values while preserving singular directions. By increasing residuals along symmetry directions and decreasing them along non-symmetry directions, it can exchange their ordering once $\delta$ reaches the half-gap threshold. For each mechanism, we also consider a coefficient-realizable, or "physical," adversarial perturbation obtained by perturbing the underlying function, density, or PDE coefficients. These constructions connect the abstract operator analysis to perturbations of the target objects themselves. The direct adversaries probe worst-case operator behavior, while the coefficient-realizable adversaries test the same mechanisms under the structural constraints of each sandbox. Construction details are provided in Appendix C.

**Recovery methods and evaluation.** Across all three settings, we recover $k$ generator directions from $\widehat{\mathcal{A}}$ using two approaches. First, we iteratively optimize the regularized objective (3). Second, we compute the $k$ right singular vectors associated with the smallest singular values of $\widehat{\mathcal{A}}$. Under the conditions of Theorem 3.1, this SVD subspace is the span of a global minimizer of the regularized objective. SVD therefore provides a direct spectral reference for evaluating iterative recovery.

For the subspace-error comparisons in Figure 1, we use the relative regularization grid $\tau = q_\tau \delta^2$, with $q_\tau \in \{0.5, 1, 2\}$. The SVD reference is independent of $\tau$. Each environment uses one fixed problem configuration, and reported errors are averaged over two random seeds. We measure recovery using the projector distance $\|\widehat{P} - P\|_{\mathrm{HS}}$ and compare it with the bound (7) evaluated using the corresponding $k$ and $\lambda$.

### 4.3 Results

**Agreement between iterative optimization and SVD.** The theorem characterizes the recovered span at a global optimum, but does not guarantee that iterative optimization reaches it. We therefore compare iterative recovery with SVD to assess whether optimization introduces additional recovery error relative to the direct spectral reference. Across the three environments in Figure 1, the two approaches yield nearly identical subspace errors. Within the theorem's regime $\delta < \lambda/2$ and $\tau > \delta^2/2$, this agreement is consistent with its characterization of the globally optimal span as the bottom-$k$ singular subspace of $\widehat{\mathcal{A}}$. Although matching errors do not establish global optimality, they show that similar recovery errors arise even without iterative optimization. This motivates the following spectral interpretation of smooth error growth and abrupt transitions.

**Smooth error growth under rotational and random perturbations.** Across the three environments in Figure 1, rotational and random perturbations produce smooth increases in subspace error as $\delta/\lambda$ grows. The (adversarial) rotational perturbation creates this effect by mixing true symmetry directions with non-symmetry directions having small positive singular values, gradually tilting the recovered subspace. In the small-perturbation regime, the observed growth is consistent with the approximately linear dependence on $\delta/\lambda$ in the theoretical bound. The physical rotational and

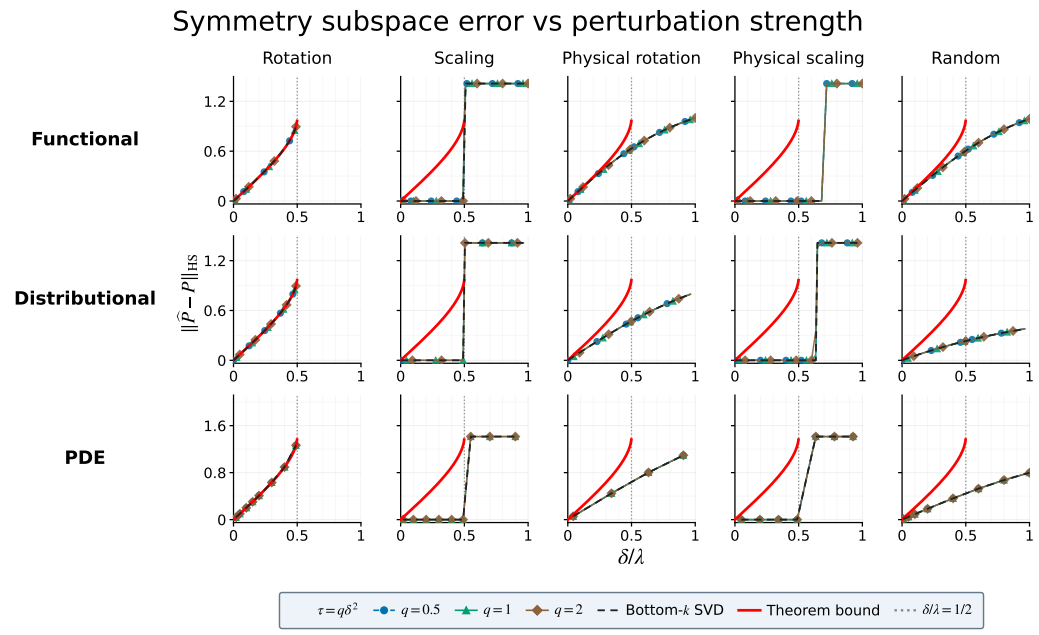


Figure 1: Symmetry subspace error $\|\widehat{P} - P\|_{\mathrm{HS}}$ versus normalized perturbation strength $\delta/\lambda$ across functional, distributional, and PDE symmetry discovery (rows), under five perturbation modes (columns). Each environment uses one fixed problem configuration, with empirical errors averaged over two random seeds. Colored curves with markers correspond to iterative optimization with $\tau = q_\tau \delta^2$ for $q_\tau \in \{0.5, 1, 2\}$; black dashed curves show the bottom-$k$ SVD reference. The red solid curve is the theoretical error bound from Theorem 3.1, shown for $\delta/\lambda < 1/2$, and the vertical dotted line marks this threshold. Iterative optimization and SVD yield nearly coincident errors across the three environments. Rotational and random perturbations produce smooth error growth, whereas scaling perturbations exhibit abrupt transitions from near-zero error to substantial subspace mismatch.

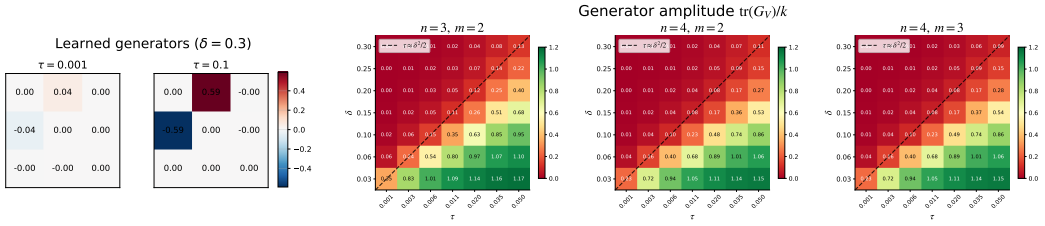


Figure 2: Left: Two instances of learned generators with $\delta = 0.3$, $m = 2$, $n = 3$, and $\tau = 0.001$ (below the $\delta^2/2$ threshold) and $\tau = 0.1$ (above the threshold), respectively. Right: Heatmap of $\mathrm{tr}(G_V)/k$, representing the amplitude of the learned generators, over a $(\delta, \tau)$-grid around $\tau = \delta^2/2$.

random perturbations exhibit the same qualitative trend, showing that gradual recovery degradation also arises from perturbing the underlying target objects. Throughout these comparisons, the iterative curves remain close to the SVD reference. Thus, the growth in error is also present in direct spectral recovery, consistent with the theorem's description of sensitivity to operator perturbations.

**Spectral transitions under scaling perturbations.** Scaling perturbations instead produce a near-zero error plateau followed by an abrupt jump. In the direct scaling case, this transition occurs at the vertical line $\delta/\lambda = 1/2$. The construction raises the residuals of symmetry directions while lowering those of competing non-symmetry directions; at the half-gap threshold, their singular values cross and the recovered bottom-$k$ subspace changes. The resulting jump therefore reflects a change in which directions are selected, rather than a gradual rotation of the same subspace. The coefficient-realizable

scaling perturbations also exhibit abrupt transitions, with their locations depending on the constraints imposed by the underlying target object. These transitions appear in both iterative optimization and SVD across the three environments, supporting the spectral explanation of recovery failure and illustrating the role of the theorem's condition $\delta < \lambda/2$.

**Generator amplitudes and regularization.** The preceding results examine subspace recovery and the role of the condition $\delta < \lambda/2$. We now turn to the complementary condition $\tau > \delta^2/2$, which ensures sufficient regularization to prevent generator collapse. We examine this through generator amplitudes in the functional symmetry sandbox. Figure 2 (left) compares learned generators under regularization strengths below and above this threshold. Under small $\tau$, the learned generators have near-zero amplitudes, whereas stronger regularization prevents this collapse.

To examine this behavior across perturbation strengths, we visualize the normalized trace of the Gram matrix, $\mathrm{tr}(G_V)/k$, in Figure 2 (right) as a metric for generator amplitudes. The ideal value of this metric is $1.0$, which indicates that the learned generators have unit norm on average. To construct the $(\delta, \tau)$-grid, we choose $\delta \in \{0.03, 0.06, 0.10, 0.15, 0.20, 0.25, 0.30\}$ and a matching set of $\tau$-values so that one diagonal corresponds to the theoretical threshold $\tau = \delta^2/2$. Above the diagonal, where $\tau < \delta^2/2$, the average amplitude is almost zero. The transition to visibly positive amplitudes emerges immediately below the diagonal. This behavior supports the role of the theorem's sufficient regularization condition in preventing generator collapse, complementing the subspace-error analysis above.

## 5 Conclusion

We developed a stability theory for generator-based continuous symmetry learning. By formulating symmetry discovery as the recovery of the kernel of a symmetry-measuring operator, we provided a common framework for functional, distributional, and PDE symmetries. Our subspace-distance bound quantifies how errors in a surrogate operator affect the recovered generator subspace, with recovery accuracy governed by the perturbation strength relative to the spectral gap. Controlled experiments across the three settings support the predicted stability behavior, including smooth error growth and abrupt spectral transitions, with close agreement between iterative optimization and SVD. A separate amplitude analysis illustrates the role of regularization in preventing generator collapse.

An immediate direction is to complement this stability result with optimization guarantees. The current analysis characterizes the subspace recovered at a global optimum. Studying the optimization landscape and convergence of gradient-based training would clarify when this subspace can be reached and how a finite computational budget affects recovery accuracy. Such results would connect the operator perturbation analysis to the behavior of concrete learning procedures.

Another priority is to connect surrogate learning from data to quantitative control of operator error. For functional symmetry, this requires accounting for derivative fidelity alongside prediction accuracy. Statistically justified operator-uncertainty bounds, developed under explicit regularity or modeling assumptions, could relate observable prediction behavior and data coverage to the quantities entering the stability theorem. Combined with spectral information from the surrogate operator, these bounds could assess recovery accuracy without access to the ground-truth function or symmetry.

The operator perspective also suggests extensions to other target objects and alternative surrogate constructions. For a new target object, the first step is to derive the infinitesimal residual that defines its symmetry-measuring operator. Even for a fixed target, different methods may approximate this operator differently. For example, directly estimating $\nabla \log q$ with a score model offers an alternative construction for distributional symmetry, motivating comparisons of the operator accuracy achieved by different surrogate models.

The supporting taxonomy situates these results within a broader landscape of symmetry discovery problems. In particular, input-dependent subgroup symmetries motivate extending recovery guarantees from a fixed subspace to a field of symmetry subspaces. More general subset-valued symmetries, where valid transformations may occupy bounded or disconnected regions, call for a complementary theory of methods that learn distributions over group elements. These directions extend the present analysis beyond global subgroup recovery while preserving its central aim: understanding how approximation errors affect the symmetries inferred from data.

## AI USE STATEMENT

In this work, we used generative AI tools for generating synthetic datasets, implementing methods, providing feedback on research methodology and experiments, supporting qualitative data analysis (writing code to plot figures), and interpreting results. We have not used generative AI tools for developing theoretical models or conceptual frameworks, formulating or providing critical ingredients for proving mathematical claims, assisting in the writing of proofs, proposing or refining hypotheses, and assisting with translation, and cleaning and reformatting datasets are not applicable to this work. Additionally, we used generative AI tools for creating figures, brainstorming, and writing and editing the paper in general. We have reviewed all AI-assisted work. For example, we proofread all LLM-generated text to ensure there is no hallucination or any other inaccurate or misleading statement about the methodology and experiment results. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

## REPRODUCIBILITY STATEMENT

We have provided clear explanations of assumptions and a complete proof in Appendix A.1 for the main theoretical claim in this paper. In Appendix C, we have also included the complete description of how to create the experimental environments and how to construct the perturbations required to validate the theorem.

# REFERENCES


Tara Akhound-Sadegh, Laurence Perreault-Levasseur, Johannes Brandstetter, Max Welling, and Siamak Ravanbakhsh. Lie point symmetry and physics-informed networks. *Advances in Neural Information Processing Systems*, 36:42468–42481, 2023.

James U Allingham, Bruno K Mlodozeniec, Shreyas Padhy, Javier Antorán, David Krueger, Richard E Turner, Eric Nalisnick, and José M Hernández-Lobato. A generative model of symmetry transformations. *Advances in Neural Information Processing Systems*, 37:91091–91130, 2024.

Gregory Benton, Marc Finzi, Pavel Izmailov, and Andrew G Wilson. Learning invariances in neural networks from training data. *Advances in neural information processing systems*, 33:17605–17616, 2020.

Manu Bhat, Jonghyun Park, Jianke Yang, Nima Dehmamy, Robin Walters, and Rose Yu. Atlasd: Automatic local symmetry discovery. In *International Conference on Machine Learning*, pp. 4153–4171. PMLR, 2025.

Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

Yuxuan Chen, Jung Yeon Park, Floor Eijkelboom, Jianke Yang, Jan-Willem van de Meent, Lawson LS Wong, and Robin Walters. Discovering symmetry groups with flow matching. *arXiv e-prints*, pp. arXiv–2512, 2025.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pp. 2990–2999. PMLR, 2016.

Krish Desai, Benjamin Nachman, and Jesse Thaler. Symmetry discovery with deep learning. *Physical Review D*, 105(9):096031, 2022.

Marc Finzi, Max Welling, and Andrew Gordon Wilson. A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups. In *International conference on machine learning*, pp. 3318–3328. PMLR, 2021.

Wanda Hou, Molan Li, and Yi-Zhuang You. Machine learning symmetry discovery for integrable hamiltonian dynamics. *Physical Review E*, 113(2):025302, 2026.

Lexiang Hu, Yikang Li, and Zhouchen Lin. Explicit discovery of nonlinear symmetries from dynamic data. *arXiv preprint arXiv:2510.01855*, 2025a.

Lexiang Hu, Yikang Li, and Zhouchen Lin. Symmetry discovery for different data types. *Neural Networks*, 188:107481, 2025b.

Pavan Karjol, Kumar Shubham, and Prathosh AP. Spectral discovery of continuous symmetries via generalized fourier transforms. *arXiv preprint arXiv:2603.07299*, 2026.

Gyeonghoon Ko, Hyunsu Kim, and Juho Lee. Learning infinitesimal generators of continuous symmetries from data. *Advances in Neural Information Processing Systems*, 37:85973–86003, 2024.

Shida Liu, Sumit Sinha, Vishaal Krishnan, and L. Mahadevan. Liedynnet: Learning lie symmetries from spatiotemporal data, 2025. URL https://openreview.net/forum?id=VgZ8BJkneV.

Thomas Mitchel, Michael J Taylor, and Vincent Sitzmann. Neural isometries: Taming transformations for equivariant ml. *Advances in Neural Information Processing Systems*, 37:7311–7338, 2024.

Artem Moskalev, Anna Sepliarskaia, Ivan Sosnovik, and Arnold Smeulders. Liegg: Studying learned lie group generators. *Advances in Neural Information Processing Systems*, 35:25212–25223, 2022.

Peter J Olver. *Applications of Lie groups to differential equations*, volume 107. Springer Science & Business Media, 1993.

Samuel E. Otto, Nicholas Zolman, J. Nathan Kutz, and Steven L. Brunton. A unified framework to enforce, discover, and promote symmetry in machine learning. *Journal of Machine Learning Research*, 26(248):1–83, 2025. URL http://jmlr.org/papers/v26/24-1315.html.

Ben Shaw, Abram Magner, and Kevin Moon. Symmetry discovery beyond affine transformations. *Advances in Neural Information Processing Systems*, 37:112889–112918, 2024.

Alonso Urbano, David W Romero, Max Zimmer, and Sebastian Pokutta. Recon: Robust symmetry discovery via explicit canonical orientation normalization. *arXiv preprint arXiv:2505.13289*, 2025.

Rui Wang, Robin Walters, and Rose Yu. Incorporating symmetry into deep dynamics models for improved generalization. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=wta_8Hx2KD.

Rui Wang, Robin Walters, and Rose Yu. Approximately equivariant networks for imperfectly symmetric dynamics. In *International Conference on Machine Learning*, pp. 23078–23091. PMLR, 2022.

Maurice Weiler and Gabriele Cesa. General e (2)-equivariant steerable cnns. *Advances in neural information processing systems*, 32, 2019.

Jianke Yang, Robin Walters, Nima Dehmamy, and Rose Yu. Generative adversarial symmetry discovery. In *International Conference on Machine Learning*, pp. 39488–39508. PMLR, 2023.

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. PMLR, 2024. URL https://proceedings.mlr.press/v235/yang24g.html.

Jianke Yang, Manu Bhat, Bryan Hu, Yadi Cao, Nima Dehmamy, Robin Walters, and Rose Yu. Discovering symbolic differential equations with symmetry invariants. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. URL https://openreview.net/forum?id=9t1dEyYfPc. Featured Certification, J2C Certification.

# A PROOFS

## A.1 SYMMETRY SUBSPACE ERROR BOUND

**Theorem A.1.** *Let $\mathcal{H} = L^2(q; \mathbb{R}^n)$ be the Hilbert space of vector fields. Let $\mathcal{A}, \hat{\mathcal{A}} : \mathcal{H} \to \mathcal{H}'$ be the ground-truth and surrogate symmetry-measuring operators defined in Section 3.1. Let $\mathcal{V}_{\text{true}} = \ker(\mathcal{A})$ be the true $k$-dimensional subspace of infinitesimal symmetry generators, and $\hat{\mathcal{V}} \subset \mathcal{H}$ be the $k$-dimensional subspace spanned by the global optimum of the surrogate objective $\mathcal{L}(V)$ (3). Let $P$ and $\hat{P}$ be the orthogonal projectors onto $\mathcal{V}_{\text{true}}$ and $\hat{\mathcal{V}}$ in $\mathcal{H}$. Assume:*

- *(Surrogate operator error) $\|\hat{\mathcal{A}} - \mathcal{A}\|_{\text{op}} \leq \delta$.*
- *(Isolated Spectral Gap) For any vector field $v \in \mathcal{H}$ strictly orthogonal to $\mathcal{V}_{true}$, the symmetry error grows strictly: $\|\mathcal{A}v\|_{\mathcal{H}'} \geq \lambda \|v\|_{\mathcal{H}}$ for some $\lambda > 0$.*

*If $\tau > \frac{\delta^2}{2}$ and $\delta < \frac{\lambda}{2}$, the distance between the true and discovered subspaces is bounded by*

$$\|\hat{P} - P\|_{\text{HS}} \leq \sqrt{k\left(1 - \sqrt{1 - 4(\delta/\lambda)^2}\right)} \approx \sqrt{2k}\frac{\delta}{\lambda} + \mathcal{O}\left(\frac{\delta^3}{\lambda^3}\right). \tag{8}$$

*Proof.* We first determine the subspace spanned by the optimum $V \in \mathcal{H}^k$. We take the Fréchet derivative of $\mathcal{L}(V)$ with respect to the vector fields. Perturbing $V$ by a test function tuple $\eta \in \mathcal{H}^k$:

$$\delta\mathcal{L}(V; \eta) = 2\text{tr}(\langle \eta, \hat{\mathcal{A}}^*\hat{\mathcal{A}}V\rangle_{\mathcal{H}}) + 4\tau\text{tr}\left(\langle \eta, V\rangle_{\mathcal{H}}(G_V - I_k)\right) \tag{9}$$

Denote $\hat{\mathcal{M}} = \hat{\mathcal{A}}^*\hat{\mathcal{A}}$. Setting the variation to zero for all $\eta$ requires that the following holds a.e. in $\mathcal{X}$:

$$(\hat{\mathcal{M}}V)(x) = 2\tau V(x)(I_k - G_V) \tag{10}$$

Next, we seek to construct a global minimum solution in $\mathcal{H}^k$ that satisfies the optimality condition and results in a full-rank $G_V$. We will find that such a minimum exists and is uniquely the $k$-dimensional right singular subspace corresponding to the lowest singular values of $\hat{\mathcal{A}}$, if $\tau > \frac{\delta^2}{2}$ and $\delta < \lambda/2$.

Let $G_V = R\Sigma^2 R^T$ be the eigendecomposition of the constant $k \times k$ Gram matrix. Assuming $G_V$ has full rank under the optimal solution for $V$, we construct an orthonormal basis for the discovered subspace by setting $U(x) = V(x)R\Sigma^{-1}$, so $\langle U, U\rangle_{\mathcal{H}} = I_k$. Substituting $V(x) = U(x)\Sigma R^T$ into the optimality condition yields:

$$\hat{\mathcal{M}}U\Sigma R^T = 2\tau U\Sigma R^T(I_k - R\Sigma^2 R^T) \implies \hat{\mathcal{M}}U = U \cdot 2\tau(I_k - \Sigma^2) \tag{11}$$

Because $2\tau(I_k - \Sigma^2)$ is a diagonal matrix of constants, the columns of $U(x)$ are exactly the $k$ eigenfunctions of the surrogate operator $\hat{\mathcal{M}}$ corresponding to its eigenvalues $\hat{\Lambda} = 2\tau(I_k - \Sigma^2)$. Rearranging $\hat{\Lambda} = \text{diag}(\hat{\lambda}_i^2)$ and $\Sigma = \text{diag}(\sigma_i)$ gives $\sigma_i^2 = 1 - \frac{\hat{\lambda}_i^2}{2\tau}$, where $\hat{\lambda}_i$ are the singular values of $\hat{\mathcal{A}}$ in ascending order. To proceed, we need to verify the assumption that the optimal $V$ yields full-rank $G_V$, i.e. $\sigma_i^2 > 0$.

Notice that by Assumption 2, the true operator $\mathcal{A}$ has the singular value 0 with multiplicity $k$ (corresponding to $\mathcal{V}_{true}$), and the rest of its singular values are contained in $[\lambda, \infty)$. Now we require that $\delta^2 < 2\tau$ and $\delta < \lambda/2$. By Weyl's Theorem for bounded operators, the singular values of $\hat{\mathcal{A}}$ are perturbed by at most $\delta$. Because $\delta < \lambda/2$, the $k$ lowest singular values of $\hat{\mathcal{M}}$ reside in $[0, \delta]$, i.e. $\hat{\lambda}_i \leq \delta$ for $i \leq k$, while the rest of the singular values of $\hat{\mathcal{A}}$ satisfies $\hat{\lambda}_i \geq \lambda - \delta > \delta$.

Since $\tau > \delta^2/2$, we have $\sigma_i^2 = 1 - \frac{\hat{\lambda}_i^2}{2\tau} > 0$ for any $i \leq k$, and the optimal subspace $\hat{\mathcal{V}}$ is a $k$-dimensional eigenspace of $\hat{\mathcal{M}}$, or equivalently, a $k$-dimensional right singular subspace of $\hat{\mathcal{A}}$. It can be further shown that the global minimum of $\mathcal{L}(V)$ is uniquely the singular subspace $\hat{\mathcal{V}}$ corresponding to $\hat{\lambda}_1, \ldots, \hat{\lambda}_k$, whereas other $\hat{\lambda}_j$ $(j > k)$ satisfying $\sigma_j^2 = 1 - \frac{\hat{\lambda}_j^2}{2\tau} > 0$ corresponds to saddle points of $\mathcal{L}(V)$.

Next, we calculate the subspace distance between $\mathcal{V}_{true}$ and $\hat{\mathcal{V}}$ in terms of their principal angles. Let the principal angles between $\mathcal{V}_{true}$ and $\hat{\mathcal{V}}$ be $0 \leq \theta_1, \ldots, \theta_k < \pi/2$. For each nonzero principal angle, we may choose orthonormal vectors $v_i \in \mathcal{V}_{true}$ and $w_i \in \mathcal{W} = \mathcal{V}_{true}^{\perp}$ such that the corresponding perturbed right singular direction has the form

$$\hat{v}_i = \cos\theta_i v_i + \sin\theta_i w_i. \tag{12}$$

There is also a companion vector in the orthogonal complement of $\hat{\mathcal{V}}$,

$$\hat{w}_i = -\sin\theta_i v_i + \cos\theta_i w_i. \tag{13}$$

Since $\mathcal{A}v_i = 0$, we have

$$\mathcal{A}\hat{v}_i = \sin\theta_i \mathcal{A}w_i, \quad \mathcal{A}\hat{w}_i = \cos\theta_i \mathcal{A}w_i. \tag{14}$$

Thus, before perturbation, the two vectors $\mathcal{A}\hat{v}_i$ and $\mathcal{A}\hat{w}_i$ are collinear in the direction $a_i = \mathcal{A}w_i$. We also note that $\|a_i\|_{\mathcal{H}'} \geq \lambda$ from Assumption 2. But for $\hat{\mathcal{A}}$, we have $\langle \hat{\mathcal{A}}\hat{v}_i, \hat{\mathcal{A}}\hat{w}_i \rangle = 0$. Given that $\|\hat{\mathcal{A}} - \mathcal{A}\|_{op} \leq \delta$, there is a maximum $s = \sin\theta_i$ that makes this orthogonality constraint achievable. However, instead of looking for the maximum $s$ achievable under fixed perturbation budget $\delta$, it will be more convenient to find the minimum perturbation needed for a fixed $s$.

Specifically, let $\gamma = \|a_i\|_{\mathcal{H}'}$ and $u = \frac{a_i}{\gamma}$. Write $s = \sin\theta_i$ and $c = \cos\theta_i$. Then the restriction of $\mathcal{A}$ to the two-dimensional domain basis $\{\hat{v}_i, \hat{w}_i\}$ looks like

$$\mathcal{A}_{\text{loc}} = \gamma u \begin{pmatrix} s & c \end{pmatrix}. \tag{15}$$

After choosing a range basis whose spanned subspace includes $u$,[2] we can express $\mathcal{A}_{\text{loc}}$ in a matrix $M$. At this point, our original problem is equivalent to finding the smallest operator-norm perturbation $F$ s.t. the perturbed operator's realization with the chosen bases has orthogonal columns, i.e.

$$\min_{\langle (M+F)e_1, (M+F)e_2 \rangle = 0} \|F\|_{op}. \tag{16}$$

Since the two columns of $N = M + F$ are orthogonal, we may choose an orthonormal basis of the range so that $N$ is diagonalized: $Ne_1 = ae_1'$, $Ne_2 = be_2'$, with $e_1' \perp e_2'$, $a, b \geq 0$. The coordinate of $u$ under this basis becomes some unit vector $h = (h_1, h_2, h_\perp)$ with $|h_1|^2 + |h_2|^2 + \|h_\perp\|^2 = 1$. Thus $Me_1 = \gamma s h,\ Me_2 = \gamma c h$, and the perturbation columns are

$$Fe_1 = ae_1' - \gamma s h, \quad Fe_2 = be_2' - \gamma c h. \tag{17}$$

Because $\|F\|_{op} \geq \|Fe_1\|$, we have $\|F\|_{op} \geq \|ae_1' - \gamma s h\| \geq \gamma s\sqrt{1 - |h_1|^2}$. Denoting $\rho = \|F\|_{op}$, we can obtain

$$|h_1|^2 \geq 1 - \frac{\rho^2}{\gamma^2 s^2}, \tag{18}$$

and similarly,

$$|h_2|^2 \geq 1 - \frac{\rho^2}{\gamma^2 c^2}. \tag{19}$$

Since $|h_1|^2 + |h_2|^2 \leq 1$, we can combine the inequalities to get

$$\frac{\rho^2}{\gamma^2}\left(\frac{1}{s^2} + \frac{1}{c^2}\right) \geq 1 \ \Rightarrow\ \rho \geq \gamma s c \tag{20}$$

We can actually construct the perturbation that achieves the $\gamma sc$ bound explicitly. Let $w$ be a unit vector orthogonal to $u$ in the range space, and define two orthonormal vectors $p = su - cw,\ q = cu + sw$, which satisfies $\langle p, q \rangle = 0$. Consider the following operator:

$$Ne_1 = \gamma s^2 p, \quad Ne_2 = \gamma c^2 q. \tag{21}$$

[2]That this action is possible implicitly assumes that $\mathcal{A}$ has a nontrivial left null subspace. If this is not true, we will end up with a strictly larger minimum perturbation for a fixed $s$, and reversely, a smaller achievable principal angle.

Or, in matrix form in the range basis $\{u, w\}$ and the domain basis $\hat{v}_i, \hat{w}_i$,

$$N = \gamma \begin{pmatrix} s^3 & c^3 \\ -s^2 c & c^2 s \end{pmatrix}. \tag{22}$$

The original matrix was

$$M = \gamma \begin{pmatrix} s & c \\ 0 & 0 \end{pmatrix}. \tag{23}$$

It can be easily verified that

$$F = N - M = \gamma sc \begin{pmatrix} -c & -s \\ -s & c \end{pmatrix} \tag{24}$$

and hence $\|F\|_{op} = \gamma sc$.

By Assumption 1, we know that the operator norm of the perturbation cannot exceed $\delta$. Plugging in the inequalities,

$$\delta \geq \gamma sc \geq \lambda sc = \lambda s\sqrt{1 - s^2}. \tag{25}$$

Solving on the relevant branch $0 \leq s < 1/\sqrt{2}$ gives

$$s^2 = \sin^2 \theta_i \leq \frac{1 - \sqrt{1 - 4(\delta/\lambda)^2}}{2}. \tag{26}$$

Note that this inequality holds for all $\theta_i$ $(i \leq k)$. If we write $\sin \Theta = \mathrm{diag}(\sin \theta_1, \ldots, \sin \theta_k)$, we have

$$\|\sin \Theta\|_F^2 \leq \frac{k}{2}\left(1 - \sqrt{1 - 4(\delta/\lambda)^2}\right) \tag{27}$$

Finally, utilizing the geometric identity for projection operators, $\|\hat{P} - P\|_{\mathrm{HS}}^2 = 2\|\sin \Theta\|_{\mathrm{F}}^2$, we obtain the exact subspace bound:

$$\|\hat{P} - P\|_{\mathrm{HS}} \leq \sqrt{k\left(1 - \sqrt{1 - 4(\delta/\lambda)^2}\right)} \tag{28}$$

Expansion at $\frac{\delta}{\lambda} = 0$ gives an asymptotic bound $\sqrt{2k}\frac{\delta}{\lambda} + \mathcal{O}\left(\frac{\delta^3}{\lambda^3}\right)$.

□

# B Additional Operator Details

## B.1 Distributional Symmetry Operator

We provide the derivation of the distributional symmetry operator introduced in Section 3.1. The symmetry relation in this case is $g_{\#}q = q$.

We can similarly linearize the residual $g_{\#}q - q$ at the identity. Expand the terms in the transformed $g_{\#}q$ where $g = \exp(\epsilon v)$:

$$\begin{aligned} q(g^{-1}x) &= q(x) - \epsilon\langle \nabla q(x), v(x)\rangle + \mathcal{O}(\epsilon^2) \\ \left|\det J_{g^{-1}}(x)\right| &= 1 - \epsilon \mathrm{tr}(\nabla v(x)) + \mathcal{O}(\epsilon^2) = 1 - \epsilon \nabla \cdot v(x) + \mathcal{O}(\epsilon^2) \end{aligned} \tag{29}$$

Then multiply these terms together to get the full expansion:

$$(g_{\#}q)(x) = q(x) - \epsilon \nabla \cdot (q(x)v(x)) + \mathcal{O}(\epsilon^2) \tag{30}$$

For the distribution to be invariant ($g_{\#}q = q$), the first-order term in $\epsilon$ must vanish. This gives the exact infinitesimal condition $\nabla \cdot (qv) = 0$. For practical considerations, we further normalize this by dividing by $q(x)$. We get the following scalar operator for distributional symmetry:

$$\mathcal{A}v = \frac{\nabla \cdot (qv)}{q} = \nabla \log q \cdot v + \nabla \cdot v. \tag{31}$$

### B.2 LieGAN Operator Formulation

Among existing methods, LieGAN (Yang et al., 2023) essentially operates in the distributional-symmetry domain, albeit targeting the joint data distribution of features and labels, $q(x, y)$. Here, we simply denote the joint data as $x \in \mathcal{X}$. The symmetry loss in LieGAN can be cast into the same form as above. In essence, LieGAN simultaneously trains a discriminator $D$ and a reparameterizable sampler for group elements $p_G$ (i.e., the "generator" in the GAN context). Its symmetry loss, w.r.t a group element sample from the generator, is given by

$$\mathbb{E}_{x\sim q,\ g\sim p_G}[\log(1 - D(gx))]. \tag{32}$$

We can show that, with certain adaptations in the design of LieGAN discriminators, and *assuming optimal discriminators*, the above loss term is equivalent to an infinitesimal counterpart

$$\mathcal{L}(V) \propto \sum_{i=1}^{k} \|\mathcal{A}v_i\|^2_{L^2(q;\mathbb{R})}, \tag{33}$$

where $\mathcal{A}$ is defined in (5). Since in practice we do not have access to the true density $q$, or equivalently, a perfect discriminator, we can similarly define a surrogate operator $\hat{\mathcal{A}}$ based on the learned discriminator. The discriminator's role here is analogous to the surrogate predictor in the setup of functional symmetry discovery. Specifically:

$$(\hat{\mathcal{A}}v)(x) = \nabla \log(1 - \hat{D}_v(x)) \cdot v(x), \tag{34}$$

where $\hat{D}_v$ is a discriminator trained to distinguish the original data distribution and the distribution after transformation along a specific direction $v$.

We would like to know what the loss (32) becomes when the sampled group element is around the identity. To this end, let the generator apply an infinitesimal transformation: $g_t x = \exp(tv)x,\ t \in [0, \epsilon]$, where $v$ has unit norm and $\epsilon$ is a small positive value bounding the size of the transformation. Denote the Lie derivative $\mathcal{L}_v q = \nabla \cdot (qv)$. The probability density of the generated data $q_{v,t}(x)$ under this specific transformation can be obtained from the continuity equation:

$$\begin{aligned} q_{v,t}(x) &= q(x) - t\mathcal{L}_v q(x) + \frac{t^2}{2}\mathcal{L}_v^2 q(x) + \mathcal{O}(t^3) \\ &= q(x)\left(1 - \mathcal{O}(\epsilon)\frac{\mathcal{L}_v q}{q}(x) + \mathcal{O}(\epsilon^2)\frac{\mathcal{L}_v^2 q}{q}(x) + \mathcal{O}(\epsilon^3)\right) \end{aligned} \tag{35}$$

Let's define $\Delta_v = \frac{\mathcal{L}_v q}{q} = \nabla \log q \cdot v + \nabla \cdot v$ and $\Xi_v = \frac{\mathcal{L}_v^2 q}{q}$. As the symmetry generator in LieGAN randomly samples the vector field direction $v$ and the size $t$ within $V$, the aggregated transformed distribution density $q_V$ can be estimated as

$$\begin{aligned} q_V(x) &= \mathbb{E}_{v,t}[q_{v,t}(x)] \approx q(x)(1 - \mathbb{E}_{v,t}[t\Delta_v(x) - \frac{t^2}{2}\Xi_v(x)]) \\ &= q(x)\left(1 - \mathbb{E}[t]\mathbb{E}_v[\Delta_v(x)] + \mathbb{E}[\frac{t^2}{2}]\mathbb{E}_v[\Xi_v(x)]\right). \end{aligned} \tag{36}$$

At this point, we introduce two modifications to the original LieGAN framework in order to align with the generator-learning objective (3). First, we require that $t$ be sampled from a biased distribution s.t. $\mathbb{E}[t] \neq 0$, as opposed to $\mathcal{N}(0, \sigma^2)$ and $\mathcal{U}(-a, a)$ used in the original implementation. In the latter case, the first-order term in $q_V(x)$ will vanish, and the most significant terms will involve higher-order differential operators than $\Delta_v$. With this assumption, we write $c = \mathbb{E}[t] = \mathcal{O}(\epsilon)$, and $\mathbb{E}[t^2] = \mathcal{O}(\epsilon^2)$.

Second, instead of uniformly sampling all directions in the Lie algebra, we require that each sample strictly follows the direction of one selected generator: $g = \exp(tv_i)$. In this case, we can decompose $\mathbb{E}_v[\Delta_v(x)]$ and $\mathbb{E}_v[\Xi_v(x)]$ as the weighted sum of $\Delta_v$ and $\Xi_v$. Conveniently, we can assign each generator $v_i$ with a uniform selection probability and obtain

$$q_V(x) = q(x)\left(1 - \frac{c}{k}\Delta(x) + \frac{\mathbb{E}[t^2]}{2k}\Xi(x)\right), \quad \Delta(x) = \sum_{i=1}^{k}\Delta_{v_i}(x),\ \Xi(x) = \sum_{i=1}^{k}\Xi_{v_i}(x). \tag{37}$$

A classical GAN result establishes that the optimal discriminator is $D(x) = \frac{q(x)}{q(x)+q_V(x)}$. Plugging in our first-order approximation of $q_V(x)$:

$$D(x) \approx \frac{q(x)}{2q(x) - \frac{c}{k}q(x)\Delta(x) + \frac{\mathbb{E}[t^2]}{2k}q(x)\Xi(x)} = \frac{1}{2 - \frac{c}{k}\Delta(x) + \frac{\mathbb{E}[t^2]}{2k}\Xi(x)}. \tag{38}$$

Using the Taylor expansion $(1-y)^{-1} \approx 1 + y + y^2$, we expand the discrminator around $\epsilon = 0$:

$$D(x) \approx \frac{1}{2}\left(1 + \mathcal{O}(\epsilon)\Delta(x) + \mathcal{O}(\epsilon^2)\Delta(x)^2 + \mathcal{O}(\epsilon^2)\Xi(x)\right) \tag{39}$$

We want to evaluate the LieGAN objective (32). Notice that we evaluate $D$ at $\exp(tv_i)(x)$ for a specific $i$. Noting that $\nabla D(x)$ is of order $\mathcal{O}(\epsilon)$, we use the following Taylor expansion for $D$:

$$D(\exp(tv_i)(x)) = D(x) + t\nabla D(x) \cdot v_i(x) + \mathcal{O}(\epsilon^3). \tag{40}$$

As we combine everything to expand $\log(1-D)$ up to second order, we get:

$$\log(1 - D(\exp(tv_i)(x))) \approx -\log 2 - \mathcal{O}(\epsilon)\Delta(x) - \mathcal{O}(\epsilon^2)\Delta(x)^2 - \mathcal{O}(\epsilon^2)\Xi(x) - \mathcal{O}(\epsilon^2)\nabla\Delta(x)\cdot v_i(x) \tag{41}$$

By the divergence theorem, the integral of a divergence over all space evaluates to zero (assuming the vector field vanishes at infinity). The relevant terms in (41) therefore vanish when we evaluate the expected loss over $x \sim q$, assuming $q(x) = 0$ at infinity:

$$\mathbb{E}_{x\sim q}[\Delta(x)] = \mathbb{E}_{x\sim q}\left[\frac{\mathcal{L}_v q(x)}{q(x)}\right] = \int \nabla \cdot (qv)dx = 0, \tag{42}$$

$$\mathbb{E}_{x\sim q}[\Xi(x)] = \mathbb{E}_{x\sim q}\left[\frac{\mathcal{L}_v^2 q(x)}{q(x)}\right] = \int \nabla \cdot ((\nabla \cdot (qv))v)dx = 0. \tag{43}$$

Similarly, we can find that two $\mathcal{O}(\epsilon^2)$ terms merge together. By divergence theorem, $\int \nabla\cdot(q\Delta v_i) = 0$. On the other hand, expanding $\nabla \cdot (q\Delta v_i)$, we have

$$\nabla \cdot (q\Delta v_i) = q\Delta\left(\frac{\nabla q}{q} \cdot v_i + \nabla \cdot v_i\right) + q\big(\nabla\Delta \cdot v_i\big) = q\Delta\Delta_{v_i} + q\big(\nabla\Delta \cdot v_i\big) \tag{44}$$

And therefore:

$$\mathbb{E}_{x\sim q}\big[\nabla\Delta(x) \cdot v_i(x)\big] = -\mathbb{E}_{x\sim q}\big[\Delta(x)\Delta_{v_i}(x)\big] \tag{45}$$

Evaluating (41) for each $v_i$ and w.r.t expectation over $x \sim q$ and factoring out $\mathcal{O}(\epsilon^2)$, we have the following infinitesimal counterpart of the original LieGAN symmetry loss (32):

$$\mathcal{L}(V) = \sum_{i=1}^{k} \mathbb{E}_q[\Delta^2 + C \cdot \Delta\Delta_{v_i}] \propto \mathbb{E}_q[\Delta^2] \tag{46}$$

where $C$ is an indefinite coefficient arising from combining the same-order terms.

To align with the per-vector-field separable loss similar to (3), we introduce a third modification to the LieGAN model implementation. In addition to enforcing the generator to output samples strictly in the vector field directions, we introduce a multi-head discriminator $D(x, i)$, where each head optimally distinguishes $q(x)$ from the specific transformed density $q_{v_i}$. Essentially, the model is organized as $k$ generator-discriminator pairs. Each generator learns a one-parameter subgroup, while the orthogonality regularization pushes generators away from each other. Under this setup, all $\Delta$ in the above expansions for the discriminator evaluation becomes $\Delta_{v_i}$, and the final expectation over the uniform selection of generator-discriminator pairs cleanly separates into a sum of independent norms:

$$\mathcal{L}(V) \propto \sum_{i=1}^{k} \mathbb{E}_q[\Delta_{v_i}(x)^2] = \sum_{i=1}^{k} \|\mathcal{A}v_i\|^2_{L^2(q;\mathbb{R})}, \tag{47}$$

where the scalar operator $\mathcal{A}$ is defined by the normalized Lie derivative of the density $q$:

$$(\mathcal{A}v)(x) = \nabla \log q(x) \cdot v(x) + \nabla \cdot v(x) = \frac{1}{q(x)}\nabla \cdot (q(x)v(x)) \tag{48}$$

It can also be verified that the adjoint operator of $\mathcal{A}$ is simply $\mathcal{A}^* = -\nabla$.

Deriving the form of the surrogate operator $\hat{\mathcal{A}}$ here requires slightly more effort than in the case of functional invariance/equivariance. In this discriminator-based formulation, the target density is not modeled explicitly, so the surrogate construction differs from that used for functional symmetry.

Instead, consider how we linearize the objective (41) around $t = 0$ when the optimal discriminator is replaced by a trained discriminator $\hat{D}_v$ for each vector field $v$:

$$F(t) = \log\left(1 - \hat{D}_v(\exp(tv)(x))\right) \tag{49}$$

$$\begin{aligned} F'(0) &= \nabla_z \log(1 - \hat{D}_v(z))\big|_{z=x} \cdot \frac{\partial}{\partial t} \exp(tv)(x)\big|_{t=0} \\ &= \nabla \log(1 - \hat{D}_v(x)) \cdot v(x) \end{aligned} \tag{50}$$

Therefore, we have the surrogate operator effectively defined as

$$\hat{\mathcal{A}} v = \nabla \log(1 - \hat{D}_v) \cdot v. \tag{51}$$

# C Experiment Details

## C.1 Sandbox experiments

This appendix provides the construction and implementation details for the functional, distributional, and PDE sandboxes in Section 4. Each setting admits an explicit ground-truth symmetry-measuring operator on a finite-dimensional candidate generator space. We construct controlled surrogate operators and compare the recovered generator subspaces with the known ground truth.

**Common notation.** Let $\mathcal{H}_d \subset \mathcal{H}$ be the $d$-dimensional candidate generator space of a sandbox, and let $\mathcal{H}'_{d'} \subset \mathcal{H}'$ be a finite-dimensional residual space containing the ranges of all ground-truth and surrogate operators considered in that sandbox. In orthonormal bases of these spaces, $\mathcal{A}$ and $\widehat{\mathcal{A}}$ are represented by matrices $M, \widehat{M} \in \mathbb{R}^{d' \times d}$. These representations preserve the singular values of their respective operators, with singular vectors expressed in the chosen bases. In particular, $\|\widehat{M} - M\|_2 = \|\widehat{\mathcal{A}} - \mathcal{A}\|_{\mathrm{op}}$ on $\mathcal{H}_d$.

Let $\sigma_1 \le \cdots \le \sigma_d$ and $\widehat{\sigma}_1 \le \cdots \le \widehat{\sigma}_d$ denote the singular values of $M$ and $\widehat{M}$ in ascending order, including zeros. We set

$$k = \dim\ker M, \qquad \lambda = \sigma_{k+1}, \qquad \delta = \|\widehat{M} - M\|_2, \qquad \rho = \delta/\lambda. \tag{52}$$

Here, $\lambda$ is the spectral gap of the ground-truth operator restricted to $\mathcal{H}_d$, and $\delta$ is the measured norm of the realized operator perturbation. We write $M = \sum_{j\ge 1} \sigma_{k+j} u_j y_j^\top$ for the positive part of its singular value decomposition, where $y_j$ and $u_j$ are the corresponding right and left singular vectors. We use $z_1, \ldots, z_k$ for an orthonormal basis of $\ker M$, and $\ell_1, \ell_2, \ldots$ for orthonormal vectors in $\ker M^\top$, referred to as left-null directions.

Let $e_i$ denote the $i$-th standard basis vector and $\mathbf{1}_N$ the all-ones vector in $\mathbb{R}^N$. For matrix generators, we use the Frobenius-orthonormal bases $K_{ij} = (e_i e_j^\top - e_j e_i^\top)/\sqrt{2}$ $(i < j)$ for skew-symmetric matrices, and $S_{ii} = e_i e_i^\top$, $S_{ij} = (e_i e_j^\top + e_j e_i^\top)/\sqrt{2}$ $(i < j)$ for symmetric matrices. For $N \ge 2$, the vector $t^{(N)} \in \mathbb{R}^N$ has equally spaced entries $t_i^{(N)} = -1 + 2(i-1)/(N-1)$. Throughout this appendix, $Y$ denotes a matrix with i.i.d. $\mathcal{N}(0,1)$ entries, with dimensions determined by the setting.

**Recovered subspace and error.** Let $P = ZZ^\top$, with $Z = [z_1, \ldots, z_k]$, be the orthogonal projector onto $\ker M$. We store the coordinates of $k$ recovered generators as the columns of $V \in \mathbb{R}^{d\times k}$. Let $P_V$ denote the orthogonal projector onto its column span. The recovery error is $\|P_V - P\|_F$, corresponding to the Hilbert-Schmidt projector distance in Theorem 3.1.

For the experiments in Section 4, we obtain $V$ using two approaches. (i) *SVD reference:* We compute the $k$ right singular vectors associated with the smallest singular values of $\widehat{M}$ and collect them in

$V = \widehat{Z}$. The recovered projector is $\widehat{P} := P_{\widehat{Z}}$. This approach is independent of $\tau$. (ii) *Iterative optimization:* We minimize the generator-learning objective (3), which in these coordinates becomes

$$\mathcal{L}(V) = \|\widehat{M}V\|_F^2 + \tau\|V^\top V - I_k\|_F^2. \tag{53}$$

We optimize (53) using Adam with a learning rate of $10^{-3}$ for 6000 steps.

For the subspace-error comparisons in Figure 1, we use the relative regularization grid $\tau = q_\tau \delta^2$, with $q_\tau \in \{0.5, 1, 2\}$. Each environment uses one fixed problem configuration, and reported errors are averaged over two random seeds. The functional sandbox uses $(m, n, k) = (2, 3, 1)$, the distributional sandbox uses $(m, k) = (2, 1)$, and the PDE sandbox uses $(m, n, k) = (2, 4, 2)$.

Reported errors are averaged over two random seeds. In the functional sandbox, each seed generates a different target coefficient matrix $W$, resulting in a different spectral gap $\lambda$. We therefore compute $\lambda$ separately for each target and average recovery errors at matching normalized perturbation strengths $\rho = \delta/\lambda$.

**Theorem bound.** The theoretical bound in Theorem 3.1 depends on $\rho$ and $k$, with the recovery guarantee applying under $\rho < 1/2$ and $\tau > \delta^2/2$. For $0 \le \rho < 1/2$, the comparison curve is

$$B_k(\rho) = \sqrt{k\left(1 - \sqrt{1 - 4\rho^2}\right)}. \tag{54}$$

**Two abstract adversaries.** We consider five perturbation modes: (a) direct rotation, (b) direct scaling, (c) physical rotation, (d) physical scaling, and (e) random coefficient perturbations. As discussed in Section 4.2, a perturbation can affect recovery by *rotating* the bottom-$k$ singular subspace or by *reordering* singular values across the gap. The direct adversaries (a) and (b) act on $M$ and allocate the entire perturbation budget $\delta$ to one of these mechanisms.

Both constructions use $r := \min\{k, \operatorname{rank} M\}$ triples $(z_j, y_j, \ell_j)$, $j \le r$. Each triple pairs a null direction $z_j$ with the right singular vector $y_j$ corresponding to the $j$-th smallest positive singular value $\sigma_{k+j}$, together with a left-null direction $\ell_j$.

**(a) Direct rotation.** For $j \le r$, choose $s_j \in (0, 1/\sqrt{2})$ and $c_j = \sqrt{1 - s_j^2}$ such that $\sigma_{k+j} s_j c_j = \delta$, or equivalently, $s_j^2 = \frac{1}{2}\big(1 - \sqrt{1 - 4\delta^2/\sigma_{k+j}^2}\big)$. Define the rotated vectors

$$\hat{z}_j = c_j z_j + s_j y_j, \qquad \hat{y}_j = -s_j z_j + c_j y_j, \qquad \hat{\ell}_j = s_j u_j - c_j \ell_j, \qquad \hat{u}_j = c_j u_j + s_j \ell_j. \tag{55}$$

The perturbation replaces each block $\sigma_{k+j} u_j y_j^\top$ by $\sigma_{k+j}\big(s_j^2 \hat{\ell}_j \hat{z}_j^\top + c_j^2 \hat{u}_j \hat{y}_j^\top\big)$:

$$\Delta M_{\mathrm{rot}} = \sum_{j=1}^{r} \sigma_{k+j}\big(s_j^2 \hat{\ell}_j \hat{z}_j^\top + c_j^2 \hat{u}_j \hat{y}_j^\top - u_j y_j^\top\big), \qquad \widehat{M} = M + \Delta M_{\mathrm{rot}}. \tag{56}$$

Restricted to $\operatorname{span}\{z_j, y_j\} \to \operatorname{span}\{\ell_j, u_j\}$, the $j$-th term equals $\sigma_{k+j} s_j c_j = \delta$ times an orthogonal $2 \times 2$ matrix. Different terms act on mutually orthogonal domains and ranges, so $\|\Delta M_{\mathrm{rot}}\|_2 = \delta$ exactly. The construction requires $\rho < 1/2$, and this sweep stops at $\rho = 0.49$.

This follows the extremal perturbation constructed in the proof of Theorem 3.1, tilting symmetry directions toward non-symmetry directions with small positive singular values. The bottom-$k$ subspace of $\widehat{M}$ becomes $\operatorname{span}\{\hat{z}_1, \ldots, \hat{z}_r, z_{r+1}, \ldots, z_k\}$, with nonzero principal angles $\arcsin s_j$ to $\ker M$. Consequently, $\|\widehat{P} - P\|_F = (2\sum_j s_j^2)^{1/2}$, which equals $B_k(\rho)$ when $r = k$ and $\sigma_{k+j} = \lambda$. Under these conditions, the construction attains the bound.

**(b) Direct scaling.** We construct the scaling adversary as

$$\Delta M_{\mathrm{sca}} = \delta \sum_{j=1}^{r} \big(\ell_j z_j^\top - u_j y_j^\top\big), \qquad \widehat{M} = M + \Delta M_{\mathrm{sca}}. \tag{57}$$

It preserves the singular directions of $M$, lifts the paired null singular values from 0 to $\delta$, and changes the paired positive singular values from $\sigma_{k+j}$ to $|\sigma_{k+j} - \delta|$.

The $2r$ rank-one terms have orthonormal right factors and orthonormal left factors, each with coefficient $\delta$. Hence $\|\Delta M_{\mathrm{sca}}\|_2 = \delta$.

This construction isolates the reordering mechanism: the singular directions remain fixed, while the gap narrows from both sides. True symmetry directions acquire residuals of size $\delta$, whereas the least-separated non-symmetry directions have residuals of size $\lambda - \delta$. Recovery is exact while $\delta < \lambda - \delta$, or $\rho < 1/2$. At $\rho = 1/2$, the singular values tie. For $\rho > 1/2$, the bottom-$k$ subspace is $\mathrm{span}\{y_1, \ldots, y_k\}$, orthogonal to the true subspace, with $\|\widehat{P} - P\|_F = \sqrt{2k}$. Thus, the condition $\delta < \lambda/2$ in Theorem 3.1 cannot be relaxed.

**Physical and random perturbations.** To model errors in approximations of the target objects, perturbations (c)–(e) modify the function coefficients, density score, or PDE coefficient matrix. We refer to perturbations induced through these underlying objects as coefficient-realizable, or "physical," perturbations. The induced operators are evaluated on the *true* data distribution, so the inner products of $\mathcal{H}$ and $\mathcal{H}'$ remain fixed.

For the affine perturbation paths, changing the underlying parameter along a fixed direction $E$ with amplitude $\alpha \geq 0$ gives

$$\widehat{\mathcal{A}} = \mathcal{A} + \alpha \mathcal{A}[E], \qquad \delta = \alpha \beta_E, \qquad \beta_E := \|\mathcal{A}[E]\|_{\mathrm{op}}, \tag{58}$$

where $\mathcal{A}[E]$ is the operator induced by a unit change along $E$, and $M[E]$ is its matrix representation. We compute $\beta_E$ once per direction. For a target ratio $\rho$, we set $\alpha = \rho\lambda/\beta_E$ and verify the resulting $\delta$ from $\widehat{M}$. The PDE axis-rotation path is nonlinear in its angle parameter, so its perturbation norm is evaluated numerically.

### C.1.1 Functional symmetry sandbox

**Model and operator.** The target is the linear function $f(X) = -\operatorname{tr}(W^\top X)$ on $X \in \mathbb{R}^{n\times n}$. The coefficient matrix $W \in \mathbb{R}^{n\times n}$ has its first $m$ rows equal to zero, while its remaining rows form a matrix $B \in \mathbb{R}^{(n-m)\times n}$ with i.i.d. $\mathcal{U}(0,1)$ entries, giving full row rank almost surely.

The candidate group $\mathrm{SO}(n)$ acts by $X \mapsto RX$, with generators $L \in \mathfrak{so}(n)$ and vector fields $v_L(X) = LX$, so $d = n(n-1)/2$. Since $W$ ignores the first $m$ rows of $X$, $f$ is invariant under rotations of those rows. The true generator subspace is therefore $\mathfrak{so}(m)$ embedded in the upper-left block of $\mathfrak{so}(n)$, with $k = m(m-1)/2$.

By (4), $(\mathcal{A}L)(X) = \nabla f(X) \cdot LX = \langle LW, X\rangle_F$. We identify $\mathcal{A}$ with the linear map $L \mapsto LW \in \mathbb{R}^{n\times n}$ and use the Frobenius inner product on both spaces. The columns of $M$ are $\mathrm{vec}(K_{ij}W)$, where vec denotes vectorization, and $\lambda = \mu_{m+1}/\sqrt{2}$, where $\mu_{m+1}$ is the smallest positive singular value of $W$.

**Controlled perturbations.** The direct adversaries (a) and (b) are constructed from $M$ as described above. For perturbations (c)–(e), the surrogate predictor is $\hat{f}(X) = -\operatorname{tr}(\widehat{W}^\top X)$ with $\widehat{W} = W + \alpha E$. Since $\mathcal{A}$ is linear in $W$, the induced perturbation is $\mathcal{A}[E] : L \mapsto LE$, and $\delta = \alpha\beta_E$ by (58).

**(c) Physical rotation.** We project the direct rotational perturbation onto the space of coefficient-induced perturbations: $E_{\mathrm{rot}} \in \arg\min_{E\in\mathbb{R}^{n\times n}} \|\Delta M_{\mathrm{rot}} - M[E]\|_F$. We then set $\widehat{W} = W + \alpha E_{\mathrm{rot}}$, giving $\delta = \alpha\beta_{E_{\mathrm{rot}}}$.

Among coefficient perturbations of this form, this construction most closely approximates the direct rotational perturbation in Frobenius norm. The surrogate $\hat{f}$ acquires a small dependence on the rows ignored by $f$, arranged to tilt its invariance subspace. Because the perturbation must be realizable through the predictor, its recovery error stays below that of (a) at equal $\delta$.

**(d) Physical scaling.** Let $W = Q_1 \operatorname{diag}(\mu_1, \ldots, \mu_n) Q_2^\top$ be a singular value decomposition with $\mu_1 = \cdots = \mu_m = 0 < \mu_{m+1} \leq \cdots \leq \mu_n$, and set $E_{\mathrm{sca}} = Q_1 \operatorname{diag}(\mathbf{1}_m, -1, 0, \ldots, 0) Q_2^\top$. Then $\widehat{W} = W + \alpha E_{\mathrm{sca}}$ raises the $m$ zero singular values of $W$ to $\alpha$ and lowers $\mu_{m+1}$ to $\mu_{m+1} - \alpha$, while keeping the singular vectors fixed. Since $\beta_{E_{\mathrm{sca}}} = 1$, we have $\delta = \alpha$.

The surrogate $\hat{f}$ acquires a small spurious sensitivity to the ignored rows while underestimating its weakest informative direction. This models an approximation that weakens an informative component without fully suppressing irrelevant components. Because the singular vectors of $W$ remain fixed, $\widehat{\mathcal{A}}$ has the same singular vectors as $\mathcal{A}$, and only its spectrum changes: the symmetry directions acquire singular value $\alpha$, while the least-separated non-symmetry directions acquire singular value $\big((\alpha^2 + (\mu_{m+1} - \alpha)^2)/2\big)^{1/2}$. Their ordering exchanges at $\alpha = \mu_{m+1}/2$, or $\rho = 1/\sqrt{2}$.

**(e) Random.** We set $\widehat{W} = W + \alpha E$ with $E = Y \in \mathbb{R}^{n\times n}$, giving $\delta = \alpha\beta_E$. This perturbation models unstructured estimation noise across all coefficients. Only part of its perturbation budget contributes to rotating the symmetry subspace.

### C.1.2 Distributional symmetry sandbox

**Model and operator.** Let $q = \mathcal{N}(0, I_m)$ and consider linear candidate generators $v_B(x) = Bx$, where $B \in \mathfrak{gl}(m)$ and $d = m^2$. Here, $\mathfrak{gl}(m)$ denotes the space of all real $m \times m$ matrices. Under $q$, $\|v_B\|_{\mathcal{H}} = \|B\|_F$. With the score $s(x) := \nabla \log q(x) = -x$ and $\nabla \cdot v_B = \operatorname{tr} B$, (5) gives

$$(\mathcal{A}B)(x) = \operatorname{tr} B + s(x)^\top Bx = \operatorname{tr} B - x^\top Bx. \tag{59}$$

We take the residual space to be $\mathcal{H}'_{d'} = \operatorname{span}\{h_\nu : |\nu| \le 4\}$, where $h_\nu(x) = \prod_i \mathrm{He}_{\nu_i}(x_i)/\sqrt{\nu_i!}$ are the normalized Hermite polynomials, orthonormal in $L^2(q)$. Here, $\mathrm{He}_j$ denotes the probabilists' Hermite polynomial of degree $j$, $\nu$ is a multi-index of nonnegative integers, and $|\nu| = \sum_i \nu_i$. The dimension is $d' = \binom{m+4}{4}$. This space contains all residuals considered below. In the bases $\{K_{ij}, S_{ij}\}$ and $\{h_\nu\}$,

$$\mathcal{A}K_{ij} = 0, \qquad \mathcal{A}S_{ij} = -\sqrt{2}\, h_{e_i+e_j} \quad (i<j), \qquad \mathcal{A}S_{ii} = -\sqrt{2}\, h_{2e_i}. \tag{60}$$

Thus, $\ker \mathcal{A} = \mathfrak{so}(m)$, $k = m(m-1)/2$, and all $m(m+1)/2$ positive singular values equal $\lambda = \sqrt{2}$. The range of $M$ lies in the degree-two block. The cubic functions $h_{2e_i+e_j}$ $(i<j)$ are orthogonal to this range and provide the left-null directions used in (a) and (b).

**Controlled perturbations.** The direct adversaries (a) and (b) are constructed from $M$ as described above. For perturbations (c)–(e), we replace the true score with the score of a surrogate density $\hat{q}$: $\hat{s} = \nabla \log \hat{q} = s + \alpha\dot{s}$, where $\dot{s}$ is a fixed score-perturbation direction. We have

$$\widehat{\mathcal{A}}B = \mathcal{A}B + \alpha\mathcal{A}[\dot{s}]B, \qquad (\mathcal{A}[\dot{s}]B)(x) := \dot{s}(x)^\top Bx, \qquad \delta = \alpha\beta_{\dot{s}}, \quad \beta_{\dot{s}} := \|\mathcal{A}[\dot{s}]\|_{\mathrm{op}}. \tag{61}$$

Residuals of $\widehat{\mathcal{A}}$ are measured in $L^2(q)$, so the surrogate score is evaluated under the fixed ground-truth data distribution.

**(c) Physical rotation.** We construct $\hat{q} = \mathcal{N}(0, C_\alpha^{-1})$ with precision matrix $C_\alpha = I + \alpha E$, where $E = \operatorname{diag}(t^{(m)})$. Then $\dot{s}(x) = -Ex$ and $(\widehat{\mathcal{A}}B)(x) = \operatorname{tr} B - x^\top C_\alpha Bx$. The perturbation strength is $\delta = \alpha\beta_{\dot{s}}$, with $(\mathcal{A}[\dot{s}]B)(x) = -x^\top EBx$.

This perturbation models a Gaussian density approximation with inaccurate per-axis variances. The surrogate retains an exact $k$-dimensional symmetry algebra corresponding to rotations that preserve $x^\top C_\alpha x$. Its Lie algebra is $C_\alpha^{-1/2}\mathfrak{so}(m)C_\alpha^{1/2}$, consisting of all $B$ satisfying $C_\alpha B + B^\top C_\alpha = 0$. Because these transformations preserve a perturbed metric, the recovered generator subspace rotates continuously away from $\mathfrak{so}(m)$ as $\alpha$ increases.

**(d) Physical scaling.** We define the surrogate density as

$$\hat{q}_\alpha(x) \propto \exp\Big[-\frac{1-3\alpha}{2}\|x\|^2 - \frac{\alpha}{4}\Big(\sum_i x_i^4 - \eta\|x\|^4\Big)\Big], \qquad \eta = 0.9/m. \tag{62}$$

Its score-perturbation direction is $\dot{s}(x) = 3x - x^{\odot 3} + \eta\|x\|^2 x$, where $x^{\odot 3}$ denotes the entrywise cube. Since $\sum_i x_i^4 \ge \|x\|^4/m > \eta\|x\|^4$ for $x \neq 0$, $\hat{q}_\alpha$ is normalizable for every $\alpha \ge 0$, and $\hat{q}_0 = q$. Here, $\delta = \alpha\beta_{\dot{s}}$.

This construction models an error in tail geometry: the surrogate has lighter tails along the coordinate axes than along the diagonals. The factor $1-3\alpha$ compensates, to first order, for the variance removed by the coordinate quartic, while $\eta$ controls how quickly non-symmetry directions lose separation. Because $\hat{q}_\alpha$ is invariant under signed coordinate permutations but not under rotations, $\widehat{\mathcal{A}}^*\widehat{\mathcal{A}}$ remains block diagonal in the decomposition of $\mathfrak{gl}(m)$ into skew-symmetric, off-diagonal symmetric, traceless diagonal, and scalar-multiple-of-identity matrices. Thus, no singular vector rotates.

**(e) Random.** We perturb the Gaussian family in (c) along a random direction: $C_\alpha = I + \alpha E$, where $E = (Y + Y^\top)/2$ is rescaled to unit spectral norm. We do not remove its trace, allowing a global scale error in addition to anisotropic errors. As in (c), $\delta = \alpha\beta_{\dot{s}}$.

### C.1.3 PDE SYMMETRY SANDBOX

**Model and operator.** Consider the anisotropic Helmholtz equation

$$F_D[u] = \operatorname{tr}(D\nabla^2 u) + \kappa u = 0 \quad \text{on } \mathbb{R}^n, \qquad D = \operatorname{diag}(aI_m, bI_{n-m}), \tag{63}$$

with $a, b, \kappa > 0$ and $a \neq b$. We refer to the first $m$ coordinates as block 1 and the remaining $n-m$ coordinates as block 2. The candidate generators are $v_L(x) = Lx$ with $L \in \mathfrak{so}(n)$, so $d = n(n-1)/2$. We use $x \sim \mathcal{N}(0, I_n)$, under which the generator inner product is the Frobenius inner product on $L$.

The data consist of exact plane-wave solutions $u(x) = \cos(\xi^\top x + \varphi)$, where $\xi = \sqrt{\kappa}\,\chi/\sqrt{\chi^\top D\chi}$, $\chi \sim \mathrm{Unif}(S^{n-1})$, and $\varphi \sim \mathrm{Unif}[0, 2\pi)$. Here, $S^{n-1}$ is the unit sphere. This solution ensemble is determined by the ground-truth coefficient matrix $D$. The prolonged residual in (6) becomes

$$(\mathcal{A}L)[u](x) = \big\langle [D, L], \nabla^2 u(x) \big\rangle_F = -\xi^\top [D, L]\xi \cos(\xi^\top x + \varphi), \tag{64}$$

where $[D, L] = DL - LD$. We measure this residual in $L^2$ over $(\xi, \varphi, x)$. The kernel is $\mathfrak{so}(m) \oplus \mathfrak{so}(n-m)$, with $k = \binom{m}{2} + \binom{n-m}{2}$, and all $m(n-m)$ positive singular values equal $\lambda = |a-b|\sqrt{\Psi_{12}}$. Here, $\Psi_{11}$, $\Psi_{12}$, and $\Psi_{22}$ denote $\mathbb{E}[\xi_i^2\xi_j^2]$ for distinct indices both in block 1, in different blocks, and both in block 2, respectively. These moments are computed by one-dimensional quadrature.

Equation (64) is linear in $[D, L]$ and in the Hessian coordinates $[\nabla^2 u]_S$ in the basis $\{S_{ij}\}$. The operator therefore has the matrix representation $M = \Gamma^{1/2} J(D)$, where $J(D)$ represents $L \mapsto [D, L]$ in the bases $\{K_{ij}\}$ and $\{S_{ij}\}$, and $\Gamma := \mathbb{E}\big[[\nabla^2 u]_S [\nabla^2 u]_S^\top\big]$ is the moment matrix of the solution ensemble. The residual space has dimension $n(n+1)/2$, giving $\dim\ker M^\top = k + n$, which provides enough left-null directions for (a) and (b).

We use $(n, m) \in \{(4, 2), (6, 3), (8, 4)\}$, $\kappa = 1$, and $(a, b) \in \{(1, 2), (1, 6)\}$. In the listed order of $(n, m)$, the spectral gaps are approximately $0.141$, $0.099$, $0.076$ for $(a, b) = (1, 2)$, and $0.357$, $0.243$, $0.183$ for $(a, b) = (1, 6)$.

**Controlled perturbations.** The direct adversaries (a) and (b) are constructed from $M$ as described above. For perturbations (c)–(e), the surrogate is a misspecified equation $F_{\widehat{D}}$ with a perturbed coefficient matrix $\widehat{D} \succ 0$. Its prolonged residual is evaluated on the ground-truth solution ensemble: $\widehat{M} = \Gamma^{1/2} J(\widehat{D})$, using the same $\Gamma$. This models error in the estimated PDE coefficients while keeping the observed solutions exact. Since $J$ is linear, $\widehat{\mathcal{A}} - \mathcal{A} = \mathcal{A}[\widehat{D} - D]$, where $(\mathcal{A}[E]L)[u] := \langle [E, L], \nabla^2 u \rangle_F$.

**(c) Physical rotation.** We set $\widehat{D} = R_\psi D R_\psi^\top$, where $R_\psi = \exp(\psi L_\times)$, $L_\times = \sum_{j=1}^{p} (e_j e_{m+j}^\top - e_{m+j} e_j^\top)$, and $p = \min\{m, n-m\}$. This rotates the coefficient matrix by angle $\psi$ in each of the $p$ coordinate planes $\operatorname{span}\{e_j, e_{m+j}\}$.

Define $E_\times = \sum_{j \le p} (e_j e_{m+j}^\top + e_{m+j} e_j^\top)$ and $E_\pm = \sum_{j \le p} (e_j e_j^\top - e_{m+j} e_{m+j}^\top)$. Then $\widehat{D} - D = (b-a)\big(\sin^2\psi\, E_\pm + \frac{1}{2}\sin 2\psi\, E_\times\big)$, so

$$\delta(\psi) = |a-b| \left\| \sin^2\psi\, \mathcal{A}[E_\pm] + \tfrac{1}{2}\sin 2\psi\, \mathcal{A}[E_\times] \right\|_{\mathrm{op}} = |\psi|\,|a-b|\,\|\mathcal{A}[E_\times]\|_{\mathrm{op}} + O(\psi^2). \tag{65}$$

We evaluate this perturbation norm numerically for $\psi \in (0, \pi/2]$.

The surrogate preserves the coefficient eigenvalues but changes the principal axes, modeling an error in the estimated orientation of an anisotropic medium. It retains an exact $k$-dimensional symmetry algebra about the tilted axes, $R_\psi\big(\mathfrak{so}(m) \oplus \mathfrak{so}(n-m)\big)R_\psi^\top$, so the recovered generator subspace varies continuously with $\psi$.

**(d) Physical scaling.** We set $\widehat{D} = D + \alpha E_{\mathrm{split}}$, where $E_{\mathrm{split}} = \mathrm{diag}\big(t^{(m)}, t^{(n-m)} - \mathbf{1}_{n-m}\big)$. The diagonal entries become $a + \alpha t_i^{(m)}$ in block 1 and $b - \alpha + \alpha t_i^{(n-m)}$ in block 2. The condition $\widehat{D} \succ 0$ requires $\alpha < \min\{a, b/2\}$.

Since $E_{\mathrm{split}}$ is diagonal, $\mathcal{A}[E_{\mathrm{split}}]$ is diagonal in the bases $\{K_{ij}\}$ and $\{S_{ij}\}$: it maps $K_{ij}$ to a multiple of $S_{ij}$, with entries $|(E_{\mathrm{split}})_{ii} - (E_{\mathrm{split}})_{jj}|\,\mathbb{E}[\xi_i^2\xi_j^2]^{1/2}$. The largest within-block entry is $\beta_{\mathrm{in}} := \max\{2\sqrt{\Psi_{11}}, 2\sqrt{\Psi_{22}}\}$. The largest cross-block entry is $3\sqrt{\Psi_{12}}$, corresponding to the last coordinate of block 1 and the first coordinate of block 2. Thus,

$$\delta = \alpha\beta_{\mathrm{split}}, \qquad \beta_{\mathrm{split}} = \max\big\{\beta_{\mathrm{in}}, 3\sqrt{\Psi_{12}}\big\}. \tag{66}$$

The surrogate preserves the principal axes but changes the PDE coefficients. The coefficients become unequal within each block, while the contrast between the block centers decreases by $\alpha$. This breaks the continuous symmetries without rotating singular vectors, isolating the reordering mechanism. The within-block singular values, corresponding to true symmetry directions, grow to at most $\alpha\beta_{\mathrm{in}}$, while the smallest cross-block singular value decreases to $(|a-b| - 3\alpha)\sqrt{\Psi_{12}}$ for $a < b$. Their first exchange occurs at

$$\rho_c = \frac{\beta_{\mathrm{split}}}{\beta_{\mathrm{in}} + 3\sqrt{\Psi_{12}}} \geq \frac{1}{2}, \tag{67}$$

after which the recovery error increases in steps.

**(e) Random.** We set $\widehat{D} = D + \alpha E$ with $E = (Y + Y^\top)/2$. The perturbation strength is $\delta = \alpha\beta_E$, where $\beta_E = \|\mathcal{A}[E]\|_{\mathrm{op}}$.

This perturbation models estimation noise in the PDE coefficient matrix, including off-diagonal couplings. Only the part of $E$ that couples the two blocks can rotate the recovered subspace, so a random direction allocates only part of its perturbation budget to rotation.

# D Additional Sandbox Results

We extend the subspace-error comparisons in Figure 1 to additional functional and distributional configurations. These experiments examine how the recovery behavior varies with the ambient dimension and the dimension of the true symmetry subspace, using the sandbox constructions described in Appendix C.1.

**Evaluation protocol.** Figures 3 and 4 report the projector distance $\|\widehat{P} - P\|_{\mathrm{HS}}$ against the normalized perturbation strength $\delta/\lambda$, with $\lambda$ computed from the corresponding ground-truth operator. We retain the five perturbation modes from the main text and compare iterative optimization with the bottom-$k$ SVD reference, where $k$ is the true symmetry dimension. The relative regularization grid is $\tau = q_\tau\delta^2$ with $q_\tau \in \{0.5, 1, 2\}$, denoted by $q$ in the figure legends. For positive $\delta$, the choice $q_\tau = 0.5$ is a boundary case and does not satisfy the strict condition $\tau > \delta^2/2$ in Theorem 3.1. The theoretical comparison curve uses the corresponding $k$ and is shown only for $\delta/\lambda < 1/2$; the direct rotational sweep is also restricted to this regime, as required by its construction.

## D.1 Functional symmetry across configurations

Figure 3 considers $(m, n) \in \{(2,3), (2,4), (3,4)\}$. The true generator subspace is $\mathfrak{so}(m)$ embedded in $\mathfrak{so}(n)$, giving symmetry dimensions $k = 1, 1, 3$, respectively. The first two configurations vary the ambient dimension while keeping the symmetry dimension fixed, whereas the third also increases the number of symmetry generators.

Direct rotational perturbations closely follow the theoretical curve in all three configurations. Physical rotational and random coefficient perturbations produce smooth error growth, with approximately

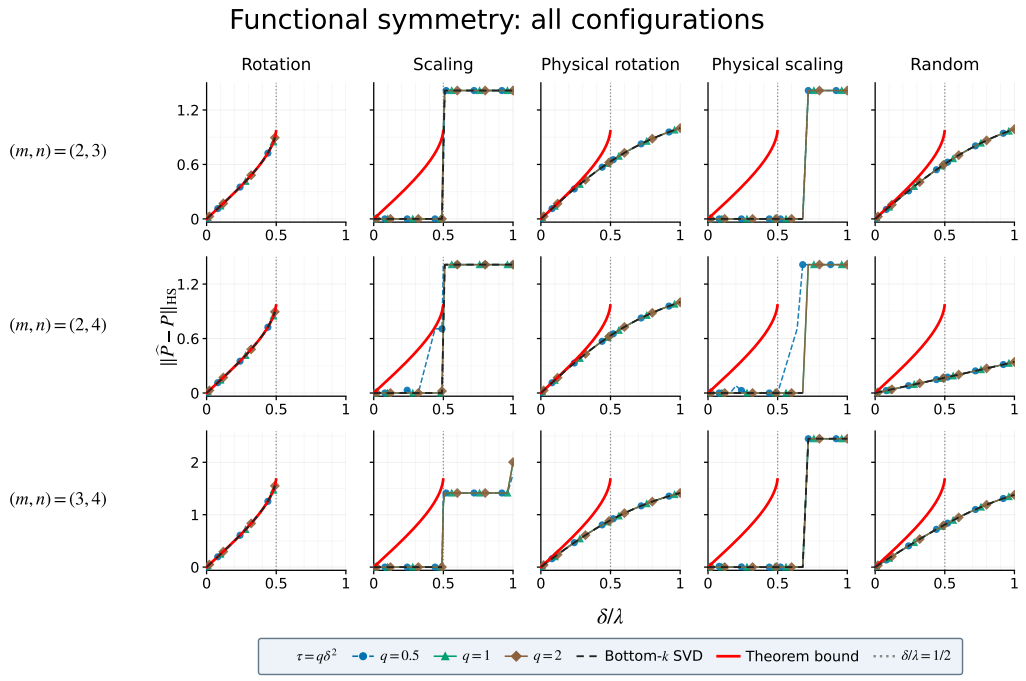


Figure 3: **Extended functional-symmetry results.** Rows correspond to $(m, n) = (2, 3)$, $(2, 4)$, and $(3, 4)$, with true symmetry dimensions $k = 1$, 1, and 3, respectively. Columns show direct rotation, direct scaling, physical rotation, physical scaling, and random coefficient perturbations. Each panel plots the projector distance $\|\widehat{P} - P\|_{\mathrm{HS}}$ against $\delta/\lambda$. Colored curves with markers show iterative optimization with $\tau = q_\tau \delta^2$ for $q_\tau \in \{0.5, 1, 2\}$; black dashed curves show the bottom-$k$ SVD reference. The red solid curve is the bound from Theorem 3.1, evaluated using the corresponding $k$ for $\delta/\lambda < 1/2$, and the vertical dotted line marks this threshold.

linear behavior at small $\delta/\lambda$ and generally smaller errors than the direct rotational adversary within the theorem's regime. The random perturbations also show appreciable variation between the two $m = 2$ configurations, despite their common symmetry dimension. Thus, the normalized perturbation strength and $k$ specify a common upper bound, not an identical recovery curve for every target and perturbation direction.

Scaling perturbations instead exhibit a near-zero plateau followed by a sharp increase in subspace error. The direct-scaling transition occurs near $\delta/\lambda = 1/2$, while the physical-scaling SVD reference transitions near $\delta/\lambda = 1/\sqrt{2}$, consistent with the coefficient-realizable construction in Appendix C.1. Iterative recovery usually tracks the SVD reference, but agreement is not uniform: the $q_\tau = 0.5$ curves for $(m, n) = (2, 4)$ depart from it near the scaling transitions, and a discrepancy is also visible at the largest direct-scaling perturbation for $(m, n) = (3, 4)$. These comparisons assess the empirical behavior of the iterative optimizer separately from the theorem's characterization of a globally optimized, full-rank solution.

### D.2 Distributional symmetry across dimensions

Figure 4 considers the isotropic Gaussian sandbox with $m \in \{2, 3, 5, 8\}$. The true generator subspace is $\mathfrak{so}(m)$, so $k = m(m-1)/2$ takes the values 1, 3, 10, and 28. The spectral gap remains $\lambda = \sqrt{2}$ in every configuration. Consequently, the theoretical comparison curve changes with $k$ even though the ground-truth spectral gap is fixed, which is important when comparing the vertical scales across rows.

The direct rotational adversary closely follows the dimension-dependent theoretical bound across all four configurations. Direct scaling produces an abrupt transition near $\delta/\lambda = 1/2$, followed by a projector distance close to $\sqrt{2k}$, consistent with replacement of the true symmetry subspace by an

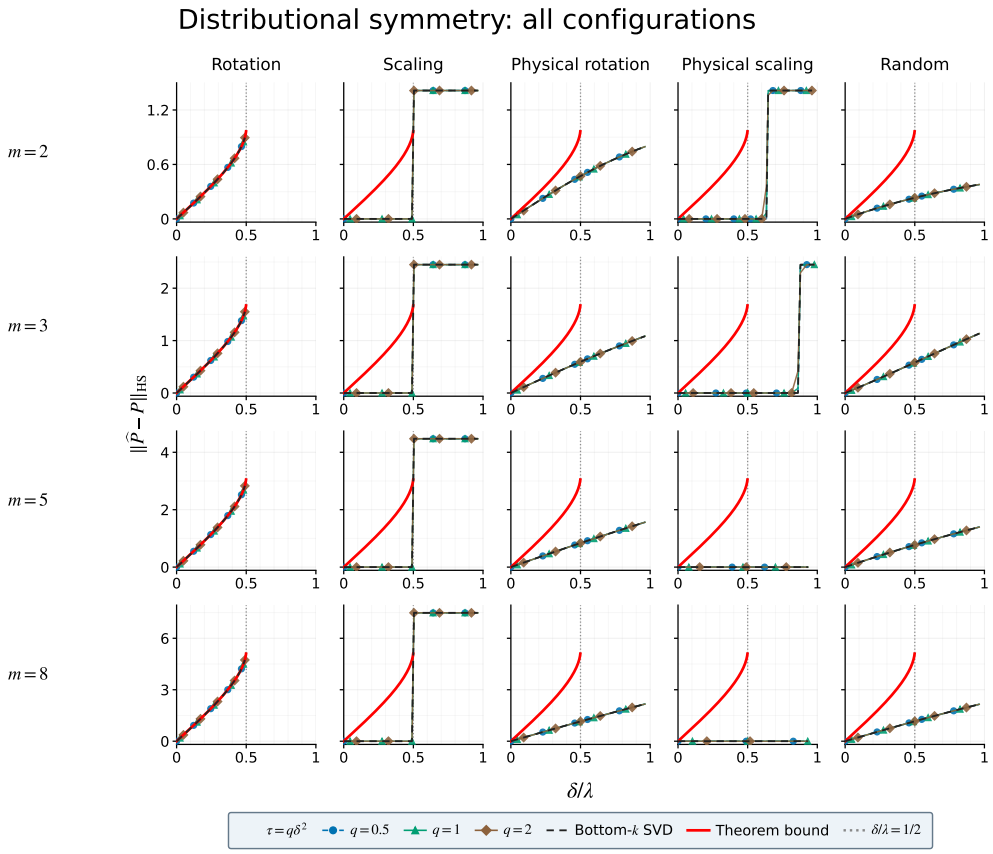


Figure 4: **Extended distributional-symmetry results.** Rows correspond to $m = 2$, 3, 5, and 8, with true symmetry dimensions $k = 1$, 3, 10, and 28, respectively, and a common spectral gap $\lambda = \sqrt{2}$. Columns show direct rotation, direct scaling, physical rotation, physical scaling, and random coefficient perturbations. Each panel plots $\|\widehat{P} - P\|_{\mathrm{HS}}$ against $\delta/\lambda$. Colored curves with markers show iterative optimization with $\tau = q_\tau \delta^2$ for $q_\tau \in \{0.5, 1, 2\}$; black dashed curves show the bottom-$k$ SVD reference. The red solid curve is the bound from Theorem 3.1 for the corresponding $k$, shown for $\delta/\lambda < 1/2$, and the vertical dotted line marks this threshold. Unlike direct scaling, physical scaling does not produce a transition within the displayed range for $m = 5$ or $m = 8$.

orthogonal competing subspace. The iterative curves remain close to the SVD reference across the tested dimensions and regularization choices.

Physical rotational and random perturbations continue to produce smooth error growth, remaining below the direct rotational adversary within the theorem's regime. Physical scaling shows a stronger dependence on the configuration: for $m = 2$ and $m = 3$, the error jumps beyond the half-gap threshold, whereas for $m = 5$ and $m = 8$, it remains near zero throughout the displayed perturbation range. These results illustrate that $\delta/\lambda < 1/2$ is a sufficient worst-case condition for stable recovery, not a universal failure threshold for every coefficient-realizable perturbation.

## E A Full Taxonomy of Symmetry Discovery Methods

This appendix expands the taxonomy introduced in Section 2, organizing continuous symmetry discovery by *target object*, *support structure*, *algorithmic paradigm*, and *action representation*. This taxonomy serves two purposes. First, it clarifies the scope and assumptions of existing work. For example, many methods identify continuous symmetry groups by learning their infinitesimal generators and target global subgroup symmetries (Yang et al., 2023; Shaw et al., 2024; Hu et al., 2025b), while recent distribution-based methods (Allingham et al., 2024; Urbano et al., 2025)

accommodate more general transformation sets. Second, it identifies directions for theoretical and methodological development. Each region of the taxonomy raises questions about when the target symmetry can be recovered, how the valid transformation set should be parameterized, and which learning objectives capture the desired symmetry relation.

We first give an abstract problem statement for continuous symmetry discovery. Let $\mathcal{X}$ denote the input/data space where transformations are evaluated, and let $\omega$ denote the target object whose symmetry is being studied, such as a function, a probability measure, or a differential operator. Let $G$ be an ambient connected Lie group of candidate transformations acting on $\mathcal{X}$.

A task-specific symmetry relation, written in pointwise form as $R_\omega(x, g) = 0$, determines whether a transformation $g \in G$ is valid at an input $x \in \mathcal{X}$. This induces the valid transformation set

$$S_x := \{g \in G : R_\omega(x, g) = 0\}. \tag{68}$$

The goal of symmetry discovery is to recover or approximate the support of this relation in $\mathcal{X} \times G$, or equivalently, the family of valid transformation sets $\{S_x\}_{x\in\mathcal{X}}$. In some settings, the action of $G$ on $\mathcal{X}$ is fixed in advance. In others, discovery also involves learning how group elements act on the data, since the same abstract group may admit different actions on the same space.

Under this formulation, we organize symmetry discovery methods along four axes:

- **Target Object and Data Modality:** The mathematical object whose symmetry is being investigated, which determines the symmetry relation and the data available for discovery.
- **Support Structure of the Symmetry Relation:** The input dependence of the valid transformation set (global versus input-dependent) and its algebraic structure (subgroup versus generic subset).
- **Algorithmic Paradigm:** The computational framework for parameterizing and optimizing over transformations.
- **Action Representation:** The class of transformations used to represent the group action on the data.

### E.1 Target Object and Data Modality

The first axis classifies methods by the mathematical object whose symmetry properties are sought. This determines the symmetry relation and shapes the learning objectives and data requirements.

**Functional Symmetry.** One motivation for symmetry discovery is to learn flexible inductive biases for invariant or equivariant neural networks (Yang et al., 2023). Given a function $f : \mathcal{X} \to \mathcal{Y}$, the goal is to discover transformations on $\mathcal{X}$ such that $f(gx) = f(x)$. For equivariance, a group action on $\mathcal{Y}$ is also modeled, and the relation becomes $f(gx) = gf(x)$, with the input and output actions distinguished when necessary.

As the underlying function $f$ is often not directly available, this setting typically uses a dataset of input-target pairs $\{(x_i, y_i)\}$. A surrogate model $\hat{f}$ is trained from these pairs and used to evaluate functional outputs and the corresponding invariance or equivariance errors (Moskalev et al., 2022). We refer to this setting as *functional symmetry*.

**Distributional Symmetry.** Another branch of the literature studies symmetries of a data distribution. Here, the target object is an unknown probability measure $q$ over the data space $\mathcal{X}$, or a density $q(x)$ when one exists. The objective is to identify transformations that preserve this measure:

$$g_\# q = q, \qquad (g_\# q)(x) = q(g^{-1}x)\,|\det J_{g^{-1}}(x)|.$$

Here, $g_\# q$ denotes the pushforward measure, and the density expression uses the Jacobian $J_{g^{-1}}$ of the inverse transformation.

When the density is not explicitly available, discovery relies on empirical samples from the target distribution. These samples may be unlabeled observations or joint input-output pairs. Methods in this category often use generative modeling approaches, including generative adversarial networks (GANs) (Yang et al., 2023), normalizing flows (Allingham et al., 2024), and diffusion or flow-matching models (Chen et al., 2025). Comparing the target and transformed distributions through

statistical divergences encourages the learned transformations to preserve the underlying probability measure. We refer to this setting as *distributional symmetry*.

A notable case is LieGAN (Yang et al., 2023): although motivated by functional symmetry, its objective enforces invariance of the joint input-output distribution. The relationship between joint-distributional and functional symmetries depends on additional assumptions about the data distribution and group action. We therefore classify LieGAN as a method for distributional symmetry discovery, according to the target of its learning objective.

**PDE Symmetry.** A third target category arises in scientific machine learning, where the object of interest is the solution set of a partial differential equation (PDE). Consider a PDE system represented abstractly as $\mathcal{F}[u] = 0$, where $\mathcal{F}$ is a differential operator and $u$ is a solution field over a space-time domain. The goal is to discover Lie point symmetries that act on the independent and dependent variables and map valid solutions to valid solutions: $\mathcal{F}[u] = 0 \Rightarrow \mathcal{F}[gu] = 0$.

Datasets in this setting typically contain PDE solutions discretized on a spatiotemporal grid. Discovery methods evaluate the residuals of transformed solutions with respect to differential operators that are either given (Ko et al., 2024) or learned from data (Liu et al., 2025). We refer to this setting as *PDE symmetry*.

These three categories are not exhaustive. Other structured objects can be incorporated into the taxonomy by specifying the corresponding symmetry relation. As developed in Section 3, generator-based discovery admits a common operator formulation across the three settings above. The target object determines the symmetry-measuring operator, while the recovery objective has a shared form. This connection motivates treating the target object and algorithmic paradigm as distinct taxonomic axes.

### E.2 Support Structure of the Symmetry Relation

Many symmetry discovery methods focus on global subgroup symmetries: finding a subgroup $H \leq G$ such that the symmetry relation holds for all inputs $x \in \mathcal{X}$ and all transformations $g \in H$. In the generic formulation above, this corresponds to $S_x = H$ for all $x$. More generally, the valid transformation sets $\{S_x\}_{x \in \mathcal{X}}$ may vary across the input domain and need not form subgroups. We refer to the structure of this family as the *support structure* of the symmetry relation.

We distinguish two aspects of support structure: *input domain locality* and *algebraic structure*. Together, they describe how the valid transformation set varies with the input and which group operations it preserves.

**Input Domain Locality.** The first sub-axis concerns whether the valid transformation set depends on the input. We use "local" as shorthand for input-dependent symmetry.

- **Global Symmetry:** The valid transformation set is the same across the input domain: there exists $S \subseteq G$ such that $S_x = S$ for all $x \in \mathcal{X}$. The discovery objective is to identify the maximal set $S$ for which the symmetry relation holds for every $(x, g) \in \mathcal{X} \times S$.
- **Input-Dependent Symmetry:** The valid transformation set $S_x$ varies with $x$. Given an ambient candidate group $G$, discovery seeks to identify the valid set $S_x \subseteq G$ at each input. This setting accommodates symmetry relations affected by asymmetric boundary conditions or irregular domain geometry. For example, Figure 5 illustrates how an obstacle in a 2D navigation task breaks the global $\mathrm{SO}(2)$ symmetry of the optimal policy, leaving input-dependent subsets of valid rotations.

**Algebraic Structure.** The second sub-axis concerns the algebraic properties of $S_x$ within the ambient group $G$.

- **Subgroup Closure:** For a given $x$, the set $S_x$ forms a subgroup of $G$, denoted $S_x \leq G$. It is closed under composition and inversion: if $g_1, g_2 \in S_x$, then $g_1 \circ g_2 \in S_x$ and $g_1^{-1} \in S_x$.
- **Generic Subset:** For a given $x$, the set $S_x \subseteq G$ does not satisfy subgroup closure. This allows a transformation direction to be valid only within a bounded range of magnitudes. For

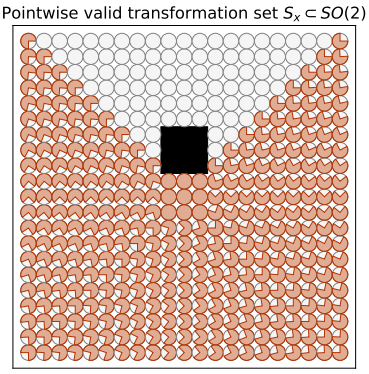


Figure 5: Symmetry of an optimal policy for navigating towards the center of a 2D domain. The policy takes a position vector as input and returns a unit vector specifying a movement direction. The obstacle, shown in black, breaks global rotational equivariance, while some rotations remain valid for inputs in the lower part of the domain. The circular plot at each grid point shows the valid transformation set $S_x \subset \mathrm{SO}(2)$, with the rightward direction corresponding to the identity transformation.

$g \in S_x$, neither repeated application $g \circ g$ nor the inverse transformation $g^{-1}$ is guaranteed to remain in $S_x$. For example, in Figure 5, the policy at an input in the lower half of the domain is equivariant under rotations up to $\pi/4$, while a larger rotation can move the input to a region where the obstacle changes the optimal direction.

Combining input domain locality and algebraic structure gives the following three scenarios:

1. **Global Subgroup Symmetry.** The valid transformation set is uniform across the domain and forms a fixed closed subgroup $H \leq G$, with $S_x = H$ for all $x \in \mathcal{X}$. Many generator-based methods target this regime, representing a fixed subgroup through a set of infinitesimal generators $V = \{v_i\}$.
2. **Input-Dependent Subgroup Symmetry.** Each valid transformation set satisfies subgroup closure, but the subgroup depends on the input: $S_x \leq G$. Recent work has begun to explore this setting, including AtlasD (Bhat et al., 2025). A symmetry model can represent this dependence through an input-dependent basis of infinitesimal generators $V(x) = \{v_i(x)\}$.
3. **Input-Dependent Subset Symmetry.** The valid transformation set $S_x \subset G$ varies with the input and lacks subgroup closure. Generator-based representations may be insufficient because they do not directly describe a transformation direction that is valid only over a bounded range. As discussed in Section E.3, methods that learn distributions over the ambient group $G$ offer a flexible representation for this setting.

Under the convention that $S_x$ denotes the maximal valid transformation set, the remaining combination, *Global Subset Symmetry*, does not arise for the object-level symmetry relations considered here. Suppose there is a fixed set $S \subseteq G$ such that $S_x = S$ for all $x \in \mathcal{X}$. For functional invariance,

$$S = \{g \in G : f(gx) = f(x),\ \forall x \in \mathcal{X}\}.$$

If $g, h \in S$, then $f((g^{-1}h)x) = f(x)$ for every $x \in \mathcal{X}$, so $g^{-1}h \in S$. Thus, $S$ is a subgroup of $G$.

More generally, the global object-level symmetry relations considered in this paper can be expressed through the stabilizer of the target object $\omega$ under an induced action of $G$ on an appropriate object space $\Omega$:

$$S = \mathrm{Stab}_G(\omega) := \{g \in G : g \cdot \omega = \omega\}. \tag{69}$$

Here, $\omega$ may be the graph of a function, a probability distribution, or a PDE solution set. Stabilizers are subgroups, so a maximal globally valid transformation set cannot lack subgroup closure. Under this convention, subset-valued symmetry arises through input dependence.

### E.3 Algorithmic Paradigm

The third axis classifies methods by how transformations are parameterized and optimized. Algorithmic choices are influenced by the target object and support structure. In particular, subgroup symmetries naturally motivate representations through infinitesimal generators, while generic subset-valued symmetries motivate representations through distributions over the ambient group.

We distinguish two broad paradigms: **generator-based discovery** and **distribution-based discovery**. Generator-based methods learn transformation directions, usually as matrix generators or, more

generally, vector fields. Once a direction is learned, the method implicitly assumes that the symmetry relation remains valid along the corresponding one-parameter subgroup. This representation is particularly suited to global subgroup symmetry and, with input conditioning, potentially to input-dependent subgroup symmetry. Its applicability is more limited for generic subset-valued symmetries, where transformations along a direction may be valid only within a bounded range.

Distribution-based methods learn probability distributions over non-infinitesimal group elements, often using generative model architectures. In the global case, this may be a distribution $p_\theta(g)$ over $G$; in the input-dependent case, it becomes a conditional distribution $p_\theta(g \mid x)$. The valid transformation set can be approximated by the support or a high-density region of the learned distribution. For example, with a density threshold $\epsilon > 0$, one may define

$$\widehat{S}_x(\epsilon) = \{g \in G : p_\theta(g \mid x) > \epsilon\}. \tag{70}$$

This representation does not require closure under composition or inversion and can express bounded intervals, disconnected regions, and other non-subgroup supports. Generator-based methods therefore provide a compact representation for subgroup symmetry, while distribution-based methods accommodate more general input-dependent transformation sets.

### E.4 Action Representation

The fourth axis concerns how an abstract group element $g \in G$ is realized as a transformation $x \mapsto gx$ on the data. It distinguishes the hypothesis classes used to represent group actions and motivates the vector-field formulation adopted in our analysis.

Early symmetry discovery methods typically assume that transformations act linearly on the input space (Yang et al., 2023). Infinitesimal generators are then represented by matrices, and the discovered Lie algebra is a finite-dimensional subspace of matrix generators. This representation is compact and mathematically convenient, but restricts the discovered symmetries to linear or affine transformations in the chosen coordinates.

More recent methods allow nonlinear group actions. These can be represented by learning linear transformations in a latent space and mapping them back to the input space (Yang et al., 2024), expanding vector fields in a predefined library of nonlinear basis functions (Shaw et al., 2024), or parameterizing infinitesimal generators as neural vector fields (Ko et al., 2024). These approaches differ architecturally but share the use of potentially nonlinear vector fields to represent transformation directions.

Action representation is distinct from support structure: a symmetry may be global or input-dependent, and subgroup- or subset-valued, regardless of whether the action is linear. For generator-based discovery, the representation determines the ambient space in which the symmetry subspace is recovered. The analysis in Section 3 is therefore formulated in a Hilbert space of vector fields, with matrix-generator methods included as a special case. Table 1 summarizes representative methods according to these four axes.