AUDIT BEFORE YOU INTERPRET: BASIS FREEDOM IN TRAINED NEURAL OPERATORS

Anonymous authors
Paper under double-blind review

ABSTRACT

A trained neural operator can make identical predictions in different internal bases, and those bases yield different interpretations and visualizations of its learned representation. We develop a framework for basis-consistent interpretation and visualization of trained neural operators. Its checkpoint audit applies to both purely data-driven and physics-informed models without requiring a linear decoder. We test candidate basis changes on public checkpoints from several architecture families. Each test applies a compensated reparameterization to the stored weights and includes a paired control expected to change the predictions. These tests allow interpretations to be checked across equivalent representations. For physics-informed cavity operators with low-rank linear decoders, we identify parameter-aligned subspaces using the model's own coefficients and distances between decoded physical fields. We fix these subspaces before consulting reference solutions for validation. The identified asymmetry subspace captures about 97% of the energy in the reference asymmetry component. Projections onto each fixed decoded subspace are independent of its internal basis. Embeddings built from field distances that are the same in every basis recover more reference neighbours than embeddings of standardized field-varimax coordinates. These results support physical interpretation through validated subspaces and visualization through distances that respect the audited basis freedom.

1 INTRODUCTION

Researchers use learned representations to study how physical systems respond to changing parameters. Latent-axis plots, decoded modes and embeddings help assign physical meaning to coordinates, compare trained models and select components for reuse. These interpretations depend on the chosen basis and the distance measure, even when the predictions remain unchanged. Consider an operator trained for lid-driven cavity flow, parameterized by the Reynolds number Re and the strength κ of a second lid. A linear decoder maps five coefficients to each predicted velocity field. Figure 1 shows how equivalent bases give these coordinates different physical labels.

In the varimax basis, different axes vary with Re, κ or their interaction. Rotating the coefficients and compensating in the decoder preserves every prediction. In the rotated basis, κ dominates every axis. Re dependence remains encoded across the coordinates. An axis's physical label therefore depends on the chosen basis. The choice of distance also matters. For a linear decoder, Euclidean distances between stored coefficients need not match distances between decoded fields when its modes have unequal norms or are non-orthogonal. We assess physical interpretations by comparing with reference solutions and testing consistency under prediction-preserving basis changes.

Previous work addresses parts of this problem. Singular value decomposition (SVD) extracts orthonormal modes ordered for deep operator networks (DeepONets) (Venturi & Casey, 2023; Peyvan et al., 2024). Varimax selects an orientation that simplifies their loading structure (Kaiser, 1958). DisentangO learns the physical latent factors identifiable up to permutation and component-wise invertible transformations under its assumptions (Liu et al., 2026). Symmetry analysis derives prediction-preserving transformations from network operations (Godfrey et al., 2022). Elhage et al. (2023) rotated a basis-independent residual stream with compensation and found the test loss almost unchanged. Cain (2026) shows that equivalent representations can have different nearest neighbours and recommends invariant quantities or an explicitly stated basis. Embeddings such as t-SNE

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

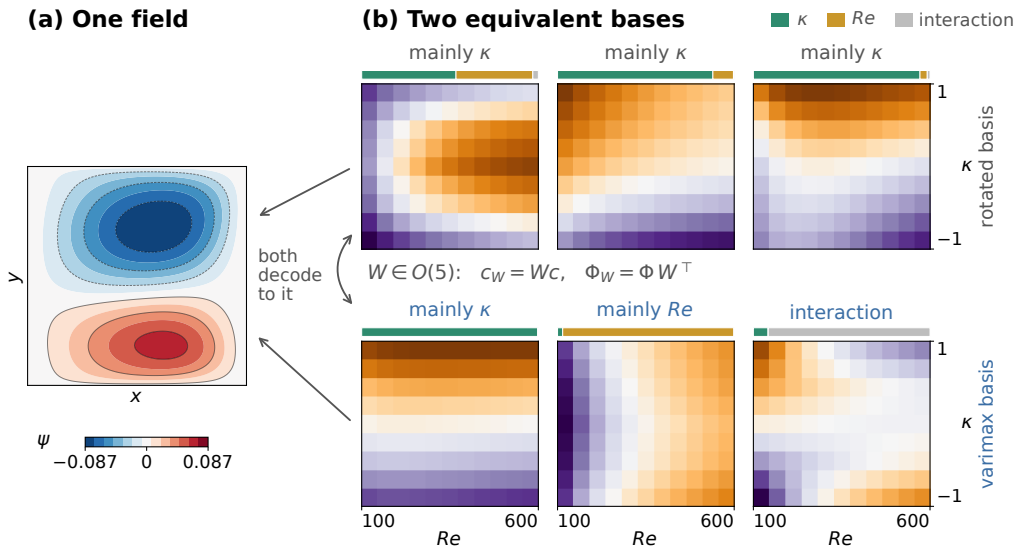


Figure 1: One decoded field, two internal bases. **(a)** The decoded streamfunction ψ at $Re = 350$, $\kappa = +1$. Both bases in (b) decode to it. **(b)** Three of five coordinate surfaces in field-orthonormal coordinates (§2.3), shown over the parameter grid (Re horizontal, κ vertical), each on its own colour scale. Bars divide variation into κ, Re and interaction shares, and titles name the effect whose share exceeds one half. The varimax axes have different dominant effects (lower row). An orthogonal rotation $\boldsymbol{W}$ makes κ dominant on all five axes (upper row). Reparameterizing the decoder $\boldsymbol{\Phi}$ as $\boldsymbol{\Phi W}^\top$ preserves every decoded field.

(van der Maaten & Hinton, 2008) depend on distances determined by their input metric and preprocessing. Appendix B discusses further related work.

We propose a checkpoint-level procedure for the physical conclusions that survive prediction-preserving basis changes. A paired audit certifies which basis changes a checkpoint admits. For linear decoders, a specified norm on decoded fields induces the *field metric*, linking latent geometry to physical responses. Field varimax identifies parameter-aligned subspaces from the model's coefficients in that metric, and reference solutions validate them. Projected response fields, embeddings and block-shared prediction are then built so that they are the same in every certified basis. The procedure combines symmetry analysis, SVD, varimax and compensated weight reparameterizations. It applies to both purely data-driven and physics-informed neural operators, because it reads only the checkpoint, and reference solutions enter after identification, for validation and completion. Any trained model with available weights and forward code can be audited, and any operator with a linear decoder and reference solutions can have its subspaces identified and validated.

Our framework makes three contributions. First, a paired checkpoint audit tests analytically derived candidate transformations through compensated reparameterizations of stored weights. Each test includes a control expected to change predictions. The audit provides numerical evidence at specified inputs, precision and tolerances without retraining or reference solutions. Second, we identify parameter-aligned subspaces for operators with linear decoders and validate their physical labels. Identification uses the model's own coefficients in the field metric. We fix the subspaces before reading reference solutions, then measure the fraction of each reference response component's energy they capture in the same metric. Third, we develop basis-consistent interpretation and visualization. Projected response fields are unchanged by basis changes within a fixed decoded subspace. Embedding inputs use field-metric distances that are unchanged under invertible basis changes. We assess consistency under the certified transformations and compare embedding neighbourhoods with those defined by reference-solution distances.

2 Methodology

We audit a checkpoint without physics (§2.1–2.2), then a second stage identifies, validates and uses physical subspaces on operators with reference solutions (§2.3–2.6).

2.1 ADMISSIBLE TRANSFORMATIONS OF THE IMPLEMENTED REPRESENTATION

The audit finds the changes of basis that a trained checkpoint leaves invisible in its output. It needs the checkpoint, its forward implementation, the operations that produce and read the representation, and probe inputs. Let $\boldsymbol{z}$ denote the representation, θ the model weights and D_θ all operations downstream of $\boldsymbol{z}$. A transformation g is *admissible* when the operations that produce and read $\boldsymbol{z}$ can be reparameterized within the architecture so that

$$D_{\theta_g}(g\boldsymbol{z}) = D_\theta(\boldsymbol{z}). \tag{1}$$

Here θ_g denotes the reparameterized weights. We test Equation (1) numerically for each candidate. Next we illustrate how a reparameterization compensates for a change of basis. Suppose that the consumer of $\boldsymbol{z}$, the first operation that reads it, begins with an affine map $L(\boldsymbol{z}) = A\boldsymbol{z} + b$. An invertible change of basis $\boldsymbol{z}' = M\boldsymbol{z} + d$ is absorbed by

$$A' = AM^{-1}, \qquad b' = b - AM^{-1}d, \qquad L'(\boldsymbol{z}') = L(\boldsymbol{z}), \tag{2}$$

where primes denote reparameterized weights, provided that the producer of $\boldsymbol{z}$ can also realize $M\boldsymbol{z}+d$ by a weight reparameterization. An elementwise nonlinearity between producer and consumer acts on each channel separately. It commutes with a reordering of the channels and does not commute with a general mixture of them. The last Fourier block studied below has no activation between its linear paths, and every invertible affine channel map is admissible there. An earlier block contains a Gaussian error linear unit (GELU). Unsigned permutations of the channels commute with it, and a generic rotation does not. The admissible transformations depend on the entire chain from producer to consumer (Figure 2(b)).

2.2 A PAIRED TEST AND ITS POSSIBLE OUTCOMES

A small output change alone does not establish that the full candidate family preserves the predictions. The tested transformation may belong to a smaller valid subgroup or affect only features that the model does not use. We pair each reparameterization with a *control*, a different transformation or the same reparameterization at another location, expected to change predictions despite compensation. For the last Fourier block, the control applies the same affine reparameterization one block earlier, where the GELU should break it.

Each reparameterization and each control is scored by how far it moves the output. For nonzero baseline output y and reparameterized output y_g, we record the relative deviation

$$\delta_\ell(g) = \frac{\|y_g - y\|_\ell}{\|y\|_\ell}. \tag{3}$$

Here ℓ is the Euclidean norm or the largest absolute entry, as listed for each architecture family in Appendix G. Probes use 64-bit floating-point arithmetic (float64) and report the maximum over draws on synthetic inputs (Appendix G).

Algorithm 1 Paired audit of one candidate transformation.

Require: Checkpoint, forward code, producer and consumers of a representation, probe inputs
1: Derive a candidate and its compensating reparameterization; name a control it should fail to absorb.
2: Fix probe inputs, draws, precision, norm, tolerance and control threshold.
3: On separate copies, record the maxima $\epsilon_{\text{reparam}}$ and $\epsilon_{\text{control}}$ of δ_ℓ.
4: If $\epsilon_{\text{reparam}}$ exceeds the tolerance, reject the candidate.
5: Otherwise, if $\epsilon_{\text{control}}$ stays below the control threshold, report the pair as inconclusive.
6: Otherwise, certify the tested reparameterization at the recorded inputs, precision and tolerances.

A pair whose control also leaves the output unchanged is inconclusive, because the test cannot separate the candidate from numerical inactivity or a weak control. A pair whose candidate preserves the output while its control changes it yields a *certificate*. The certificate is numerical evidence, at the declared inputs, precision and tolerances, that the sampled compensated reparameterization preserves the output and its named control changes it. The analytical derivation identifies the associated transformation family. The *reparameterization class* is the set of weight reparameterizations the analyst declares for a location, for example the affine channel maps. We call the certified family maximal in the declared reparameterization class when the derivation rules out any larger family, and a certified subgroup otherwise. An axis-level statistic attributed to the represented function

must stay unchanged under the certified transformations. A single certified change that alters the statistic rejects that attribution. The analyst specifies candidates and controls (Appendix A.2), and the checkpoint-only tests require no physical responses or screening criterion.

2.3 CAVITY REPRESENTATION AND SUBSPACE IDENTIFICATION

We study lid-driven cavity flow parameterized by the Reynolds number Re and the strength κ of a second, opposing lid, which we call the asymmetry parameter. Each operator is a physics-informed auto-decoder, trained from the partial differential equation (PDE) and boundary residuals alone, with no reference solutions (Appendix C). The operators reach a relative error of about 15% against the reference solutions (Table 10). The last layer of each operator is a rank-5 linear decoder. It decodes $r = 5$ modes of a 200-dimensional latent vector into a correction to a prescribed field, the parameter-dependent boundary lift plus the fixed mean. The correction is $\boldsymbol{\Phi}\boldsymbol{c}(p)$, where $p = (Re, \kappa)$, $\boldsymbol{c}(p)$ is the coefficient vector and $\boldsymbol{\Phi}$ the decoded mode matrix. An *axis* is one coefficient direction, its *coefficient surface* is that coefficient over the parameters, and its *decoded mode* is the corresponding column of $\boldsymbol{\Phi}$, a velocity field. A *block* is a group of axes, and its *decoded subspace* is the span of their decoded modes. An *internal rotation* changes the axes inside a block and leaves its decoded subspace unchanged. The orthogonal group $O(r)$ contains the real $r \times r$ matrices that rotate or reflect coefficient vectors, $\top$ denotes transpose, and for every $\boldsymbol{W} \in O(r)$,

$$\boldsymbol{c}_{\boldsymbol{W}} = \boldsymbol{W}\boldsymbol{c}, \qquad \boldsymbol{\Phi}_{\boldsymbol{W}} = \boldsymbol{\Phi}\boldsymbol{W}^{\top}, \qquad \boldsymbol{\Phi}_{\boldsymbol{W}}\boldsymbol{c}_{\boldsymbol{W}} = \boldsymbol{\Phi}\boldsymbol{c}. \tag{4}$$

These coefficient–decoder pairs represent the same field in different bases, and Table 4 sorts quantities by whether they are the same for every pair.

Coefficient distances need not measure differences between decoded fields, because the decoded modes are neither normalized nor orthogonal. We analyse the representation in the *field metric*. We compare decoded corrections after subtracting the prescribed parameter-dependent lift and the fixed mean. Let $\boldsymbol{G} = \boldsymbol{\Phi}^{\top}\boldsymbol{M}\boldsymbol{\Phi}$ be the decoder Gram matrix, with $\boldsymbol{M}$ the spatial mass matrix, the identity on the uniform velocity grid. The *field-orthonormal coordinates* $\boldsymbol{d}(p) = \boldsymbol{G}^{1/2}\boldsymbol{c}(p)$ decode through $\boldsymbol{\Phi}\boldsymbol{G}^{-1/2}$, whose columns are orthonormal, so $\|\boldsymbol{d}(p) - \boldsymbol{d}(p')\|_2 = \|\boldsymbol{\Phi}(\boldsymbol{c}(p) - \boldsymbol{c}(p'))\|_{\boldsymbol{M}}$. Under Equation (4) $\boldsymbol{d}_{\boldsymbol{W}} = \boldsymbol{W}\boldsymbol{d}$, and every invertible reparameterization $\boldsymbol{c} \mapsto \boldsymbol{A}\boldsymbol{c}$, $\boldsymbol{\Phi} \mapsto \boldsymbol{\Phi}\boldsymbol{A}^{-1}$ gives the same field-orthonormal coordinates up to an orthogonal matrix (Proposition 3). Distances in the field metric are the same for every such pair, and orthogonal blocks of field-orthonormal coordinates decode to orthogonal field subspaces.

Field varimax supplies a basis of labelled axes. It applies varimax to the main-effect loadings of the model's field-orthonormal coefficient table on the parameter grid (Appendix D.3). Functional analysis of variance (ANOVA) labels an axis Reynolds or asymmetry when that main effect accounts for more than half its field energy, and *interaction* otherwise. On every run, interaction-labelled axes carry most of their energy in the interaction component (Table 7). Every retained run yields two Reynolds axes, two asymmetry axes and one interaction axis. Their index sets I_g define blocks of dimensions $2/2/1$, and the architecture fixes only the total rank. We call their labelled decoded spans the *identified subspaces*. Physical validation, projected response fields and block-shared completion use these subspaces only through their projectors, which are unchanged by internal rotations in $O(2) \times O(2) \times O(1)$, where $\times$ denotes a direct product. The varimax axes span these subspaces and carry the labels. A second rule, *operator-response selection*, rotates the varimax basis to maximize the share of the model's field derivatives captured by each parameter's block (Appendix L).

We audit public checkpoints as provided and train cavity operators from PDE residuals. We used eight *development* operators, named by training seed, to design the protocol. Seven are *retained*: the smallest singular component of their centred stored coefficient table carries at least 2% of its total energy. This criterion uses no reference solutions or errors; s7 fails it (Table 7). Four *confirmatory* operators (s11, s23, s99, s65537) were trained after the protocol and criteria were fixed and appear only in the replication study (Appendix J.3). *Model-only* identification uses only the model's coefficients and decoder and is fixed before reading reference solutions (Appendix E.2). The *reference-initialized* protocol constructs the varimax basis and labels from full-grid reference projections (Table 3) and is used only for the invariance intervention. We set the screening threshold and adopted the field metric after initial results were known. Appendices C and L document the study sequence and registered results.

2.4 PHYSICAL VALIDATION

We validate each block's parameter label by measuring how much of that parameter's effect on the reference solutions lies in the block's decoded subspace. The reference-initialized basis is constructed, labelled and validated on the same reference solutions, with no statistical holdout. We isolate the parameter effects by decomposing the lift-subtracted reference solutions into Reynolds, asymmetry and interaction components using ANOVA under the empirical product measure. With u denoting the lift-subtracted reference solution and bars denoting averages over the parameter grid, the components are

$$\begin{aligned} u_{Re}(Re) = \overline{u}(Re,\cdot) - \bar{u}, \qquad u_{\kappa}(\kappa) = \overline{u}(\cdot,\kappa) - \bar{u}, \\ u_{\mathrm{int}}(Re,\kappa) = u(Re,\kappa) - \bar{u} - u_{Re}(Re) - u_{\kappa}(\kappa). \end{aligned} \tag{5}$$

A block captures a component when projecting the component's snapshots onto the block's decoded subspace keeps most of their energy. Let $\boldsymbol{R}$ be the rotation of the identified axes in field-orthonormal coordinates, U_h contain the component-h snapshots and $B_g = (\boldsymbol{\Phi}\boldsymbol{G}^{-1/2}\boldsymbol{R})_{:,I_g}$ be the decoder columns of block I_g. With $\|\cdot\|_F$ the Frobenius norm, the captured fraction of component energy, called *component recall*, is

$$R_{g\leftarrow h} = \frac{\|P_g U_h\|_F^2}{\|U_h\|_F^2}, \qquad P_g = B_g(B_g^\top B_g)^{-1}B_g^\top. \tag{6}$$

Identification and validation use the same field metric. The three blocks are orthogonal in that metric, so the recalls of one component over the blocks add up to its capture by the full span.

2.5 BASIS-CONSISTENT ANALYSES OF THE REPRESENTATION

A *projected response field* is the part of a physical response captured by an identified subspace. For a lift-subtracted response field r, such as the reference asymmetry component at $\kappa = +1$, it is $P_g r$, with P_g defined in Equation (6). Its *captured energy share* is $\|P_g r\|_2^2/\|r\|_2^2$. The replacement of B_g by $B_g Q$ for any invertible Q leaves both quantities unchanged because P_g depends only on $\mathrm{span}(B_g)$. The selection of subspaces depends on the rule and the metric. Field varimax returns the same labelled decoded subspaces after an invertible reparameterization under the conditions of Proposition 4. Restarts from other bases test this numerically (Appendix K). Maps show the streamfunction ψ of the projected velocity.

An *embedding* maps the sampled coefficients $\{\boldsymbol{c}(p)\}$ to two dimensions. On a grid of parameters 41×33, we compare three t-SNE inputs, the centred stored coefficients $\boldsymbol{c} - \bar{\boldsymbol{c}}$, the per-axis standardized field-varimax coordinates and the field-orthonormal coordinates $\boldsymbol{d} - \bar{\boldsymbol{d}}$. Euclidean distances between stored coefficients are unchanged under $O(r)$ and do not use $\boldsymbol{G}$. Per-axis standardization makes distances depend on the basis, even within an identified block. Reference distances compare reference solutions with the same prescribed field subtracted. We use scikit-learn's t-SNE (Pedregosa et al., 2011) with settings in Appendix H.1.3.

Two statistics measure separate properties. *Neighbour overlap* measures basis consistency through the Jaccard index of each point's 10 nearest neighbours in two embeddings. We average over the dense grid and compare with the overlap between different initializations. *Reference-neighbour recall* measures agreement with the reference solutions. Among the 99 reference nodes, it is the fraction of each node's 5 nearest neighbours by reference distance also among its 5 nearest embedding neighbours. We average over nodes and three initializations. The *rank correlation of distances* is the Spearman correlation between pairwise input or embedding distances and reference distances at these nodes.

2.6 COMPATIBLE PREDICTION WITHIN THE LEARNED SPAN

A downstream predictor can depend on the basis used to represent each identified subspace. We fit kernel ridge regression to the field-orthonormal coordinates of each fixed identified subspace. All coordinates within a subspace share one kernel and one tuning criterion. The fitted coefficients rotate with the internal basis, and decoding cancels the rotation, leaving predictions unchanged. Appendix D.1 gives sufficient conditions (Observation 1, Lemma 1). In the *completion task*, the predictor uses reference solutions on one row and one column of the parameter grid to predict the remaining nodes. Their reference solutions enter only the evaluation (Appendix E.1).

3 RESULTS

3.1 BASIS FREEDOM OF TRAINED REPRESENTATIONS

The freedom a representation admits follows the computation the public code implements, from the layer that produces a feature to every layer that reads it (Figure 2, Table 14). In three cases that computation differs from the architecture's description. On every public checkpoint we audited, the paired stored-weight tests separate compensated reparameterizations from their controls (Figure 2(a)). At every feature location but the last Fourier block, the group that the layer shapes suggest changes the output. The checkpoints, all purely data-driven, are Fourier neural operators (FNO), Tucker-factorized FNOs (TFNO), classic U-Nets (Ronneberger et al., 2015) and ConvNeXt U-Nets (Liu et al., 2022) from The Well (Ohana et al., 2024; Polymathic AI, 2025), and branch–trunk operators from a flow benchmark (Tali et al., 2025; Rabeh et al., 2025; 2026; BGLab, 2025). Most tests act on feature representations, and the rest on internal Tucker factors and an unread bias. The audit reads only the implemented computation, and the physics-informed cavity operators pass the same paired test as these purely data-driven models (Appendix E.4.1).

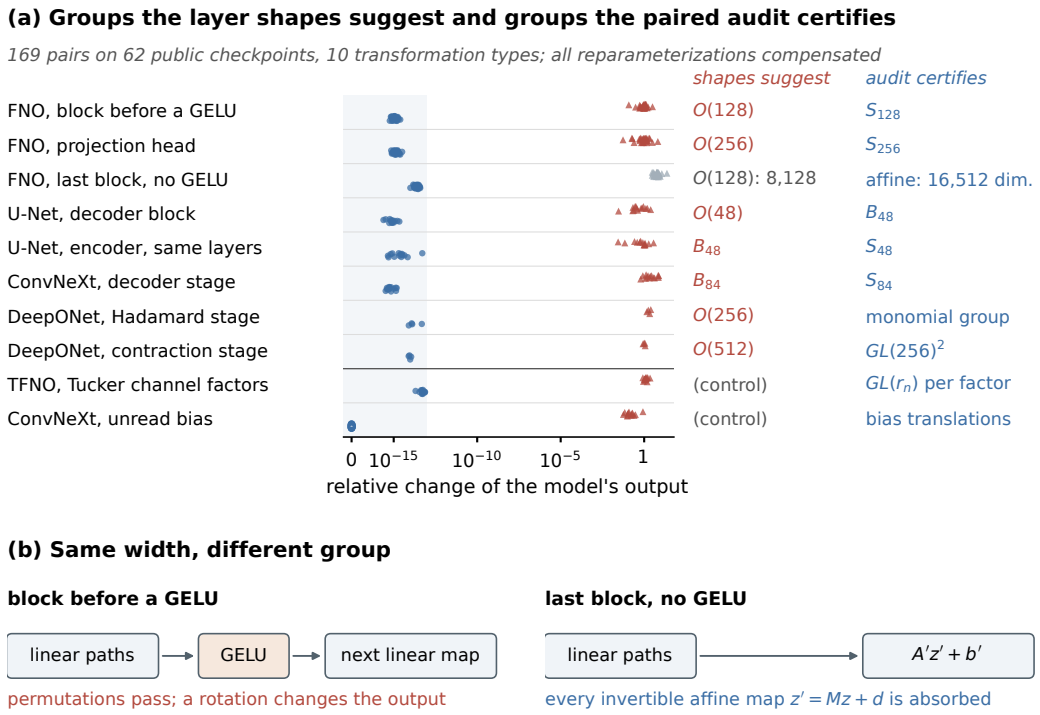


Figure 2: Paired weight reparameterizations identify checkpoint-specific basis freedom. **(a)** Output deviation (Equation (3)) on every public checkpoint, one row per transformation type. Red triangles apply the group suggested by the layer shapes or by a sibling block, blue circles the group the paired audit certifies, both with compensation. $O(n)$, S_n and B_n denote rotations, permutations and signed permutations of n channels. At the projection head the red series is the sampled rotation control, and its signed-permutation subgroup (Table 14) also changes the output. At the last Fourier block the suggested rotations are admissible, and the certified affine family has about twice their continuous degrees of freedom. The grey triangles apply the same reparameterization one block earlier. In the last two rows the triangles are the recorded controls, and open circles are exact zeros. **(b)** Two blocks of equal width admit different groups. After the earlier block a GELU intervenes and the certified permutation subgroup is discrete. With no activation after the last block, the next linear map absorbs every invertible affine channel map.

The last Fourier block admits every invertible affine channel map within the declared reparameterization class. The public model omits the optional per-block multilayer perceptron, which leaves this block without an activation, and the audit certifies the resulting family numerically (Figure 2(b)). An intervening activation changes the result one block earlier. There a compensated unsigned permutation leaves the output unchanged at numerical precision, and a rotation changes it. The audit certifies this permutation subgroup without showing that no larger group exists (Appendix G). Two

blocks of equal width support different interpretations. Per-channel statistics at the earlier block are preserved up to channel relabelling under the certified permutation subgroup. At the last block they depend on a choice among a continuous family of bases.

The paired control also exposes cases that a small output deviation cannot resolve. In one U-Net, 28 of the 48 readout features are numerically inactive, with normalization scales below 10^{-13}. Perturbing their readout rows leaves the output unchanged at numerical precision, and the same perturbation on the active rows changes it substantially (Figure 6). This contrast distinguishes numerical inactivity from an architectural symmetry at the tested tolerance (Appendix G). On a real snapshot, reparameterizing the inactive rows moves those channels to the top of a readout-weight importance score and leaves an activity-weighted score unchanged. In ConvNeXt, inspection identifies an additional normalization rule: one layer class implements two channel conventions, so decoder normalization admits unsigned permutations while resampling leaves a bias unread.

3.2 Physical interpretation of identified subspaces

Identification in the field metric separates the reference responses that identification in the stored coefficients mixes, and it returns closely aligned subspaces across retained runs (Figure 3).

Blocks identified in the stored coefficients overlap in the decoded field, so one reference response is counted in several of them. Blocks of field-orthonormal coordinates decode to orthogonal subspaces and count it once (Figure 3(a), Appendix H.1.1).

The field-varimax subspaces agree across training runs to within a few degrees in every block. The coefficient-metric Reynolds and interaction blocks differ by up to tens of degrees (Figure 3(b), Appendix J.1). The four confirmatory operators reproduce the parameter-aligned block allocation, the asymmetry recall and the close subspace alignment (Appendix J.3). The identification rule is a choice separate from the metric. Field orthonormalization makes the decoded blocks of any orthogonal basis mutually orthogonal, and SVD axes in the field metric agree across runs as closely as the varimax blocks do. The rule decides the allocation and how much of each reference component a block holds. Field varimax recalls more of the Reynolds component than SVD does (Table 18). Naming an axis inside a subspace needs a further rule. The two field-varimax asymmetry axes split into a mode nearly antisymmetric about the cavity midline and a mode closer to symmetric. The antisymmetric mode carries the larger amplitude on every run (Appendix J.2). On Burgers operators, analysed after the protocol was fixed on the cavity, field varimax gives every model one viscosity axis that carries most of the exact viscosity response (Appendix I.1).

3.3 Visualizing representation geometry

Embeddings of field-orthonormal coordinates recover more reference neighbours than embeddings of standardized field-varimax coordinates, and the field-orthonormal input distances are the same in every basis (Figure 4). Table 1 scores each input on its own distances at the reference nodes before any embedding, so the comparison does not rest on t-SNE. A common visualization, t-SNE applied to per-axis standardized coordinates (Figure 4(a,b)), changes under a rotation that leaves every decoded field unchanged, even when the coordinates are the field-varimax coordinates of the identified subspaces. Standardization gives every axis unit variance. The Reynolds axes, which carry little field energy, then receive the weight of the asymmetry axes, and the distances stop following the fields.

Table 1: Field-orthonormal coordinates combine distances that are the same in every basis with high reference-neighbour recall. Each input is scored on its own distances at the reference nodes and after t-SNE, with ranges over the seven retained runs and statistics as in §2.5. The variance ratio, the largest per-axis variance over the smallest, is taken before standardization. The rotated row uses eight bases per run.

input	input recall	input correlation	variance ratio	embedding recall
stored coefficients	0.66–0.71	0.76–0.79		0.59–0.64
standardized, varimax basis	0.54–0.55	0.65	330–344	0.50–0.51
standardized, rotated bases	0.69–0.95		7–134	0.67–0.94
field-orthonormal	0.94	0.94		0.94
block-normalized	0.51	0.63		0.49

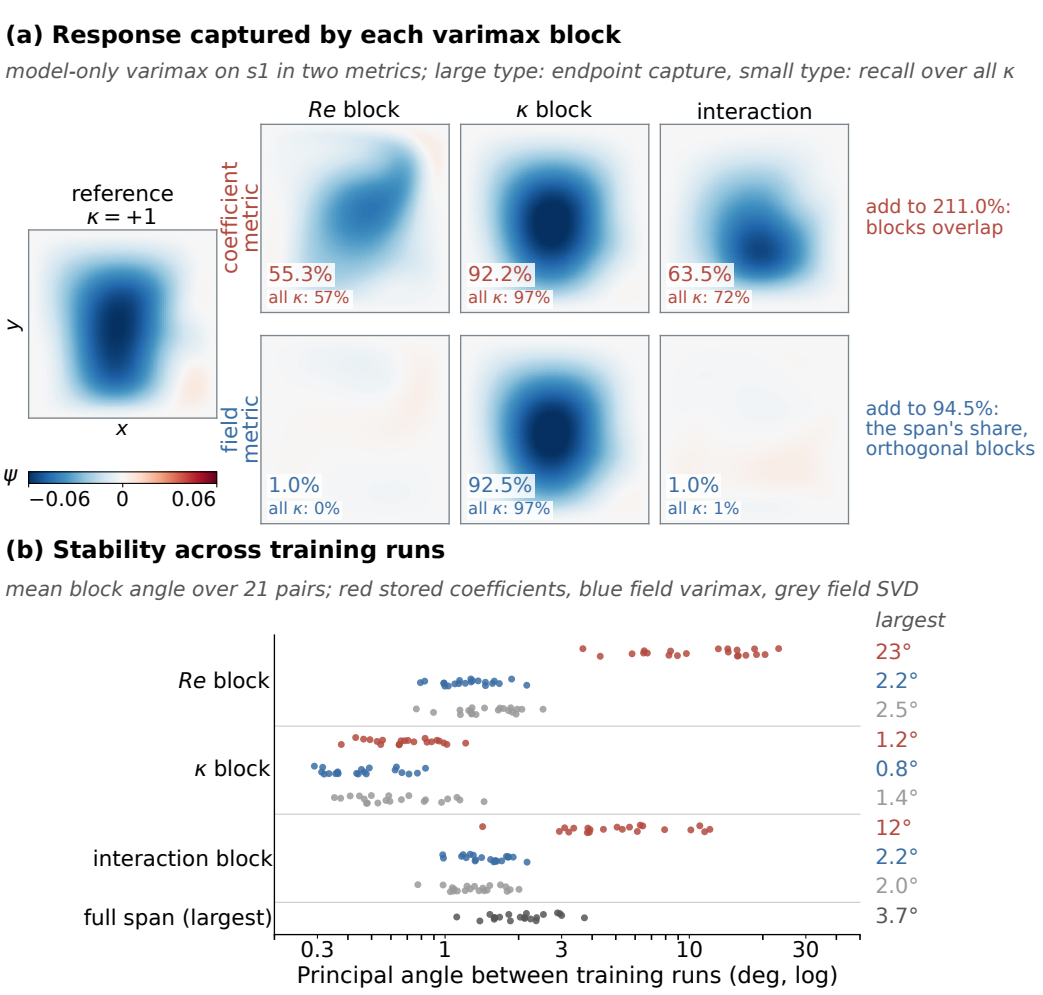


Figure 3: Identification in the field metric separates the asymmetry response and reproduces across training runs. **(a)** Model-only varimax on s1. The reference asymmetry response at $\kappa = +1$ and its projection on each block (streamfunction ψ, one colour scale), for varimax in the coefficient metric (upper row) and in the field metric (lower row). Large percentages give the *endpoint capture*, the captured energy share (§2.5) of this one response, and each row ends with their sum. Small percentages give component recall (Equation (6)). **(b)** Mean principal angle between each block of two training runs for varimax in either metric and for SVD axes in the field metric. The last row is the largest angle between full spans, which the metric does not change.

A basis rotation moves a field-orthonormal or stored-coefficient layout less than a new t-SNE initialization does. Standardized coordinates change their neighbourhoods under rotations inside the identified blocks (Appendix H.1.3). The ordering of the inputs holds on all seven retained runs, for PCA, metric MDS, three perplexities and the Burgers operators (Table 21). Field-orthonormal distances are invariant under every invertible reparameterization (Proposition 3). Each strand holds one value of κ, and Re varies along it. Scaling each identified block to unit mean square spreads the Reynolds variation across the layout and loses the reference neighbours (Table 1).

3.4 Reusing the identified subspaces

Block-shared fitting gives predictions that are the same for every basis chosen inside the identified subspaces. The test holds the identified subspaces fixed, rotates their internal bases and repeats fitting and tuning on all eight checkpoints. Per-coordinate tuning changes the prediction by about 11–12%, and block sharing reduces the change to numerical precision (Appendix E.4.2). Features of other trained models behave alike (Appendix F). Completion from one observed row and column of

reference solutions within the identified subspaces improves on the original operator, which observes none, and the complete comparison includes a more accurate full-span interpolant (Appendix E.3).

4 Discussion and Limitations

Physical structure can remain interpretable even when individual axis labels depend on the basis. An axis statistic, such as a disentanglement score, describes the chosen basis (Appendix H.5). Basis freedom varies across feature locations, so interpretations must respect the transformations certified at the location analysed. Our audit tests this freedom without requiring a linear decoder or reference solution. For operators with linear decoders, the reference validation links identified subspaces to physical responses. Projections onto fixed decoded subspaces remain unchanged under internal basis changes, and a subspace's physical label rests on component recall. Invariant field distances and block-shared fitting extend this consistency to visualization and reuse.

The metric defines physical similarity within a basis-consistent analysis. In the cavity, larger variations in asymmetry dominate field distances. Block normalization makes Reynolds-number variation more visible and changes which states appear as neighbours. Using the same field metric for identification, validation and visualization ties these analyses to a common physical comparison.

The audit requires weights, forward code, and analyst-derived candidates and controls. Physical subspace and embedding evidence come from low-rank cavity and Burgers operators with linear decoders. The learned span limits the responses that can be captured. Subspace identification depends on the rule, metric and parameter measure, with reproducibility subject to stated optimization conditions. The cavity and Burgers operators saw no reference solution in training, so their validation is independent. A purely data-driven operator needs held-out reference solutions for the same check. Without reference solutions, physical labels remain hypotheses. Embedding invariance concerns input distances, and layouts vary with initialization.

Appendix A.2 details candidate derivation. Proposition 4, Tables 8 and 9, and Appendix K.1 give the reproducibility conditions and checks. Appendix H.2 gives the decoded-rank diagnostics. Appendix A.4 records the development history, and Table 6 the evidence and scope of each claim.

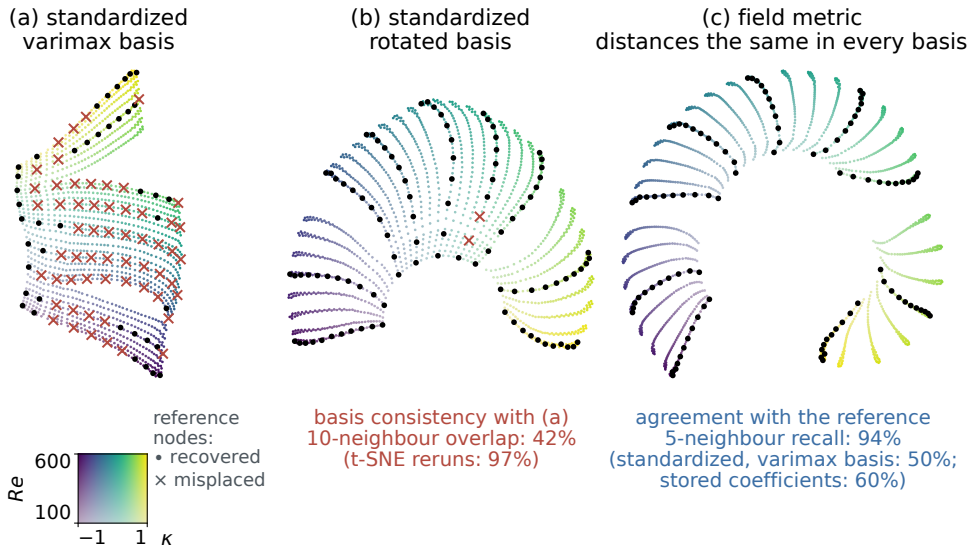


Figure 4: Embeddings of field-orthonormal coordinates use input distances that are the same in every basis and recover more reference neighbours than embeddings of standardized field-varimax coordinates. The panels show t-SNE of operator s1's coefficients on the dense grid. Panel **(a)** uses standardized field-varimax coordinates, panel **(b)** the same standardization in a rotated basis that decodes the same fields, and panel **(c)** field-orthonormal coordinates. Markers are the 99 reference nodes. A node is recovered (dot) when the panel keeps at least three of its five reference neighbours, and misplaced (cross) otherwise. Bracketed values give the same statistic for the standardized and stored inputs.

Artificial intelligence (AI) use statement

The theoretical framing, proofs, and study design originated with the authors. Claude (Anthropic) and ChatGPT (OpenAI) assisted with literature searches, cross-checking mathematical arguments and references, manuscript editing, and code implementation. These tools also supported validation planning and analysis checks. The authors remain responsible for reviewing and verifying the final scientific content, numerical evidence, and conclusions before submission.

Ethics statement

The audit reads public checkpoints from their original hosts. It pins each checkpoint by revision and verifies its digest. The findings concern the public weights and forward implementations. One supplementary example reads a public validation file of The Well (CC-BY-4.0), pinned by revision and digest. The study involves no human subjects or personal data and introduces no new dataset.

Reproducibility statement

The anonymous reproduction package is available for review at `https://anonymous.4open.science/r/iclr_demo-4C6C/`. It includes the audit and training code, the configurations and the 99 reference solutions used for evaluation. It ships the twelve cavity checkpoints we trained, the eight development operators and the four confirmatory operators. Operators are trained from PDE and boundary residuals without reference solutions. Reference solutions enter post-training selection in the reference-initialized protocol, physical validation, completion fitting, controls and evaluation; model-only selection is frozen before any reference solution is read. Appendix C documents training, seeds, and evaluation grids. The package re-derives artifact-backed numerical claims offline, reruns checkpoint-only certificates on public checkpoints, and supports cavity checkpoint regeneration and retraining. Public checkpoints are fetched from their original hosts, pinned by revision and verified by digest; they are not redistributed in the package.

References

Samuel K. Ainsworth, Jonathan Hayase, and Siddhartha Srinivasa. Git Re-Basin: Merging models modulo permutation symmetries. In *International Conference on Learning Representations*, 2023.

Igor A. Baratta, Joseph P. Dean, Jørgen S. Dokken, Michal Habera, Jack S. Hale, Chris N. Richardson, Marie E. Rognes, Matthew W. Scroggs, Nathan Sime, and Garth N. Wells. DOLFINx: The next generation FEniCS problem solving environment. Zenodo, `doi:10.5281/zenodo.10447666`, 2023.

Gal Berkooz, Philip Holmes, and John L. Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Annual Review of Fluid Mechanics*, 25(1):539–575, 1993.

BGLab. Released DeepONet checkpoints for the FlowBench flow-past-object benchmark. Hugging Face model repository `BGLab/DeepONet-FlowBench-FPO`, MIT licence, revision `1c09f5dc7efb748cbe2481c5f9a6db5feaf3326f`, 2025. URL `https://huggingface.co/BGLab/DeepONet-FlowBench-FPO/tree/1c09f5dc7efb748cbe2481c5f9a6db5feaf3326f`. The model card at this revision cites the preprint Rabeh et al. (2025) and links the FlowBench dataset (Tali et al., 2025). The journal version is Rabeh et al. (2026).

Kaushik Bhattacharya, Bamdad Hosseini, Nikola B. Kovachki, and Andrew M. Stuart. Model reduction and neural networks for parametric PDEs. *The SMAI Journal of Computational Mathematics*, 7:121–157, 2021.

Trenton Bricken, Adly Templeton, Joshua Batson, Brian Chen, Adam Jermyn, Tom Conerly, Nicholas L. Turner, Cem Anil, Carson Denison, Amanda Askell, Robert Lasenby, Yifan Wu, Shauna Kravec, Nicholas Schiefer, Tim Maxwell, Nicholas Joseph, Alex Tamkin, Karina Nguyen, Brayden McLean, Josiah E. Burke, Tristan Hume, Shan Carter, Tom Henighan, and Christopher Olah. Towards monosemanticity: decomposing language models with dictionary learning. Transformer Circuits Thread, 2023. URL `https://transformer-circuits.pub/2023/monosemantic-features/index.html`.

Jericho Cain. Gauge freedom and metric dependence in neural representation spaces. arXiv preprint arXiv:2603.06774, 2026.

Ricky T.Q. Chen, Xuechen Li, Roger Grosse, and David Duvenaud. Isolating sources of disentanglement in variational autoencoders. In *Advances in Neural Information Processing Systems*, 2018.

Francisco Chinesta, Pierre Ladevèze, and Elías Cueto. A short review on model order reduction based on proper generalized decomposition. *Archives of Computational Methods in Engineering*, 18(4):395–404, 2011. doi: 10.1007/s11831-011-9064-7.

Albert Cohen and Ronald DeVore. Approximation of high-dimensional parametric PDEs. *Acta Numerica*, 24:1–159, 2015.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International Conference on Machine Learning*, 2016.

Julian D. Cole. On a quasi-linear parabolic equation occurring in aerodynamics. *Quarterly of Applied Mathematics*, 9(3):225–236, 1951.

Hoagy Cunningham, Aidan Ewart, Logan Riggs, Robert Huben, and Lee Sharkey. Sparse autoencoders find highly interpretable features in language models. In *International Conference on Learning Representations*, 2024.

Cian Eastwood and Christopher K.I. Williams. A framework for the quantitative evaluation of disentangled representations. In *International Conference on Learning Representations*, 2018.

Hamidreza Eivazi, Soledad Le Clainche, Sergio Hoyas, and Ricardo Vinuesa. Towards extraction of orthogonal and parsimonious non-linear modes from turbulent flows. *Expert Systems with Applications*, 202:117038, 2022.

Nelson Elhage, Robert Lasenby, and Christopher Olah. Privileged bases in the transformer residual stream. Transformer Circuits Thread, 2023. URL `https://transformer-circuits.pub/2023/privileged-basis/index.html`.

Charles Godfrey, Davis Brown, Tegan Emerson, and Henry Kvinge. On the symmetries of deep learning models and their internal representations. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Ulrike Grömping. Relative importance for linear regression in R: The package relaimpo. *Journal of Statistical Software*, 17(1):1–27, 2006. doi: 10.18637/jss.v017.i01. URL `https://www.jstatsoft.org/article/view/v017i01`.

Jan S. Hesthaven and Stefano Ubbiali. Non-intrusive reduced order modeling of nonlinear problems using neural networks. *Journal of Computational Physics*, 363:55–78, 2018.

Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. β-VAE: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations*, 2017.

Irina Higgins, David Amos, David Pfau, Sébastien Racanière, Loic Matthey, Danilo Rezende, and Alexander Lerchner. Towards a definition of disentangled representations. arXiv:1812.02230, 2018.

Eberhard Hopf. The partial differential equation $u_t + uu_x = \mu u_{xx}$. *Communications on Pure and Applied Mathematics*, 3(3):201–230, 1950. doi: 10.1002/cpa.3160030302.

Xiang Huang, Zhanhong Ye, Hongsheng Liu, Beiji Shi, Zidong Wang, Kang Yang, Yang Li, Min Wang, Haotian Chu, Fan Yu, Bei Hua, Lei Chen, and Bin Dong. Meta-Auto-Decoder for solving parametric partial differential equations. In *Advances in Neural Information Processing Systems*, volume 35, pp. 23426–23438, 2022. doi: 10.52202/068431-1702.

Aapo Hyvärinen and Petteri Pajunen. Nonlinear independent component analysis: Existence and uniqueness results. *Neural Networks*, 12(3):429–439, 1999.

Henry F. Kaiser. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23 (3):187–200, 1958.

Ilyes Khemakhem, Diederik P Kingma, Ricardo Pio Monti, and Aapo Hyvärinen. Variational autoencoders and nonlinear ICA: A unifying framework. In *International Conference on Artificial Intelligence and Statistics*, 2020.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In *International Conference on Machine Learning*, pp. 3519–3529, 2019.

Abhishek Kumar, Prasanna Sattigeri, and Avinash Balakrishnan. Variational inference of disentangled latent concepts from unlabeled observations. In *International Conference on Learning Representations*, 2018.

Sanghyun Lee and Yeonjong Shin. On the training and generalization of deep operator networks. *SIAM Journal on Scientific Computing*, 46(4):C273–C296, 2024.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. Physics-informed neural operator for learning partial differential equations. *ACM/IMS Journal of Data Science*, 1(3):1–27, 2024. doi: 10.1145/3648506. Article 9.

Selena Ling, Nicholas Sharp, and Alec Jacobson. VectorAdam for rotation equivariant geometry optimization. In *Advances in Neural Information Processing Systems*, 2022.

Ning Liu, Lu Zhang, Tian Gao, and Yue Yu. Disentangled representation learning for parametric partial differential equations. In *International Conference on Learning Representations*, 2026. arXiv:2410.02136.

Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A ConvNet for the 2020s. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Raetsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. Challenging common assumptions in the unsupervised learning of disentangled representations. In *International Conference on Machine Learning*, pp. 4114–4124, 2019.

Francesco Locatello, Ben Poole, Gunnar Rätsch, Bernhard Schölkopf, Olivier Bachem, and Michael Tschannen. Weakly-supervised disentanglement without compromises. In *International Conference on Machine Learning*, pp. 6348–6359, 2020.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George E. Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021.

Lu Lu, Xuhui Meng, Shengze Cai, Zhiping Mao, Somdatta Goswami, Zhongqiang Zhang, and George E. Karniadakis. A comprehensive and fair comparison of two neural operators (with practical extensions) based on FAIR data. *Computer Methods in Applied Mechanics and Engineering*, 393:114778, 2022.

Brek Meuris, Saad Qadeer, and Panos Stinis. Machine-learning-based spectral methods for partial differential equations. *Scientific Reports*, 13:1739, 2023.

Luca Moschella, Valentino Maiorca, Marco Fumero, Antonio Norelli, Francesco Locatello, and Emanuele Rodolà. Relative representations enable zero-shot latent space communication. In *International Conference on Learning Representations*, 2023.

Ruben Ohana, Michael McCabe, Lucas Meyer, Rudy Morel, Fruzsina J. Agocs, Miguel Beneitez, Marsha Berger, Blakesley Burkhart, Stuart B. Dalziel, Drummond B. Fielding, Daniel Fortunato, Jared A. Goldberg, Keiya Hirashima, Yan-Fei Jiang, Rich R. Kerswell, Suryanarayana Maddu, Jonah Miller, Payel Mukhopadhyay, Stefan S. Nixon, Jeff Shen, Romain Watteaux, Bruno Régaldo-Saint Blancard, François Rozet, Liam H. Parker, Miles Cranmer, and Shirley Ho. The Well: a large-scale collection of diverse physics simulations for machine learning. In *Advances in Neural Information Processing Systems*, volume 37, pp. 44989–45037, 2024. doi: 10.52202/079017-1430.

Mario Ohlberger and Stephan Rave. Reduced basis methods: Success, limitations and future challenges. In *Proceedings of ALGORITMY 2016*, pp. 1–12, 2016.

Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. DeepSDF: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 165–174, 2019.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(85): 2825–2830, 2011.

William Peebles, John Peebles, Jun-Yan Zhu, Alexei A. Efros, and Antonio Torralba. The Hessian penalty: A weak prior for unsupervised disentanglement. In *European Conference on Computer Vision (ECCV)*, 2020.

Ahmad Peyvan, Vivek Oommen, Ameya D. Jagtap, and George Em Karniadakis. RiemannONets: Interpretable neural operators for Riemann problems. *Computer Methods in Applied Mechanics and Engineering*, 426:116996, 2024. doi: 10.1016/j.cma.2024.116996.

Polymathic AI. Released FNO, TFNO, U-Net and ConvNeXt U-Net baseline checkpoints for The Well. Hugging Face model repositories `polymathic-ai/{FNO,TFNO,UNetClassic,UNetConvNext}-*`. At every audited revision the model card declares no licence and the repository holds no licence file. Revisions are recorded in `experiments/external_checkpoint/checkpoint_provenance.json`, 2025.

Christophe Prud'homme, Dimitrios V Rovas, Karen Veroy, Luc Machiels, Yvon Maday, Anthony T Patera, and Gabriel Turinici. Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods. *Journal of Fluids Engineering*, 124(1):70–80, 2002.

Ali Rabeh, Suresh Murugaiyan, Adarsh Krishnamurthy, and Baskar Ganapathysubramanian. Predicting time-dependent flow over complex geometries using operator networks. arXiv preprint arXiv:2512.04434, 2025. URL `https://arxiv.org/abs/2512.04434`.

Ali Rabeh, Suresh Murugaiyan, Adarsh Krishnamurthy, and Baskar Ganapathysubramanian. Predicting time-dependent flow over complex geometries using operator networks. *Computer Methods in Applied Mechanics and Engineering*, 456:118931, 2026. doi: 10.1016/j.cma.2026.118931.

Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

Geoffrey Roeder, Luke Metz, and Durk Kingma. On linear identifiability of learned representations. In *International Conference on Machine Learning*, pp. 9030–9039, 2021.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, 2015.

Jacob Seidman, Georgios Kissas, Paris Perdikaris, and George J. Pappas. NOMAD: Nonlinear manifold decoders for operator learning. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Louis Serrano, Lise Le Boudec, Armand Kassaï Koupaï, Thomas X. Wang, Yuan Yin, Jean-Noël Vittaut, and Patrick Gallinari. Operator learning with neural fields: Tackling PDEs on general geometries. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Rui Shu, Yining Chen, Abhishek Kumar, Stefano Ermon, and Ben Poole. Weakly supervised disentanglement with guarantees. In *International Conference on Learning Representations*, 2020.

Ronak Tali, Ali Rabeh, Cheng-Hau Yang, Mehdi Shadkhah, Samundra Karki, Abhisek Upadhyaya, Suriya Dhakshinamoorthy, Marjan Saadati, Soumik Sarkar, Adarsh Krishnamurthy, Chinmay Hegde, Aditya Balu, and Baskar Ganapathysubramanian. FlowBench: A large scale benchmark for flow simulation over complex geometries. *Data-centric Machine Learning Research*, 2(14): 1–35, 2025. URL `https://data.mlr.press/assets/pdf/v02-14.pdf`.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008.

Simone Venturi and Tiernan Casey. SVD perspectives for augmenting DeepONet flexibility and interpretability. *Computer Methods in Applied Mechanics and Engineering*, 403:115718, 2023.

Sifan Wang, Hanwen Wang, and Paris Perdikaris. Learning the solution operator of parametric partial differential equations with physics-informed DeepONets. *Science Advances*, 7(40):eabi8605, 2021.

Dominik Zietlow, Michal Rolínek, and Georg Martius. Demystifying inductive biases for (β-)VAE based architectures. In *International Conference on Machine Learning*, pp. 12945–12954, 2021.

Appendix A maps each main-text claim to the result that establishes, tests, and bounds it. The remaining supplement presents theory and core evidence, additional analyses, and implementation details. The repository documents further secondary studies, indexed in Additional analyses.

A Claims and their evidence

Table 6 links claims to evidence and limitations. The repository README contains the runnable-study catalogue and each study's reproduction command.

Table 4 classifies quantities and procedures by their invariance under basis changes. Its baselines use the finite element method (FEM) and proper orthogonal decomposition (POD).

A.1 The audit protocol in one table

Table 2 gives the audit's four stages, checks and claim levels. The checkpoint-only tier executes the first stage and, when a downstream method is specified, the third. The physics-grounded tier adds the second and fourth.

Algorithm 1 is the complete decision rule of the checkpoint-only audit, and its execution adds three details. Before step 1, the analyst traces the representation's producer and every operation that reads it. After an inconclusive pair, the analyst tests for numerical inactivity or names a stronger control. Every run retains the outcome, candidate, control, deviations, settings and provenance.

Algorithm 2 lists the model-only identification and prediction procedure of §2.3–2.6. Table 3 separates the basis source from the derivative source for the two protocols and for the response-source sensitivity analysis.

Appendices H.2 and H.5 specify the parameter measures and field metrics. Rank and training-covariance diagnostics support the physics-grounded and optimization analyses. The checkpoint-only tier requires the architectural and numerical checks (Appendix C). An auxiliary within-span state, whether a covariance, a spectrum, a metric or an anisotropic prior, must transform covariantly under the basis change and be audited jointly with the basis (Corollary 1, Appendix D.1).

Table 2: The full audit comprises four stages. Table 4 supports stage 2 by classifying quantities and procedures according to their invariance over equivalent representations.

audit stage	checks	tier	claim level
1. Certify the transformation	derive the transformations that leave the decoded output unchanged; test the compensated reparameterization and a discriminating control (Algorithm 1)	checkpoint-only	certified invariance; analytical scope stated separately
2. Identify subspaces	check rank; state the identification rule, its metric, the retained object and the transformations that preserve it; validate decoded projectors against reference responses	physics-grounded	resolved span and identified subspaces; axes only when the retained object is the basis
3. Classify the downstream method	determine whether the downstream method is sensitive to the remaining freedom; test complete-procedure equivariance (Observation 1, Appendix D.1)	checkpoint-only	downstream invariance
4. Measure predictive performance	compare matched-subspace and unrestricted baselines, including rotations within identified blocks	physics-grounded	value conditional on observations, predictor and family

Algorithm 2 Model-only subspace identification and block-shared prediction.

Require: Frozen checkpoint and admissible transformations; parameter grid; velocity grid; measure, labelling threshold, candidate kernels, ridge and tie-breaking rule

1: Evaluate the model coefficient table and the decoder Gram matrix $\boldsymbol{G}$ on the velocity grid.
2: Form field-orthonormal coordinates $\boldsymbol{d} = \boldsymbol{G}^{1/2}\boldsymbol{c}$; compute their varimax basis and ANOVA labels.
3: Form the coordinate sets I_g and the decoder blocks $B_g = (\boldsymbol{\Phi}\boldsymbol{G}^{-1/2}\boldsymbol{R})_{:,I_g}$.
4: Retain the labelled decoded subspaces; the group of internal rotations that fix their projectors is $H = \prod_g O(|I_g|)$.
5: Freeze the basis, labels and subspaces before reference-solution access.
6: Read the observed fields, subtract the prescribed field, and project onto the frozen field-orthonormal decoder.
7: Select one kernel per block by summed context-only LOO loss; fit Equation (14) (Appendix E.1).
8: Decode the predicted coefficients and add the prescribed field.

Table 3: Basis initialization and derivative source are separate choices. Field varimax uses no derivatives, so only its basis source varies. The operator-response rule uses model derivatives in both protocols. The response-source sensitivity replaces them with reference derivatives in the same field metric.

protocol	initial basis and labels	derivatives	first reference read
model-only	model coefficients	model	after identification is frozen
reference-initialized	reference projections	model	basis construction
response-source	reference projections	reference	basis construction

Table 4 includes adapted disentanglement, completeness and informativeness (DCI), mutual information gap (MIG), and separated attribute predictability (SAP) scores. Appendix H.5 specifies their evaluation conventions.

Requirements for the full physics-grounded audit. In addition to checkpoint access and a realizable compensated reparameterization, the full audit requires a resolved rank, a declared field metric and identification rule, reference responses for validation, and a declared parameter measure and downstream procedure. Internal access is necessary to apply the reparameterizations. Outputs alone do not suffice. Reference solutions enter the post-training analysis. Operator training uses physics residuals. These requirements exceed those of the checkpoint-only tier (Appendices C, H and G).

A.2 Requirements for automated audits

Evaluating reparameterizations, computing deviations and recording outcomes are mechanical operations. Tracing the forward computation, deriving the group and choosing a control require analytical judgment. Once the producer's admissible reparameterizations are specified, a rule engine can constrain the candidate group in step 1 of Algorithm 1. Each consumer admits a set of trans-

Table 4: The quantities and procedures have different invariance properties under admissible reparameterizations. These properties distinguish field quantities, basis statistics and basis-dependent predictions. "Within blocks" denotes invariance under the rotations inside the identified blocks. The last column is the cavity instance of each row, ordered from invariant at the top to basis-dependent at the bottom.

quantity or procedure	$O(r)$-invariant	measured property	cavity example
decoded field; PDE residual; FEM error	yes	the represented physics	decoded family; PDE and FEM error
shared-kernel prediction	yes	equivariant downstream output	spread below 10^{-16} over 400 rotations of the whole basis on all eight runs (rotation landscape, Table 5)
span projection error; whole-span capture of a physical response	yes	quality of the represented span	whole span, capture $94/98/68\%$, matched by POD
capture by labelled factor projectors	within blocks	rule- and metric-relative physical alignment	varimax $78/97/19\%$ (stored coefficients), $74/97/41\%$ (field)
derivative alignment	within blocks	basis projectors under a tangent metric and allocation	84% (field varimax) versus 90% (physical-response rule)
blockwise completion	within blocks	prediction from identified blocks	field varimax 0.138 against the original operator's 0.148
variance-based purity; DCI/MIG/SAP	no	basis under a parameter measure	field varimax purity 89%; DCI 0.96 (0.79 random)
per-coordinate completion	no	coordinate-dependent prediction	individual axes inside blocks: no intrinsic claim
preprocessing and rendering before an embedding	stored coefficients or ridge-whitened: yes; per-axis standardized: no; renderers: conditional	preprocessing–basis interaction	–

formations (Godfrey et al., 2022), and the intersection over consumers gives the available group. Activation parity and positive homogeneity determine whether sign flips and positive scalings can pass through a nonlinear operation.

The specified rules return the certified transformation at all 10 audited locations, including the two U-Net locations (Appendix G). The consumer chains were written after the audit, and the fifth family required one additional rule. This test establishes the sufficiency of the specified rules. Automatic discovery would additionally require access to a supported computation graph for the tracing that precedes that step. A weight file alone does not provide that graph. Candidate specification remains an analyst task.

A.3 Studies in the repository

Table 5 maps each study that the supplement draws on to its runnable example under `examples/` in the repository, where protocols, records and results live. Four of them supported no claim: steady-Burgers training, a warm-start comparison, decorrelation training and a length-scale diagnostic.

A.4 Scope, development history and secondary limitations

The registered coefficient-metric protocol (Appendix L), the held-out ConvNeXt family (Appendix G) and the restart thresholds (Appendix K.1) were fixed in advance. The field metric and field varimax as the primary rule were adopted after first results on the eight development operators were known, and the screening criterion was fixed after the completion errors of s7 had been computed (Appendix C). We trained the four confirmatory operators after these choices and fixed their analysis before they existed (Appendix J.3). The tables report the development and confirmatory operators separately. The Burgers operators were analysed after those choices (Appendix I). The audit runs in seconds to minutes per checkpoint, and deriving its candidates is analyst work (Appendix G). The physical analysis uses rank-5 operators; Appendix H.4 examines rank through truncation and a reference POD, without retraining. The embedding comparison covers one family of inputs under five embeddings on two PDE families (Appendix H.1.3), and it concerns local neighbourhoods.

Table 5: Studies and their runnable examples in the repository.

study	example
audit execution, thresholds, timing and the rule check of Appendix A.2	`audit_execution`
channel-importance pictures on a public U-Net	`importance_pictures`
decoder reparameterizations for nonlinear decoders	`decoder_architectures`
model-only identification, frozen records	`reference_free_selection`
completion accuracy, prediction invariance, reconstruction commands	`targeted_prediction`
restart protocol and thresholds	`projector_identifiability`
rotation landscape: shared and per-coordinate kernels under rotations of the whole basis	`gauge_landscape`
branched coefficient map, repeated audit	`branched_coefficient_map`
second trunk architecture, repeated comparison	`architecture_twin`
latent regression against projected field prediction; reduced solves	`projection_control`, `reduced_solve_equivariance`
embeddings: pictures, preprocessing and rendering	`tsne_pictures`, `visualization_preprocessing`
derivative alignment: local consistency and allocation	`local_attribution_consistency`, `allocation_robustness`
Allen–Cahn reference solve and control	`allen_cahn_control`
field-metric spectrum of the reference family	`field_spectrum`
disentanglement scores over equivalent bases; s7; mode perturbations	`orbit_functionals`, `degenerate_basis_regularity`, `gauge_intervention`
Burgers: completion protocols, rank sweep, interaction share	`completion_protocols`, `burgers_rank_ladder`, `burgers_interaction_share`
training under alternative parameterizations; extrapolation in the Reynolds number	`gauge_covariance_training`, `re_extrapolation`
architecture ablations; nonseparable model; non-PDE family	`ablations`, `third_model_class`, `off_pde_family`
studies that supported no claim	`steady_burgers_sensitivity`, `warm_start_null`, `decorrelation_cotraining`, `lengthscale_mismatch`

A.5 Claims, evidence and limitations

Table 6 lists each claim of the main text with its evidence and its limitation.

Table 6: The evidence supports each main-text claim within its stated scope. Claims 1–6 concern interpretation and visualization. Claims 7–8 concern the supporting prediction application. Evidence is of three kinds: analytical transformation families, numerical certificates at declared inputs and tolerances, and comparisons against reference solutions. Additional controls are described with their corresponding experiments. The runnable-study catalogue is in the repository.

claim	evidence	limitation
1. The checkpoint audit gives numerical evidence, at the declared inputs and tolerances, for the derived admissible transformations.	169 pairs on 62 checkpoints from five families: reparameterization deviations $< 10^{-13}$, controls ≥ 0.027; Table 14.	Maximality is analytical and class-relative. The unread-bias result concerns additive reparameterizations. The count keeps the channel-factor Tucker certificates; sampled all-factor reparameterizations clear the same tolerance with well-conditioned maps (Appendix G).
2. Paired controls identify inactive features and distinguish implemented channel conventions.	$28/48$ inactive U-Net features; active control deviation 2.2; activity-weighted channel importance unchanged by inactive-row reparameterizations (Figure 7); two ConvNeXt normalization branches, §3.1.	A pair with both deviations below threshold is inconclusive; inactive features and weak controls can both cause this outcome.
3. Field varimax supplies a basis of labelled axes from the model alone; the identified subspaces are the labelled decoded subspaces, and the basis used inside them downstream remains open.	Algorithm 2; P_g depends only on $\mathrm{span}(B_g)$ (§2.5); internal-rotation and restart studies, Appendices E.4.2 and K.1. Proposition 2 applies to operator-response selection, whose objective is invariant under internal rotations.	Basis reproducibility is conditional (Proposition 4); symmetry alone does not establish a unique projector tuple or physical alignment.
4. The identified subspaces support a physical interpretation.	Asymmetry recall about 97% on every retained run; blocks orthogonal in the decoded field, where stored-coefficient blocks overlap; mean cross-run principal angles at most 2.2° (Figure 3; Tables 17 and 29; Appendix J); replicated on four operators trained after the protocol was fixed (Table 26).	Model-only and reference-initialized varimax bases differ by up to 19° (Table 29); the interaction block recalls 41–42% of its component.
5. Projected response fields do not depend on the basis chosen inside a fixed decoded subspace.	P_g depends only on $\mathrm{span}(B_g)$ (§2.5); endpoint maps and captured energy shares, Figure 3(a) and Appendix H.1.1.	Basis independence holds for a fixed subspace; which subspace is identified depends on the rule and the metric. Endpoint shares at $\kappa = +1$ differ from component recall.

claim	evidence	limitation
6. Embeddings computed from field-orthonormal coordinates use input distances that are the same in every basis and preserve more reference neighbours than standardized field-varimax coordinates.	Proposition 3 (§2.3; any invertible reparameterization, Appendix D.5); distances agree across bases to below 10^{-8}; reference 5-neighbour recall 0.94 against 0.50–0.51 standardized field-varimax and 0.59–0.64 stored coefficients on seven of seven runs; standardized neighbourhoods change under rotations inside identified blocks; displayed layouts recompute exactly (Table 19); the same ordering of the inputs before any embedding, for PCA, metric MDS and three perplexities, and on eleven Burgers operators (Tables 1 and 21).	The t-SNE layout at perplexity 30 follows reference distances over all pairs less closely than the stored-coefficient input's layout (0.75 against 0.84–0.87); the input distances and the PCA, MDS and perplexity-50 layouts follow them more closely (Tables 1 and 21). Layouts vary between initializations; no dimension or topology claim.
Supporting application: reuse of the identified subspaces		
7. Block-shared fitting and selection give invariant decoded predictions within identified subspaces.	Observation 1, Lemma 1; fixed-subspace intervention: direct field deviation 0.110–0.124 with per-coordinate tuning, $< 5 \times 10^{-16}$ after block sharing (§3.4; Appendix E.4.2).	Equivariance must cover preprocessing, fitting and selection. It establishes basis independence; prediction accuracy requires a separate comparison.
8. Completion within model-only field-varimax subspaces improves on the original operator from 19 observed fields.	Field varimax 0.138, original operator 0.148, matched full-span Gaussian 0.191; original minus field varimax $+0.0102$ $[0.0092, 0.0111]$ on $7/7$ runs and $8/8$ with s7 (Tables 10 and 11); identified blocks beat nearly all dimension-matched random subspaces (Appendix E.4.2).	Full-span thin-plate interpolation gives 0.125; operator-response selection reaches the same accuracy as field varimax (Table 10).

B Extended related work

Operator representations and basis selection. Operator learning includes branch–trunk (Lu et al., 2021), spectral (Li et al., 2021; 2024) and physics-informed (Wang et al., 2021) constructions. Our branch-free model directly optimizes latent coefficients, following the auto-decoder approach (Park et al., 2019; Huang et al., 2022; Serrano et al., 2023). A fixed basis with a learned coefficient map is also used by POD neural networks (Hesthaven & Ubbiali, 2018), networks based on principal component analysis (Bhattacharya et al., 2021) and POD-DeepONet (Lu et al., 2022). These snapshot methods obtain their bases from reference solutions. Our basis is learned from the governing residuals.

Basis construction does not automatically resolve orientation. A branch–trunk contraction $\sum_k b_k T_k$, with branch coefficients b_k and trunk features T_k, is invariant under $T \mapsto AT$, $b \mapsto A^{-\top} b$ for invertible A; $A^{-\top} = (A^{-1})^{\top}$. With the latent metric and surrounding representation fixed, orthonormality restricts this coefficient–basis reparameterization to $O(r)$. SVD of the training data (Venturi & Casey, 2023), hierarchical orthonormalization of trained trunk functions for spectral solvers (Meuris et al., 2023) and two-step training with Gram–Schmidt orthonormalization, which improves stability and generalization (Lee & Shin, 2024), give the basis additional structure. SVD-based representatives used for physical interpretation (Peyvan et al., 2024) are unique only up to signs when singular values are distinct. Repeated singular values retain subspace freedom. Our audit makes these residual ambiguities explicit and checks whether the downstream procedure respects them. Appendix H.1.2 applies SVD mode extraction to our trained operators and compares its axes with field varimax. We did not train with SVD or two-step orthonormalization. The audit applies to such a model once it is trained. The linear solution span still has approximation limits governed by Kolmogorov widths (Seidman et al., 2022), as in classical reduced-order modeling (Berkooz et al., 1993; Cohen & DeVore, 2015; Ohlberger & Rave, 2016; Chinesta et al., 2011; Prud'homme et al., 2002).

Disentanglement and physical factors. DisentangO (Liu et al., 2026), the closest PDE-specific work, combines a multi-task neural operator with a variational model of task-dependent operator parameters to recover latent physical factors. Under auxiliary-variable assumptions related to nonlinear independent component analysis (ICA) (Khemakhem et al., 2020), it identifies these factors up to permutation and component-wise invertible transformations. DisentangO learns its factors during supervised multi-task training. We study a different object, the freedom that remains in a checkpoint after training. The audit can be run on a DisentangO model as on any other checkpoint,

and it would test which transformations of the learned factors the implementation admits. A rotation after training changes a checkpoint's coordinates and preserves its decoded fields. An added training objective can change the learned function. In our one-parameter operators a decorrelation term raised the error (decorrelation study, Table 5).

This distinction complements the statistical literature. Unsupervised disentanglement requires inductive bias (Locatello et al., 2019). Suitable supervision can restore identifiability under additional assumptions (Shu et al., 2020; Locatello et al., 2020). Variational-autoencoder objectives (Higgins et al., 2017; Eivazi et al., 2022) exploit particular inductive biases. Zietlow et al. (2021) show that dataset perturbations preserving generative factors can disrupt the variance structure on which those biases rely. Group-based formulations connect factors to equivariance (Higgins et al., 2018), and the Hessian penalty penalizes mixed second derivatives of the generator (Peebles et al., 2020). General nonlinear ICA remains non-identifiable without further structure (Hyvärinen & Pajunen, 1999; Khemakhem et al., 2020). Other model classes admit identifiability up to a linear map (Roeder et al., 2021). Physical response fields provide a reference for comparing variance-based scores, derivative alignment and downstream error.

Symmetry analysis and representation comparison. Godfrey et al. (2022) derive intertwiner groups linking a network's nonlinearities to output-preserving transformations of its internal representations. Their analysis supplies the architectural foundation for the checkpoint audit. Cain (2026) studies general linear changes of basis at a linear readout and the resulting metric dependence of representation geometry. Compensating the readout preserves predictions, and cosine similarities and nearest neighbours can still change. Elhage et al. (2023) trained a transformer whose LayerNorm was modified to be basis-independent, with no mean subtraction and a single learnable scale. They then multiplied every matrix that reads from the residual stream by a random orthogonal matrix R and every matrix that writes to it by $R^\top$. The test loss changed far less than it varies across initialization seeds. The checkpoint audit applies the same compensated reparameterization to trained operators. It adds paired controls that must change the output, an analysis of every consumer that reads the representation and a diagnostic for inactive features. Subspace identification adds validation against physical responses and retuning of the downstream predictors.

Several approaches construct invariance or account for it in comparisons: relative representations (Moschella et al., 2023), equivariant architectures (Cohen & Welling, 2016), orthogonally invariant similarity measures such as centered kernel alignment (Kornblith et al., 2019), and permutation alignment through re-basin methods (Ainsworth et al., 2023). We use the relevant symmetry as a constraint on an interpretation claim. The audit uses checkpoint-specific controls to separate implemented symmetry from numerical inactivity. Subspace identification then determines the remaining transformations that a compatible predictor should respect. The cavity intervention tests that design, and the matched completion comparisons assess its predictive use. The mathematical results formalize the distinctions needed by this procedure. Observation 1 states a standard equivariance identity, Proposition 1 distinguishes variance-based purity from derivative alignment, and Proposition 2 relates the rotations an identification rule leaves free to the downstream method's invariance.

Sparse dictionaries and feature interpretations. Sparse autoencoders learn overcomplete dictionaries whose features are often interpreted individually (Cunningham et al., 2024; Bricken et al., 2023). A sparsity objective distinguishes orientations, and independent validation supplies evidence for a feature's meaning. An audit must identify the transformations admitted by the forward computation, the freedoms resolved by identification, and the reference evidence supporting the feature labels. A paired control should change the output. A collapsed-feature diagnostic can identify numerically inactive directions. The physical-response tests used here require a suitable external referent and therefore do not transfer directly to language-model features.

C Training and evaluation protocol

Interpretation of the intervals. Per-basis intervals resample the 63 interior cross centers. These centers form a census of the chosen grid, and any two tasks share at least 63 of their 80 target cells. We therefore interpret their bootstrap intervals as descriptive variation across this set of tasks. They do not estimate uncertainty over independently sampled test populations.

Seed-level summaries use the paired seed means, a one-sided sign test and a t interval over the seven retained rank-5 runs. These describe training variation within one architecture. The full spans agree to 1.1–3.7° across all 21 pairs and interpolated u_{rel} agrees to 0.001. The stability of the identified blocks is reported in Appendix J. Neither interval represents uncertainty over other architectures or PDE families.

Multiplicity. We apply no multiplicity correction. The reported tests include two nested cavity sign tests (with p-values 0.008 and 0.0039 on seven and eight runs), one Burgers sign test, per-basis intervals, and percentiles relative to two null distributions. The eight-run result assesses inclusion of the screened run. The cavity sign tests reach their sample-size-dependent minima because all differences have the same sign. Effect magnitudes and seed-level intervals provide the complementary scale.

Preliminary diagnostics. Covariance of training. This diagnostic starts from two initializations with identical decoded fields, related by a sampled $\boldsymbol{W} \in O(r)$ through $(\boldsymbol{\Phi}, \boldsymbol{c}) \mapsto (\boldsymbol{\Phi}\boldsymbol{W}^\top, \boldsymbol{W}\boldsymbol{c})$. Train with identical data order and random seeds and measure the relative difference between decoded fields throughout an early window. An equivariant procedure preserves the relation up to floating-point effects. An immediate departure indicates basis dependence. Comparing only converged accuracy can miss this distinction when both runs approach the same solution. Appendix D.4 measures the roles of the optimizer, parameterization and later roundoff amplification.

Parameter sensitivity of the residual. This diagnostic varies one parameter over its intended range while holding the other inputs fixed. It measures the induced residual change with the training quadrature and compares it with the achieved training loss. A small change indicates a weak identification signal under the tested protocol. A contribution below the achieved scalar loss may still provide a systematic gradient.

Training. The operators used to audit physical identification train from PDE and boundary residuals. Training uses 50,000 steps of adaptive moment estimation (Adam) (Kingma & Ba, 2015). The learning rate decreases from 10^{-3} to 10^{-5} by cosine annealing. Weight decay is 5×10^{-6}, and the gradient norm is clipped at 1.0. Each step draws 16 parameter samples uniformly over the trained box. Each sample uses 5,000 interior collocation points and 2,000 on each of four walls. Residuals use a log-cosh penalty, loss weights are unity except the mean-zero pressure constraint at 0.5, and we omit limited-memory Broyden–Fletcher–Goldfarb–Shanno refinement to keep the optimization budgets comparable.

The cavity rotation, internal-rotation and random-projector audits use $\{1, 7, 42, 314, 1729, 2026, 5150, 31337\}$. Every seed is reported, including the run screened by the criterion below (Table 7).

Rank conditioning. Writing $\boldsymbol{C} \in \mathbb{R}^{n_R n_K \times 5}$ for the centred rank-5 coefficient table over the $n_R \times n_K$ parameter grid (n_R, n_K count Reynolds and asymmetry nodes), let $e_i = \sigma_i^2 / \sum_{k=1}^{5} \sigma_k^2$ be the coefficient-energy fractions, where $\sigma_1 \geq \cdots \geq \sigma_5$ are the singular values. A basis is retained when $e_5 \geq 0.02$. The checkpoint s7 falls below it (Table 7) and is screened out of the seven-seed completion and cross-seed physical comparisons, which report eight-run aggregates beside them. It remains in the eight-checkpoint invariance intervention. This criterion measures coefficient-energy conditioning. It does not imply mathematical rank deficiency or describe decoded-field conditioning. The statistic e_5 uses only the trained coefficient table. The threshold is our choice, and we made it knowing the completion errors of s7, which was among the first three runs. We fixed it in the commit that launched the five later seeds, before any of them finished training or was evaluated, and have not changed it since. It was not publicly preregistered. The threshold lies in an empty gap of the statistic (Table 7). The second row of Table 7 gives, for each run, the share of the interaction-labelled axis's variation that the ANOVA interaction component carries. Run s7 has two such axes.

Table 7: The screening statistic separates seven retained runs from the screened run, and the interaction-labelled axis carries most of its variation in the interaction component on every run. The threshold 0.02 lies in an empty gap. The eighth basis lies $5.5\times$ below the lowest retained basis, and any threshold in that gap gives the same seven-versus-one partition. The thresholds 0.015, 0.020, 0.025 and 0.030 all return it. Across all eight bases, the ranges of e_1, e_2, e_3 and e_4 are 0.356–0.411, 0.278–0.321, 0.172–0.192 and 0.077–0.099, respectively.

seed	s1	s42	s314	s1729	s2026	s5150	s31337	s7 (screened)
e_5	0.0714	0.0563	0.0767	0.0608	0.0579	0.0505	0.0624	**0.0092**
interaction share	0.916	0.915	0.914	0.916	0.916	0.919	0.917	0.974, 0.573

The expanded Burgers audit uses $\{1, 7, 11, 23, 42, 99, 314, 1729, 2026, 5150, 31337\}$. Other compact tables report seed means. Section 3.2 measures agreement across training seeds.

Reference solutions. We solve the steady incompressible Navier–Stokes equations in DOLFINx version 0.10.0 (Baratta et al., 2023) with the regularized lid profile of Equation (8). The two-sided family uses a κ-scaled bottom lid through a boundary lift. We use inf-sup-stable Taylor–Hood elements (P_2 velocity, P_1 pressure; P_k denotes degree-k polynomial finite elements) on a uniform 160×160 triangular mesh in double precision. A damped Newton iteration uses direct linear solves and a residual tolerance of 10^{-9}, with at most 80 iterations. We retain only converged solutions. Non-convergence near the steady/unsteady transition at high Re bounds the reported parameter boxes. We sample fields on a 100×100 spatial grid over the (Re, κ) lattice. Refinement to a 240×240 mesh confirms convergence of the reported physical quantities (Appendix H.2).

The cavity operators used for physical identification train without reference solutions. Post-training analyses use reference solutions, numerical or analytic, to construct snapshot-POD bases, ANOVA components, local tangents and reference-initialized varimax bases, to fit and evaluate reference-response rules and completion regressors, and to measure physical attribution and accuracy.

The completion task (Appendix E.1) uses FEM fields on the 11×9 parameter grid. Generic regressors fit the episode context and predict off-context targets. The per-coordinate and blockwise internal-rotation comparisons tune length scales on context only and enumerate all 63 cross centers (Appendix E.4.2). The separate rotation landscape (Table 5) calibrates per-coordinate length scales on the full reference lattice, including later target cells. These information budgets differ.

We evaluate u_{rel} on the full spatial grid, restricting parameter points to the trained box. We report off-node, on-node and whole-box averages where that choice affects a comparison (Table 24). Each configuration uses the final schedule-end checkpoint.

Reproduction of the evidence. The anonymous review package includes the twelve cavity checkpoints we trained, the eight development operators and the four confirmatory operators, under `cavity/checkpoints` with their resolved network configurations. It includes the 99 reference solutions on the evaluation grid, the confirmatory record and an example that replays the preregistered analysis on the shipped weights. Studies ran in separate environments. A study whose numbers depend on a library implementation records its versions beside its results, as the embedding study does (Appendix H.1.3), and the package's verification record gives the environment of the shipped checks. Archived-statistic verification re-derives the registered claims from the shipped records (verification record). Checkpoint-level reproduction reruns model-only identification, the completion task and the internal-rotation intervention from the shipped weights, and the confirmatory example replays its record. The package supplies the training code and configurations for retraining.

The external comparator requires all 169 certificate pairs with finite deviations satisfying their declared thresholds. Missing or refused certificates fail the comparison. We replayed the certificates on the public weights once, on 12 September 2026. Later entries of the verification record rerun the archived-statistic level, and smoke tests cover a subset. The repository's experimental setup links the configurations, commands and complete verification records.

C.1 GOVERNING EQUATIONS, BOUNDARY CONDITIONS AND EXACT CONSTRAINTS

Governing equations. All cavity operators solve the steady incompressible Navier–Stokes equations in nondimensional form on the unit square $\Omega = [0,1] \times [0,1]$, with coordinates (x, y) (unit density; velocity scaled by the peak lid speed, length by the cavity height, so the viscosity is $1/Re$):

$$(\boldsymbol{u} \cdot \nabla)\, \boldsymbol{u} + \nabla p_{\text{fluid}} - \frac{1}{Re} \nabla^2 \boldsymbol{u} = \boldsymbol{0}, \qquad \nabla \cdot \boldsymbol{u} = 0, \qquad \boldsymbol{u} = (u, v). \tag{7}$$

Here p_{fluid} is pressure and $\partial\Omega$ is the domain boundary; $\nabla\cdot$ and ∇^2 denote spatial divergence and the Laplacian. The loss and collocation protocol are in Appendix C. Continuity is identically zero for the streamfunction models. The pressure is a separate network output entering only through ∇p_{fluid}, so it is determined up to a constant and is not hard-constrained.

Boundary conditions. With $\xi = x$ the horizontal coordinate, the regularized lid profile is

$$u_{\text{lid}}(\xi) \;=\; 16\,\xi^2(1-\xi)^2, \tag{8}$$

which peaks at 1 mid-lid and vanishes at both corners (removing the corner velocity discontinuity of the classical uniform lid). The boundary conditions are

$$u(x,1) = u_{\text{lid}}(\xi), \qquad u(x,0) = \kappa\, u_{\text{lid}}(\xi), \qquad u = 0 \text{ on the side walls}, \qquad v = 0 \text{ on } \partial\Omega, \tag{9}$$

so $\kappa \in [-1, 1]$ scales an equal-profile bottom lid (κ=0 recovers the classical single-lid cavity, κ=1 the symmetric two-sided one).

Exact constraints via boundary lifting. With ψ denoting streamfunction, the ansatz of Equation (12) enforces incompressibility, no penetration and the prescribed lid velocities architecturally. Stationary-wall tangential no-slip is imposed softly.

The lift

$$\psi_{\text{lift}}(\xi, y;\, \kappa) \;=\; u_{\text{lid}}(\xi)\,\big[\, y^2(y-1) \;+\; \kappa\, y(1-y)^2 \,\big] \tag{10}$$

satisfies $\partial_y \psi_{\text{lift}} = u_{\text{lid}}$ at y=1 and $\kappa\, u_{\text{lid}}$ at y=0, and vanishes on all four walls (the side walls because u_{lid} vanishes at $\xi \in \{0, 1\}$, which is why the regularized lid is required for the two-sided family). The boundary factor multiplying the learned correction is

$$b(\xi, y) \;=\; \begin{cases} 16\,\xi(1-\xi)\, y\,(1-y)^2 & \text{single-lid operators,} \\ 64\,\xi(1-\xi)\, y^2(1-y)^2 & \text{two-sided } (\kappa) \text{ operators.} \end{cases} \tag{11}$$

Both factors vanish on every wall, making ψ constant along each wall and enforcing no penetration exactly. Both satisfy $\partial_y b = 0$ at $y = 1$. The two-sided factor also satisfies it at $y = 0$. The learned correction therefore preserves the prescribed lid velocities. Tangential no-slip on stationary walls is imposed by a boundary residual. Incompressibility is exact because $u = \partial_y \psi$, $v = -\partial_x \psi$ satisfy continuity identically.

Low-rank representation. The trunk returns 200 features at the spatial vector $\boldsymbol{x} = (x, y)$. The first 100 form the streamfunction features $T(\boldsymbol{x})$ and the last 100 the pressure features $T_{\text{fluid}}(\boldsymbol{x})$, each padded with zeros to length 200. The cavity operator represents the two outputs as

$$\begin{aligned} \psi(\boldsymbol{x}; p) &= \psi_{\text{lift}}(\boldsymbol{x}; p) + b(\boldsymbol{x})\big(\langle \boldsymbol{z}(p), T(\boldsymbol{x})\rangle + \beta_\psi\big), \\ p_{\text{fluid}}(\boldsymbol{x}; p) &= \langle \boldsymbol{z}(p), T_{\text{fluid}}(\boldsymbol{x})\rangle + \beta_{\text{fluid}}, \end{aligned} \tag{12}$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product and $\beta_\psi, \beta_{\text{fluid}}$ are the two shared output biases. The lift and boundary factor are defined above and act on the streamfunction only. The pressure has no lift and no boundary factor, and the mean-zero penalty of Appendix C fixes its constant. The latent vector $\boldsymbol{z}$ feeds both outputs, so a reparameterization that preserves $\boldsymbol{z}$ preserves both. The decoder Gram matrix $\boldsymbol{G}$ uses the velocity alone. The 200-dimensional latent vector has the expansion

$$\boldsymbol{z}(p) = \bar{\boldsymbol{z}} + \sum_{k=1}^{r} c_k(p) \boldsymbol{v}_k = \bar{\boldsymbol{z}} + \boldsymbol{V}^\top \boldsymbol{c}(p), \qquad r = 5. \tag{13}$$

Here $\bar{\boldsymbol{z}}$ is the latent mean, and the rows of $\boldsymbol{V}$ form the effective orthonormal basis defined in Appendix D.4. The trunk maps this basis to the decoded mode matrix $\boldsymbol{\Phi}$. Appendix D.5 derives the admitted orthogonal action with the latent metric and surrounding representation fixed.

D THEORY AND COVARIANCE TESTS

D.1 INVARIANCE UNDER BASIS CHANGES AND ESTIMATOR EQUIVARIANCE

We use the coefficient–decoder action of Equation (4). Observation 1 states the standard identity behind basis-independent decoded predictions.

Observation 1 (equivariance implies invariant decoded predictions)**.** *Let F map coefficients, decoder and basis-independent side information s to predicted coefficients. If $F(\boldsymbol{W}\boldsymbol{c}, \boldsymbol{\Phi}\boldsymbol{W}^\top, s) = \boldsymbol{W}F(\boldsymbol{c}, \boldsymbol{\Phi}, s)$ for every admissible $\boldsymbol{W} \in O(r)$, then $\boldsymbol{\Phi}\boldsymbol{W}^\top F(\boldsymbol{W}\boldsymbol{c}, \boldsymbol{\Phi}\boldsymbol{W}^\top, s) = \boldsymbol{\Phi}F(\boldsymbol{c}, \boldsymbol{\Phi}, s)$.*

For an identified decomposition the condition is required only within the identified blocks, and it concerns the complete fitting and tuning procedure, normalization, penalties and tie-breaking included (Lemma 1). The guarantee holds for fixed identified subspaces. The input hypothesis in Lemma 1(i) is necessary. When coefficients enter as *inputs* and targets, the joint action is $f \mapsto \boldsymbol{W}f(\boldsymbol{W}^\top\cdot)$. For linear maps this is conjugation $\boldsymbol{A} \mapsto \boldsymbol{W}\boldsymbol{A}\boldsymbol{W}^\top$. In a class closed under $\boldsymbol{A} \mapsto \boldsymbol{W}\boldsymbol{A}$, the rotated-data minimizer $\widehat{\boldsymbol{A}}_{\boldsymbol{W}}$ need not equal $\boldsymbol{W}\widehat{\boldsymbol{A}}\boldsymbol{W}^\top$.

For an explicit example, let e_1, e_2 be the unit coordinate vectors in $\mathbb{R}^2$ and minimize $\|\boldsymbol{A}e_2 - e_2\|_2^2 + \lambda\|\boldsymbol{A}\|_F^2$ over $\boldsymbol{A}e_1 = 0$, with $\lambda > 0$. The class is closed under left multiplication, and its unique minimizer is $\widehat{\boldsymbol{A}} = (1+\lambda)^{-1}e_2e_2^\top$. Let $\boldsymbol{W}$ swap the two coordinates. Both input and target become e_1, so the unchanged constraint gives $\widehat{\boldsymbol{A}}_{\boldsymbol{W}} = 0$. The conjugated minimizer is $\boldsymbol{W}\widehat{\boldsymbol{A}}\boldsymbol{W}^\top = (1+\lambda)^{-1}e_1e_1^\top \neq 0$.

Proof of Observation 1. Substituting the hypothesis into the decoded output gives $\boldsymbol{\Phi}\boldsymbol{W}^\top F(\boldsymbol{W}\boldsymbol{c}, \boldsymbol{\Phi}\boldsymbol{W}^\top, s) = \boldsymbol{\Phi}\boldsymbol{W}^\top\boldsymbol{W}F(\boldsymbol{c}, \boldsymbol{\Phi}, s) = \boldsymbol{\Phi}F(\boldsymbol{c}, \boldsymbol{\Phi}, s)$, since $\boldsymbol{W}^\top\boldsymbol{W} = \boldsymbol{I}$. □

The following cases give sufficient conditions for an estimator to satisfy the equivariance hypothesis. Clause (ii′) restricts the guarantee to the *block stabilizer*, the group of rotations inside the identified blocks.

Lemma 1 (estimators that satisfy the hypothesis)**.** *$F(\boldsymbol{W}\boldsymbol{c}, \boldsymbol{\Phi}\boldsymbol{W}^\top, s) = \boldsymbol{W}F(\boldsymbol{c}, \boldsymbol{\Phi}, s)$ under the following conditions, for all $\boldsymbol{W} \in O(r)$ except in (ii′), where $\boldsymbol{W}$ belongs to the block stabilizer.* ***(i)*** *F minimizes empirical risk over a class closed under $f \mapsto \boldsymbol{W}f$. The loss and penalty are $O(r)$-invariant, the inputs are unchanged by the group action, and the minimizer is unique or selected by an $O(r)$-equivariant tie-break.* ***(ii)*** *F is a shared-weight smoother whose weights depend only on the design and query.* ***(ii′)*** *F is a shared-weight smoother whose response-dependent weights are selected by a block-invariant rule from a candidate set that is fixed in advance, or built by a rule with the same invariance. The conclusion then holds for the block stabilizer. Invariance under all of $O(r)$ requires a rule invariant under that larger group.* ***(iii)*** *F returns $\boldsymbol{c}_\theta$ after isotropic gradient descent on parameters θ. The group acts on the parameters by orthogonal matrices $\rho(\boldsymbol{W})$, which are isometries of the parameter space, and $\boldsymbol{c}_{\rho(\boldsymbol{W})\theta} = \boldsymbol{W}\boldsymbol{c}_\theta$. The objective is invariant under the joint action on parameters and data, and the initialization is covariant. The same guarantee holds for Newton updates with an invertible Hessian, or an equivariant generalized inverse and tie-break.*

Proof of Lemma 1. (i) The risk of f on targets $\boldsymbol{c}$ equals that of $\boldsymbol{W}f$ on targets $\boldsymbol{W}\boldsymbol{c}$, so minimizers correspond by $f \mapsto \boldsymbol{W}f$. Uniqueness (as under a strictly convex regularized objective) or an equivariant tie-break then gives $\hat{f}_{\boldsymbol{W}} = \boldsymbol{W}\hat{f}$. Without such a rule, the optimal risk is invariant and the minimizer sets transform equivariantly. The hypothesis requires an equivariant selection rule. (ii) For smoother weights $w_i(p)$ at observed inputs p_i, $i = 1, \dots, N$, replacing $\boldsymbol{c}$ by $\boldsymbol{W}\boldsymbol{c}$ in $\hat{\boldsymbol{c}}(p) = \sum_i w_i(p)\,\boldsymbol{c}(p_i)$ gives $\sum_i w_i(p)\,\boldsymbol{W}\boldsymbol{c}(p_i) = \boldsymbol{W}\hat{\boldsymbol{c}}(p)$. Invariance of extrapolated predictions then follows algebraically. (ii′) Our downstream methods choose length scales by minimizing a leave-one-out criterion on the context, so w_i is a function of $\boldsymbol{c}$ and (ii) does not apply as stated. For a candidate kernel let $\boldsymbol{A} = \boldsymbol{K} + \lambda\boldsymbol{I}$ be positive definite and define

$$\boldsymbol{D} = \mathrm{Diag}\big((\boldsymbol{A}^{-1})_{11}, \dots, (\boldsymbol{A}^{-1})_{NN}\big), \qquad \boldsymbol{N} = \boldsymbol{D}^{-1}\boldsymbol{A}^{-1}.$$

Here $\boldsymbol{K}$ is the context kernel matrix, Diag constructs a diagonal matrix and $\boldsymbol{Y} \in \mathbb{R}^{N\times r}$ contains response rows. Left multiplication normalizes each row, giving LOO residuals $\boldsymbol{N}\boldsymbol{Y}$. For a coordinate-index set B, the summed block loss is $\mathcal{L}_B(\boldsymbol{Y}) = \|\boldsymbol{N}\boldsymbol{Y}_B\|_F^2$ (division by N does not

change its minimizer). Responses are stored as rows, so a rotation $\boldsymbol{Q}_B \in O(|B|)$ of a block of size $|B|$ acts as $\boldsymbol{Y}_B \mapsto \boldsymbol{Y}_B \boldsymbol{Q}_B^\top$. The residuals become $\boldsymbol{N}\boldsymbol{Y}_B\boldsymbol{Q}_B^\top$, with the same Frobenius norm, so every candidate loss is unchanged. The candidate grid is fixed in advance (Appendix E.1) and ties are broken by candidate order. The same length scales are then selected, and the fitted map is equivariant under the stabilizer of the blocks. A grid built from per-coordinate statistics of the responses need not give the same candidates in two bases. Taking B to be all r axes gives the fully shared case, equivariant under all of $O(r)$. A *per-coordinate* criterion need not be invariant under rotations mixing coordinates, permitting basis-dependent hyperparameter choices. Section 3.4 tests this consequence at fixed projectors. Appendix E.4.2 gives the complete rotation experiment. (iii) Write J for the objective on the original data and $J_{\boldsymbol{W}}$ for its counterpart on the rotated data. Joint invariance means $J_{\boldsymbol{W}}(\rho(\boldsymbol{W})\theta) = J(\theta)$ for every θ. Dependence on the coefficients only through the decoded field is a sufficient special case. Differentiating with respect to θ gives $\rho(\boldsymbol{W})^\top \nabla J_{\boldsymbol{W}}(\rho(\boldsymbol{W})\theta) = \nabla J(\theta)$. Because $\rho(\boldsymbol{W})$ is orthogonal, this reads $\nabla J_{\boldsymbol{W}}(\rho(\boldsymbol{W})\theta) = \rho(\boldsymbol{W})\nabla J(\theta)$. This is where the isometry enters. Under a general invertible action the gradient would transform with the inverse transpose, and an *isotropic* step would not commute with the action. Gradient descent $\theta^{t+1} = \theta^t - \eta \nabla J(\theta^t)$, with learning rate η and iteration t, satisfies $\theta^t_{\boldsymbol{W}} = \rho(\boldsymbol{W})\theta^t$ at every iterate, by induction from the covariant initialization. Hence $\boldsymbol{c}^t_{\boldsymbol{W}} = \boldsymbol{W}\boldsymbol{c}^t$. Descent on the coefficients themselves is the case $\theta = \boldsymbol{c}$, $\rho(\boldsymbol{W}) = \boldsymbol{W}$. The Hessian conjugates, $\nabla^2 J_{\boldsymbol{W}}(\rho(\boldsymbol{W})\theta) = \rho(\boldsymbol{W})\,\nabla^2 J(\theta)\,\rho(\boldsymbol{W})^\top$, so a Newton step with an invertible Hessian is equivariant as well. For a singular Hessian, the same statement requires an equivariant generalized inverse and tie-break. □

Clause (iii) concerns an estimator fitted on fixed coordinates. Training the operator itself is a different setting. Forward QR is not an orthogonal action on the raw parameters, and Adam's step is not isotropic, so the hypotheses of the clause fail there (Appendix D.4).

Corollary 1 (auxiliary-state equivariance)**.** *Let $\mathcal{A}$ be within-span auxiliary state and $\rho_{\mathrm{aux}}(\boldsymbol{W})$ its group action. If $F(\boldsymbol{W}\boldsymbol{c}, \boldsymbol{\Phi}\boldsymbol{W}^\top, \rho_{\mathrm{aux}}(\boldsymbol{W})\mathcal{A}, s) = \boldsymbol{W}F(\boldsymbol{c}, \boldsymbol{\Phi}, \mathcal{A}, s)$, then the decoded output is the same in every basis. A coefficient covariance Σ must transform as $\Sigma \mapsto \boldsymbol{W}\Sigma\boldsymbol{W}^\top$: altering $\boldsymbol{\Phi}\Sigma\boldsymbol{\Phi}^\top$ changes the represented state. A change of basis preserves this decoded covariance.*

Corollary 2 (the shared-kernel regressor)**.** *Kernel ridge regression with one length-scale pair and one penalty shared by all r axes is a shared-weight smoother, and selecting that pair by a leave-one-out criterion summed over all axes is an invariant rule, so Lemma 1(ii′) gives $\hat{\boldsymbol{c}}_{\boldsymbol{W}} = \boldsymbol{W}\hat{\boldsymbol{c}}$ and identical decoded completions in exact arithmetic.*

This corollary applies to the shared-kernel baseline. Corollary 1 separately covers methods with additional within-span state.

D.2 Selection-rule and downstream-method compatibility

Lemma 2 (necessity of non-equivariance)**.** *Basis-dependent decoded predictions require failure of equivariance in the complete coefficient-prediction procedure under the declared joint transformation of coefficients, decoder and auxiliary state. When $\boldsymbol{\Phi}$ has full column rank, that failure is also sufficient. Per-coordinate hyperparameters, constraints, priors or preconditioners can cause such failure when held fixed as the basis changes. Anisotropic state need not break equivariance when it transforms covariantly (Corollary 1).*

Proof of Lemma 2. If the complete procedure is equivariant, Observation 1 and Corollary 1 give identical decoded predictions. The contrapositive proves the claim. For the converse, write $\widehat{\boldsymbol{c}}$ and $\widehat{\boldsymbol{c}}_{\boldsymbol{W}}$ for the predictions in the two bases. The decoded predictions differ by $\boldsymbol{\Phi}(\boldsymbol{W}^\top\widehat{\boldsymbol{c}}_{\boldsymbol{W}} - \widehat{\boldsymbol{c}})$. A decoder with full column rank is injective, so this difference vanishes only if $\widehat{\boldsymbol{c}}_{\boldsymbol{W}} = \boldsymbol{W}\widehat{\boldsymbol{c}}$. Every cavity decoder studied here has full column rank (Appendix H.2). A rank-deficient decoder can leave the output unchanged when predicted coefficients change. Individual operations need not be invariant, and their presence alone does not establish failure of the complete procedure. For example, a covariance transformed as $\Sigma \mapsto \boldsymbol{W}\Sigma\boldsymbol{W}^\top$ can preserve equivariance despite being anisotropic. □

Proposition 1 (small variance contributions can have large derivatives)**.** *Let p_1, p_2 be independent uniform random variables on $[0,1]$ and $c(p_1,p_2) = \big(p_1 + \frac{\alpha}{2\pi k}\sin(2\pi k p_2), p_2\big)$ for fixed $\alpha > 0$ and integer $k \geq 1$. The contribution of p_2 to the first-axis variance is exactly $\alpha^2/(8\pi^2k^2)$, so this cross-parameter variance contribution vanishes as k^{-2}. The derivative satisfies $\partial c_1/\partial p_2 = \alpha\cos(2\pi k p_2)$ and $\|\partial c_1/\partial p_2\|_{L^2} = \alpha/\sqrt{2}$ for every k.*

Proof. Writing Var for variance, $\mathrm{Var}[\frac{\alpha}{2\pi k}\sin(2\pi k p_2)] = \frac{\alpha^2}{4\pi^2 k^2}\mathrm{Var}[\sin(2\pi k p_2)] = \frac{\alpha^2}{8\pi^2 k^2}$ for integer $k \geq 1$, since $\sin(2\pi k p_2)$ has mean 0 and mean square $\frac{1}{2}$ over a whole number of periods. Differentiating gives $\partial c_1/\partial p_2 = \alpha\cos(2\pi k p_2)$, whose mean square is likewise $\frac{1}{2}$ for every k, so its L^2 norm is $\alpha/\sqrt{2}$ independently of k. □

Identification rule–downstream method compatibility. For m labelled blocks with dimensions $r_1 + \cdots + r_m = r$, generalize the stabilizer to $H = \prod_{j=1}^{m} O(r_j)$. Using the coordinate projectors E_j of Equation (19), define $\Pi_j(\boldsymbol{W}) = \boldsymbol{W}^\top E_j \boldsymbol{W}$. The *orbit* $[\boldsymbol{W}] = H\boldsymbol{W}$, the set of all $\boldsymbol{M}\boldsymbol{W}$ with $\boldsymbol{M} \in H$, contains the equivalent representatives, and $H\backslash O(r)$ is the set of these orbits. In the cavity application H is the group of orthogonal changes of the field-orthonormal axes, in the labelled axis order, that fix every retained labelled projector. The varimax basis is a computational representative of its orbit.

Proposition 2 (identification rule–downstream method compatibility)**.** *An H-invariant objective can identify at most a labelled projector tuple. The bases within its blocks remain unresolved. Uniqueness of the identified tuple requires additional evidence. An H-equivariant downstream method yields a well-defined decoded prediction on the identified coset. Without such equivariance (or direct H-invariance of the reported output), the identification rule alone gives no representative-independent guarantee.*

Proof. For any $\boldsymbol{Q} \in H$, block diagonality gives $\Pi_j(\boldsymbol{Q}\boldsymbol{W}) = \boldsymbol{W}^\top\boldsymbol{Q}^\top E_j \boldsymbol{Q}\boldsymbol{W} = \boldsymbol{W}^\top E_j \boldsymbol{W} = \Pi_j(\boldsymbol{W})$. Thus every representative in the orbit $H\boldsymbol{W}$ has the same labelled projector tuple, and an H-invariant objective cannot distinguish their within-block bases. The converse also holds. If $\Pi_j(\boldsymbol{W}_1) = \Pi_j(\boldsymbol{W}_2)$ for every j, then $\boldsymbol{M} = \boldsymbol{W}_2\boldsymbol{W}_1^\top$ satisfies $\boldsymbol{M}E_j = E_j\boldsymbol{M}$. The E_j are coordinate projectors that sum to the identity, so $\boldsymbol{M} = \sum_j E_j\boldsymbol{M}E_j$ is block diagonal with orthogonal blocks, and $\boldsymbol{W}_2 = \boldsymbol{M}\boldsymbol{W}_1$ lies in $H\boldsymbol{W}_1$. Labelled projector tuples and orbits correspond one to one. The operator-response objective of Equation (19) is an example: $\|E_j\boldsymbol{W}g\|_2^2 = g^\top\Pi_j(\boldsymbol{W})g$, so it depends on $\boldsymbol{W}$ only through the tuple. This does not establish a unique preferred coset: a constant objective, for example, leaves every tuple optimal. If the downstream method is H-equivariant, its output transforms covariantly between these representatives and therefore defines the same decoded prediction on the quotient space. Direct invariance of the reported output also suffices. Without either condition, a prediction need not be well-defined on the quotient space. A per-coordinate method, for example, may assign different hyperparameters after an in-block rotation and produce different outputs for the same projector tuple. □

Field varimax is not of this kind. It selects a basis whose criterion and grouped subspaces are unchanged by signed column permutations, and its criterion changes under other rotations inside a block (Appendix J.2). Grouping the selected axes by their labels gives a labelled projector tuple, and an internal rotation preserves that tuple. It need not preserve the varimax criterion, and the labelling rule applied to a rotated basis need not reproduce the axis labels.

D.3 Invertible reparameterizations in the field metric

A linear decoder with no orthonormality constraint admits every invertible $\boldsymbol{A}$ in the general linear group $GL(r)$, including anisotropic scaling and shear. The reparameterization is $\boldsymbol{c}_{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{c}$, $\boldsymbol{\Phi}_{\boldsymbol{A}} = \boldsymbol{\Phi}\boldsymbol{A}^{-1}$. It preserves each decoded field and turns the decoder Gram matrix into $\boldsymbol{G}_{\boldsymbol{A}} = \boldsymbol{A}^{-\top}\boldsymbol{G}\boldsymbol{A}^{-1}$.

Proposition 3 (field-orthonormal coordinates absorb invertible reparameterizations)**.** *Let $\boldsymbol{G} = \boldsymbol{\Phi}^\top\boldsymbol{M}\boldsymbol{\Phi}$ be positive definite. For every $\boldsymbol{A} \in GL(r)$, $\boldsymbol{d}_{\boldsymbol{A}} = \boldsymbol{S}_{\boldsymbol{A}}\boldsymbol{d}$ with $\boldsymbol{S}_{\boldsymbol{A}} = \boldsymbol{G}_{\boldsymbol{A}}^{1/2}\boldsymbol{A}\boldsymbol{G}^{-1/2} \in O(r)$.*

Proof. By definition $\boldsymbol{d}_{\boldsymbol{A}} = \boldsymbol{G}_{\boldsymbol{A}}^{1/2}\boldsymbol{c}_{\boldsymbol{A}} = \boldsymbol{G}_{\boldsymbol{A}}^{1/2}\boldsymbol{A}\boldsymbol{G}^{-1/2}\boldsymbol{d}$, and $\boldsymbol{S}_{\boldsymbol{A}}^\top\boldsymbol{S}_{\boldsymbol{A}} = \boldsymbol{G}^{-1/2}\boldsymbol{A}^\top\boldsymbol{G}_{\boldsymbol{A}}\boldsymbol{A}\boldsymbol{G}^{-1/2} = \boldsymbol{G}^{-1/2}\boldsymbol{A}^\top\boldsymbol{A}^{-\top}\boldsymbol{G}\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{G}^{-1/2} = \boldsymbol{I}$. □

The hypothesis holds on every run (Appendix H.2), so $\boldsymbol{A}$ leaves the distances between field-orthonormal coordinates exactly unchanged. Distances between stored coefficients do change, because $\|\boldsymbol{A}(\boldsymbol{c} - \boldsymbol{c}')\|_2$ depends on $\boldsymbol{A}$. For $\boldsymbol{A} = \boldsymbol{W} \in O(r)$ the proposition gives $\boldsymbol{d}_{\boldsymbol{W}} = \boldsymbol{W}\boldsymbol{d}$. The next proposition reduces agreement of the identified subspaces after a reparameterization to a condition on the varimax iteration.

Field-varimax specification. The field-orthonormal table $\boldsymbol{D}$ holds the coordinates $\boldsymbol{d}(p)$ on the 11×9 grid of Re and κ values, one column per axis, $r = 5$. We subtract the grand mean over the grid from each column and form two main-effect tables, the means over κ (an $11 \times r$ table) and the means over Re (a $9 \times r$ table). We scale them by $\sqrt{9}$ and $\sqrt{11}$, so that their squared entries sum to the ANOVA main-effect sums of squares. Stacked, they form the loading matrix L of shape $m \times r$ with $m = 20$. No column is normalized. With $\Lambda = L\boldsymbol{R}$, the varimax criterion is $V_L(\boldsymbol{R}) = \sum_{ij} \Lambda_{ij}^4 - \frac{1}{m} \sum_j \big(\sum_i \Lambda_{ij}^2\big)^2$. One step F_L replaces $\boldsymbol{R}$ by the orthogonal polar factor $\boldsymbol{U}\boldsymbol{V}^\top$ of $\boldsymbol{B}(\boldsymbol{R}) = L^\top h(\Lambda)$, from the singular value decomposition $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$, where $h(\Lambda) = \Lambda^{\circ 3} - \Lambda\,\mathrm{diag}(\mathbf{1}^\top \Lambda^{\circ 2})/m$, $\circ$ denotes entrywise powers and $\mathbf{1}$ is the all-ones vector. The iteration starts at $\boldsymbol{R} = \boldsymbol{I}$. It stops when the sum of the singular values of $\boldsymbol{B}$ grows by less than 10^{-9} relative to the previous step, or after 200 steps. Two-way ANOVA shares of its variation over the grid then label each column of $\boldsymbol{D}\boldsymbol{R}$. A main effect above one half names the column Reynolds or asymmetry, and any other column is interaction. A block is the set of columns with one label. No sign or ordering convention is needed, because a label ignores the sign of a column and a set has no order. The Burgers family applies the same labelling rule at 0.6 and calls the residual label mixed (Appendix I).

Proposition 4 (field varimax after a reparameterization)**.** *Let L be the main-effect loadings of the field-orthonormal table, $V_L(\boldsymbol{R})$ the varimax criterion of $L\boldsymbol{R}$ for $\boldsymbol{R} \in O(r)$, and F_L one step of the varimax iteration. A reparameterization that changes the coordinates by $\boldsymbol{S} \in O(r)$ changes the loadings to $L\boldsymbol{S}^\top$. Then $V_{L\boldsymbol{S}^\top}(\boldsymbol{R}) = V_L(\boldsymbol{S}^\top \boldsymbol{R})$, and $F_{L\boldsymbol{S}^\top}(\boldsymbol{R}) = \boldsymbol{S}\, F_L(\boldsymbol{S}^\top \boldsymbol{R})$ whenever the update matrix of the step is nonsingular. Hence, if every update matrix met along the two iterations is nonsingular, the iteration started at the identity in the reparameterized coordinates equals $\boldsymbol{S}$ times the iteration started at $\boldsymbol{S}^\top$ in the original ones. If both starts reach the same maximizer up to a signed column permutation, the rotated coordinate tables agree up to that permutation, and the labelled decoded subspaces coincide.*

Proof. The loadings are linear in the table, and the table changes from $\boldsymbol{D}$ to $\boldsymbol{D}\boldsymbol{S}^\top$, which gives $L\boldsymbol{S}^\top$. The criterion depends on $\boldsymbol{R}$ only through $L\boldsymbol{S}^\top\boldsymbol{R}$, which proves the first identity. A varimax step returns the orthogonal polar factor of $\boldsymbol{B}(\boldsymbol{R}) = L^\top h(L\boldsymbol{R})$, where h acts on the rotated loadings. For the reparameterized loadings this matrix is $\boldsymbol{S}\,\boldsymbol{B}(\boldsymbol{S}^\top\boldsymbol{R})$. A nonsingular $\boldsymbol{B}$ has the unique orthogonal polar factor $\boldsymbol{B}(\boldsymbol{B}^\top\boldsymbol{B})^{-1/2}$, which the routine computes from a singular value decomposition, and $\boldsymbol{S}\boldsymbol{B}\big((\boldsymbol{S}\boldsymbol{B})^\top\boldsymbol{S}\boldsymbol{B}\big)^{-1/2} = \boldsymbol{S}\,\boldsymbol{B}(\boldsymbol{B}^\top\boldsymbol{B})^{-1/2}$. This proves the second identity. For a singular $\boldsymbol{B}$ the factor is not unique and the identity can fail. The hypothesis concerns the varimax updates and is separate from positive definiteness of the decoder Gram matrix. The stopping rule uses the singular values of $\boldsymbol{B}$, which $\boldsymbol{S}$ does not change, so induction on the steps gives the statement about the two starts. If the limits agree up to a signed permutation $\boldsymbol{P}$, the rotated tables are $\boldsymbol{D}\boldsymbol{S}^\top\boldsymbol{S}\boldsymbol{R}\boldsymbol{P} = \boldsymbol{D}\boldsymbol{R}\boldsymbol{P}$. A label depends on one column of the rotated table and ignores its sign, so the labelled blocks coincide. For the reparameterization of Proposition 3, $\boldsymbol{S} = \boldsymbol{S}_{\boldsymbol{A}}$ and the decoded modes are $\boldsymbol{\Phi}_{\boldsymbol{A}}\boldsymbol{G}_{\boldsymbol{A}}^{-1/2}\boldsymbol{S}\boldsymbol{R}\boldsymbol{P} = \boldsymbol{\Phi}\boldsymbol{G}^{-1/2}\boldsymbol{R}\boldsymbol{P}$, because $\boldsymbol{A}^{-1}\boldsymbol{G}_{\boldsymbol{A}}^{-1/2}\boldsymbol{S}_{\boldsymbol{A}} = \boldsymbol{G}^{-1/2}$. □

The proposition has two conditions. The first, nonsingular updates, holds on every iteration we compare, with little room (Table 8). A varimax update cubes the loadings, and the interaction axis carries a very small share of the main-effect loadings, so the update matrices are nonsingular and badly conditioned. The polar factor is unique at every step and sensitive to rounding: the step identity holds to about 10^{-8}, and both iterations stop at the same step in every comparison. The second condition holds when the criterion has one maximizer up to signed permutation and the iteration reaches it from every start. We do not prove that. The restart study of Appendix K supports it numerically: every start drawn uniformly from the rotation group returned the same blocks on all eight checkpoints. Agreement of restarts is evidence and does not establish global uniqueness. Appendix D.5 tests the distance statement and the recovery of the blocks under sampled reparameterizations.

D.4 Covariance of training under a change of basis

Optimization scope. Let $\boldsymbol{U} \in \mathbb{R}^{r \times 200}$ denote the raw trainable mode parameters and $\boldsymbol{V} = R(\boldsymbol{U}) = \mathrm{qf}(\boldsymbol{U}^\top)^\top$ the effective row-orthonormal basis, with row k written $\boldsymbol{v}_k^\top$. In the QR factorization $\boldsymbol{U}^\top = Q_{\mathrm{qr}} R_{\mathrm{qr}}$, Q_{qr} has orthonormal columns and R_{qr} is upper triangular; qf returns Q_{qr}. The effective transformation $\boldsymbol{V}' = \boldsymbol{W}\boldsymbol{V}$, $\boldsymbol{c}' = \boldsymbol{W}\boldsymbol{c}$ preserves $\boldsymbol{V}^\top\boldsymbol{c}$. The naive raw reparameterization $\boldsymbol{U}' = \boldsymbol{W}\boldsymbol{U}$ need not realize it, since $R(\boldsymbol{W}\boldsymbol{U}) \neq \boldsymbol{W}R(\boldsymbol{U})$ in general.

Table 8: The nonsingularity hypothesis of Proposition 4 on the compared iterations of all eight checkpoints. The relative singular value is the smallest over the largest singular value of a varimax update matrix, and the step residual is the largest entry of $\boldsymbol{R}'_k - \boldsymbol{S}\boldsymbol{R}_k$ over the steps k of the two iterations.

	sampled reparameterizations	uniform restarts
comparisons	160	2400
smallest relative singular value	1.3×10^{-9}	8.5×10^{-11}
largest step residual	9.4×10^{-9}	4.5×10^{-9}
same stopping step	160 of 160	2400 of 2400

The paired initializations store $\boldsymbol{U}_{\boldsymbol{W}} = \boldsymbol{W}\boldsymbol{V}$. If forward QR returns $R(\boldsymbol{U}_{\boldsymbol{W}}) = S_{\boldsymbol{W}}\boldsymbol{W}\boldsymbol{V}$ for a diagonal sign matrix $S_{\boldsymbol{W}}$, the coefficient map is reparameterized as $\boldsymbol{c}_{\boldsymbol{W}} = S_{\boldsymbol{W}}\boldsymbol{W}\boldsymbol{c}$. This compensated reparameterization preserves the decoded field (Appendix E.4.1). The implementation stores coefficient vectors as rows of C and accepts $\boldsymbol{T} = \boldsymbol{W}^\top$. It sets $\boldsymbol{U}_{\boldsymbol{T}} = \boldsymbol{T}^\top R(\boldsymbol{U})$. Let $S_{\boldsymbol{T}}$ be the diagonal sign matrix that forward QR returns for this input, $R(\boldsymbol{U}_{\boldsymbol{T}}) = S_{\boldsymbol{T}}\boldsymbol{T}^\top R(\boldsymbol{U})$. The coefficient rows become $C_{\boldsymbol{T}} = C\boldsymbol{T}S_{\boldsymbol{T}}$. The identity control uses the same operation with $\boldsymbol{T} = \boldsymbol{I}$, including its QR signs. With forward QR disabled, the reparameterization is simply $\boldsymbol{U}_{\boldsymbol{T}} = \boldsymbol{T}^\top\boldsymbol{U}$, $C_{\boldsymbol{T}} = C\boldsymbol{T}$.

The gradient regression test compares $\nabla_C\mathcal{L}_{\boldsymbol{T}}$ with $(\nabla_C\mathcal{L}_{\boldsymbol{I}})\boldsymbol{T}$ and $\nabla_{\boldsymbol{U}}\mathcal{L}_{\boldsymbol{T}}$ with $\boldsymbol{T}^\top\nabla_{\boldsymbol{U}}\mathcal{L}_{\boldsymbol{I}}$, using the maximum absolute entry of each difference divided by that of its baseline gradient. Here $\mathcal{L}_{\boldsymbol{T}}$ and $\mathcal{L}_{\boldsymbol{I}}$ are the losses after the transformed and identity reparameterizations. Each sums squared decoded velocities at fixed inputs.

Relative gradient differences are $O(1)$ with QR and $\approx 10^{-16}$ without it. These comparisons use the naive linear action: they omit the QR sign compensation and do not differentiate the compensated reparameterization. They therefore do not establish failure of covariance under every field-preserving parameter transformation. Covariance of raw-parameter gradients and of the optimization trajectory must be assessed separately from decoded-field invariance.

The optimizer introduces a further covariance condition. Lemma 1(iii) covers plain gradient descent and the qualified Newton update under an equivariant parameterization. Adam's per-coordinate second moments can break $O(r)$ equivariance. We therefore also test an adaptation of VectorAdam (Ling et al., 2022) to latent blocks, which shares one second moment across each rotated block. Forward QR still requires a separate covariance check, and reference-solution accuracy is evaluated independently.

Training comparisons. The archived comparisons train the same rank-5 operator from the compensated paired initializations above, with forward QR active or disabled. They measure decoded-field differences across bases along each optimization trajectory and assess reference-solution accuracy separately. These measurements concern the specified parameter actions and training configurations.

D.5 Centred coefficients, the scope of $O(r)$ and sampled invertible reparameterizations

Centering removes the translational degrees of freedom. With $\bar{\boldsymbol{c}}$ the mean under the declared parameter measure, $\boldsymbol{c} \leftarrow \boldsymbol{c} - \bar{\boldsymbol{c}}$ and $\bar{\boldsymbol{z}} \leftarrow \bar{\boldsymbol{z}} + \boldsymbol{V}^\top\bar{\boldsymbol{c}}$ preserve every decoded field. All reported quantities use this centred basis. When the centred table spans $\mathbb{R}^r$ (tested by the e_5 rule), the coefficient–mode reparameterization class, with the latent metric and specified surrounding representation fixed, admits exactly $O(r)$. The reparameterization $\boldsymbol{c}_A = \boldsymbol{A}\boldsymbol{c}$, $\boldsymbol{V}_A = \boldsymbol{A}^{-\top}\boldsymbol{V}$ preserves $\boldsymbol{V}^\top\boldsymbol{c}$. Row orthonormality requires $\boldsymbol{V}_A\boldsymbol{V}_A^\top = \boldsymbol{A}^{-\top}\boldsymbol{A}^{-1} = \boldsymbol{I}$, hence $\boldsymbol{A} \in O(r)$. This does not exhaust joint network reparameterizations: scaling $\boldsymbol{c}$ and $\bar{\boldsymbol{z}}$ by a nonzero scalar α and dividing the trunk by α preserves the contraction and changes the latent metric. The certified orthogonal subgroup alone rules out an intrinsic axis interpretation.

A raw-parameter $GL(5)$ probe changes the forward output by less than 10^{-15} without orthonormalization and at least 1.7 with it under the specified reparameterization. QR-aware rotation compensation is certified in Appendix E.4.1. At rank five the affine, general-linear and orthogonal reparameterization groups have dimensions 30, 25 and 10, respectively.

The field-metric analysis does not need this restriction (Propositions 3 and 4). We tested both statements on the seven retained runs. We drew 140 invertible maps. They are anisotropic scalings with condition numbers up to 1000, shears, and products of the two. Each map reparameterizes a coefficient–decoder pair and keeps every decoded field fixed. Field varimax returned the same allocation and the same blocks on every draw. Varimax on the reparameterized stored coefficients kept its allocation on fewer than half of the draws (Table 9). The screened run s7 behaves alike.

Table 9: Identification after an invertible reparameterization of the coefficient–decoder pair, over the sampled maps on the seven retained runs. Angles compare a block after the reparameterization with the same block before it.

	field varimax	varimax, stored coefficients
residual of Proposition 3 in the coordinates	$< 10^{-13}$	–
draws that keep the $2/2/1$ allocation	140 of 140	66 of 140
block angle	$< 0.02^\circ$ (largest)	27.5° (median)
change of component recall	$< 10^{-4}$	–
factor between pairwise distances	1	up to 997

E Full predictive-audit results

Model-only and reference-initialized comparisons use the completion task of Appendix E.1. Appendix C defines the intervals. Appendix K tests the reproducibility and the sensitivity of the identification.

E.1 The block-shared predictor and the completion task

At fixed identified subspaces, Gaussian kernel ridge regression is fitted to the field-orthonormal coordinates of each block. For block g, C_g contains the observed coordinates as rows and K_g is the kernel matrix of the observed parameters. The prediction at parameter p is

$$\widehat{c}_g(p)^\top = k_g(p)^\top (K_g + \lambda I)^{-1} C_g, \tag{14}$$

where hats denote fitted or predicted quantities, I is the identity, $k_g(p)$ contains kernel values against the observed parameters, and $\lambda = 10^{-3}$ is the ridge penalty. The kernel is $k(p,q) = \exp\big(-\sum_i (p_i - q_i)^2/\ell_i^2\big)$ on inputs rescaled to $[-1,1]$, $Re \mapsto (Re - 350)/250$ and κ unchanged. Each length scale ℓ_i takes one of the six values 0.25, 0.5, 1, 2, 4 and 8, the same set for both inputs. The candidate grid holds the 36 pairs, ordered by ℓ_{Re} and then ℓ_κ. The ridge penalty is fixed, and a tie is resolved by the first pair in that order. One fitted kernel is shared by all coordinates of a block, and its two input length scales may differ. They minimize the leave-one-out (LOO) error summed over the block, which in field-orthonormal coordinates is the decoded-field error of the block: each observed field is predicted from the others. An internal rotation preserves the summed squared LOO residuals of every candidate kernel, and the candidate set and its deterministic tie-break are basis independent, so the kernel choice is unchanged. The predicted coordinates rotate with the basis, and decoding cancels the rotation. Appendix E.4 tests this invariance.

Completion predicts the rest of a parameter family from a few observed fields. A learned span is reused this way when new reference solutions are scarce. The 11×9 parameter grid defines 63 interior crosses. Each cross observes the 19 complete velocity fields on one row and one column of the grid, and the remaining 80 fields are the targets. Off-cross fields enter only the evaluation and cannot select bases, labels, subspaces or kernels. Reference-initialized comparisons use decompositions built from full-grid reference solutions (§2.3). In those comparisons only the kernel selection is restricted to the observed cross. Crosses share targets, so they are not independent replications (Appendix C).

Completion replaces the learned coefficient map with a fit to the observed fields, tuned on the observed cross alone. The matched full-span baseline uses the same Gaussian family, candidate grid and ridge penalty, with one length-scale pair for all five coordinates. Operator-response selection shares the block dimensions and tuning procedure with field varimax. All three use the same observations, frozen span and loss on the observed cross. An isotropic thin-plate interpolant provides

a second full-span kernel family, and the original frozen operator predicts with zero reference observations. The interpolant is the thin-plate spline of SciPy's `RBFInterpolator` with a linear polynomial term and no smoothing. It is fitted jointly to the five field-orthonormal coordinates on the same rescaled inputs and has no scale to select. The completion study ran in float64 on two threads, and the verification record gives the library versions of the environment in which the package reproduces it. The task requires combining responses seen only separately. A cross observes each parameter's variation once, and the off-cross targets lie where both vary.

E.2 Model-only identification

We recorded the model-only identification procedure in the repository before execution. The trained coefficient map on the fixed 11×9 parameter grid supplies the coefficient table. The decoder Gram matrix is computed on the velocity grid from its known coordinates, so no reference file is opened. Varimax and the labelling rule of §2.3 act on the field-orthonormal table. Block sizes are not imposed. Operator-response selection uses central differences of the same coordinates, which exclude the parameter-dependent lift and fixed latent mean by construction, and ten fixed-seed Powell starts. The maximizing start is chosen without reference errors. File-access guards forbid reference solutions and result caches throughout identification, and the choices for all eight checkpoints are saved before evaluation. The decoder grid is checked against the reference grid afterwards. Table 29 compares the identified subspaces across metrics, rules, runs and basis sources, and the experimental setup provides the frozen record and the commands for archived-statistic and checkpoint-level verification.

E.3 Prediction accuracy on held-out fields

Table 10 is the accuracy record on the targets of Appendix E.1 and the 100×100 spatial grid. We report relative error against reference solutions, $u_{\mathrm{rel}} = \|u_{\mathrm{pred}} - u_{\mathrm{ref}}\|_2 / \|u_{\mathrm{ref}}\|_2$, concatenating the velocities (u, v) and averaging over target fields, cross centers and seeds in that order. Repeated fields keep their cross multiplicity. Original predictions use the checkpoint's coefficient map, effective decoder and fixed field. Reference solutions enter completion fitting and evaluation after model-only identification is frozen.

Table 10: The accuracy record of the model-only protocol: relative error of five procedures on the same targets for the eight training seeds of one architecture. The original operator predicts with zero reference observations. The three Gaussian procedures and thin-plate use the observed fields of Appendix E.1, and the last uses a different kernel family. The last column projects each target's reference solution onto the rank-5 span and involves no prediction, so it is the floor of every method that predicts inside the span. Means are over the seven retained runs and over all eight.

		Gaussian, field-orthonormal coordinates				span
seed	original	field varimax	selection	full span	thin-plate	projection
observed fields	0	19	19	19	19	all
s1	0.14855	0.13919	0.13747	0.18900	0.12544	0.07597
s42	0.14798	0.13748	0.13874	0.19046	0.12554	0.07585
s314	0.14866	0.13986	0.13587	0.19056	0.12554	0.07599
s1729	0.14884	0.13707	0.13912	0.19051	0.12552	0.07613
s2026	0.14762	0.13826	0.13875	0.19087	0.12555	0.07606
s5150	0.14817	0.13703	0.13614	0.19120	0.12524	0.07552
s31337	0.14825	0.13806	0.13875	0.19163	0.12546	0.07589
mean, seven	0.14830	0.13814	0.13784	0.19061	0.12547	0.07592
s7 (screened)	0.15064	0.13294	0.13796	0.16738	0.12326	0.07144
mean, eight	0.14859	0.13749	0.13785	0.18770	0.12519	0.07536

Completion within the field-varimax subspaces improves on the original operator on every run and on the matched full-span Gaussian predictor by a wide margin. Operator-response selection reaches the same accuracy. Thin-plate interpolation of the full span gives the lowest error. The three Gaussian procedures share the observations, candidate kernels and context-only selection, so the block structure accounts for their difference.

Table 11: Paired original-minus-completion differences of the run means. Positive values favour completion. Intervals are two-sided 95% Student t intervals over the seven retained runs.

	seven retained runs			eight runs
completion procedure	mean	interval	positive	mean, positive
field varimax	$+0.0102$	$[+0.0092, +0.0111]$	7/7	$+0.0111$, 8/8
operator-response selection	$+0.0105$	$[+0.0091, +0.0119]$	7/7	$+0.0107$, 8/8
full-span Gaussian	-0.0423	$[-0.0433, -0.0413]$	0/7	-0.0391, 0/8
full-span thin-plate	$+0.0228$	$[+0.0224, +0.0232]$	7/7	$+0.0234$, 8/8

E.4 Within-block invariance of the specified predictors

E.4.1 Paired cavity field control

We test Equation (4) in the forward implementation on the seven retained cavity checkpoints. The coefficient table is centred and its mean absorbed into the fixed latent mean. For each checkpoint, 12 orthogonal draws are applied to separate model copies using the stored-weight reparameterization, including compensation for the QR sign convention. The paired control copies exactly the candidate's rotated coefficient table and retains the original modes. Thus both reparameterizations apply the same effective orthogonal change to the coefficients. Only the candidate compensates the decoder.

We evaluate velocity at $(Re, \kappa) = (150, -1), (350, 0), (550, 1), (275, -0.375)$ on a 17×17 spatial grid in $[0.025, 0.975]^2$, concatenate the fields, and use relative L^2 deviation in float64. Across 84 pairs, the compensated maximum is below 5×10^{-16}. The uncompensated deviations range from 0.175 to 1.214. Every pair satisfies the declared 10^{-12} reparameterization and 10^{-3} control thresholds. This checks the implemented orthogonal reparameterization. Orthonormality supplies the analytical restriction to $O(5)$ within the centred, metric-preserving coefficient–mode class with the surrounding representation fixed (Appendix D.5).

The downstream methods

For the predictor in Equation (14), use $\lambda = 10^{-3}$ and choose the Reynolds and asymmetry length scales ℓ_{Re}, ℓ_κ of block B by

$$(\hat{\ell}_{Re}, \hat{\ell}_\kappa)_B = \operatorname*{arg\,min}_{\ell_{Re}, \ell_\kappa} \sum_i \left\| \boldsymbol{d}_B(p_i) - \widehat{\boldsymbol{d}}_B^{(-i)}(p_i) \right\|_2^2, \tag{15}$$

where the superscript $(-i)$ omits observation i when fitting and the loss uses the block Euclidean norm of field-orthonormal coordinates. The block-index convention is that of Lemma 1. For $\boldsymbol{Q}_B \in O(|B|)$, the objective and tie-break are unchanged and $\widehat{(\boldsymbol{Q}_B \boldsymbol{d}_B)} = \boldsymbol{Q}_B \widehat{\boldsymbol{d}}_B$. A shared thin-plate regressor is unchanged by any internal rotation, as Lemma 1(ii) predicts. Its Burgers deviation is 5.6×10^{-17} (Burgers completion study, Table 5).

E.4.2 Internal rotations at fixed subspaces

Write $\boldsymbol{R}$ for the reference-initialized field-varimax axes. For the seven retained runs the group of internal rotations of the identified subspaces is $O(2) \times O(2) \times O(1)$. The group is the same for the model-only and the reference-initialized basis of a run, because both allocate $2/2/1$, and the intervention rotates the reference-initialized representative. The screened run s7 allocates one Reynolds, three asymmetry and one interaction axis, giving $O(1) \times O(3) \times O(1)$. For each run we fix the subspaces and sample 300 representatives $\boldsymbol{R}\boldsymbol{Q}_i$, with $\boldsymbol{Q}_i$ drawn from the Haar probability measure of that group. Stored matrices act on coordinate rows as $D \mapsto D\boldsymbol{R}\boldsymbol{Q}_i$ and on decoded columns as $\boldsymbol{\Phi}\boldsymbol{G}^{-1/2} \mapsto \boldsymbol{\Phi}\boldsymbol{G}^{-1/2}\boldsymbol{R}\boldsymbol{Q}_i$, so $(\boldsymbol{R}\boldsymbol{Q}_i)^\top (\boldsymbol{R}\boldsymbol{Q}_j) = \boldsymbol{Q}_i^\top \boldsymbol{Q}_j$ is block diagonal. The singleton axis has entry exactly ± 1 over all draws. The same audit around operator-response selection gives the same conclusions.

Direct decoded-prediction comparison. The intervention of §3.4 holds the identified subspaces of the reference-initialized basis fixed, rotates their internal bases and reruns the whole fitting and tuning procedure each time. The prediction change is the largest relative distance between decoded

predictions made in different internal bases, $\Delta_{\text{pred}} = \max_{Q \in \mathcal{S}_H} \|\widehat{\mathbf{u}}_Q - \widehat{\mathbf{u}}_I\|_2 / \max(\|\widehat{\mathbf{u}}_I\|_2, \varepsilon)$, where $\mathcal{S}_H$ holds the sampled internal rotations relative to the first representative and $\varepsilon = 10^{-30}$. It uses all eight sets of 300 representatives, including s7. For each representative and both tuning procedures, the test reruns kernel selection and predicts all 80 targets at each of the 63 cross centers. It decodes both velocity components on the 100×100 spatial grid, including the fixed mean and boundary lift. Target multiplicity across crosses is retained.

Per-coordinate deviations are 0.110–0.124. Block sharing reduces Δ_{pred} below 5×10^{-16} in every checkpoint. The prespecified float64 tolerance is 10^{-10} for block-shared tuning, and the per-coordinate control must exceed 10^{-6}. A separate calculation with the decoder Gram matrix checks the literal field-deviation norm to 10^{-10}. Figure 5 shows the completion-error variation at fixed subspaces. Per-coordinate standard deviations are 0.0027–0.0071 across the eight runs. Blockwise standard deviations are below 5×10^{-17}, with full ranges below 2×10^{-16}.

Reconstructing the varimax basis after a complete reparameterization of the coefficient–decoder pair is a separate matter, which Proposition 4 settles under its two conditions. A predictor that reconstructs the varimax basis and is equivariant under its signed permutations is another route to basis-independent predictions, which we do not test.

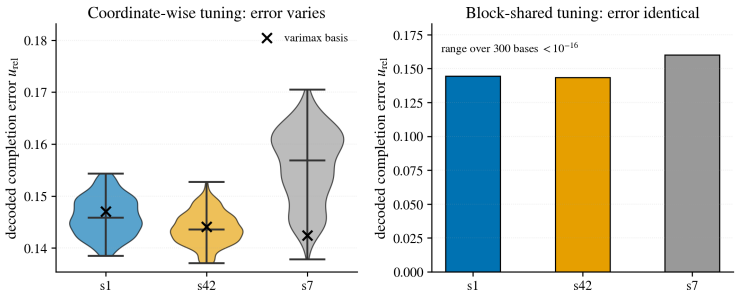


Figure 5: Block sharing reduces completion-error variation across internal bases to numerical precision. Reference-initialized field-varimax subspaces, 63 crosses. *Left:* per-coordinate errors over 300 internal bases per run (s.d.: standard deviation). Crosses mark the varimax basis. *Right:* sharing one kernel within each block makes the predictor invariant under internal rotations. Three runs are shown.

Comparison with random subspaces. Each run supplies 500 random labelled subspace tuples with the same dimensions and number of fitted length-scale pairs as the identified blocks. They are scored on the same 63 crosses with context-only kernel selection. On every retained run, field varimax and operator-response selection both have lower error than nearly all random tuples, with and without off-cross observations. On the screened run s7, a third of the random tuples have lower error than varimax. Table 30 compares the two rules directly, and the random-tuple selection protocol records the per-run values and a cross-fitted selector among the random tuples.

F THE READOUT COMPARISON ACROSS ARCHITECTURES

This appendix tests downstream basis dependence in features extracted from existing checkpoints (§2.6). It is an effective-readout basis study, separate from the stored-weight certificates counted in Appendix G. No model is trained for this probe. Besides the cavity and Burgers, it uses operators for the Graetz problem, steady convection–diffusion of heat in a fully developed channel flow with the Péclet number as parameter.

Feature location and producer. For both the plain PINN (Raissi et al., 2019) and parameter-conditioned PINN, the probe reads H immediately before the Fourier-modified trunk's final affine head, $u_{\text{raw}} = \boldsymbol{W}_{\text{out}} H + b_{\text{out}}$, with affine weights $\boldsymbol{W}_{\text{out}}$, bias b_{out} and pre-lift output u_{raw}. The producer is a gated nonlinear network: each hidden update has the form $H_\ell = Z_\ell \odot U + (1 - Z_\ell) \odot V$, with activated encoders U, V and activated gates Z_ℓ at layer ℓ; $\odot$ denotes elementwise multiplication. The formal compensation $H \mapsto QH$, $\boldsymbol{W}_{\text{out}} \mapsto \boldsymbol{W}_{\text{out}} Q^\top$, with $Q \in O(\text{width})$ and b_{out} fixed, preserves the effective readout. It does not establish that a stored-weight reparameterization of this producer can realize QH.

The experiment rotates the extracted feature matrix and leaves model weights unchanged. The branch–trunk rows likewise rotate extracted branch outputs. The Burgers row rotates low-rank coefficients. All four readout forms are therefore reported here as effective-basis comparisons, without an architectural $O(\text{width})$ claim for the PINN hidden layers. The cavity's compensated weight reparameterization is tested separately in Appendix E.4.1. Table 12 reports effect sizes and random-initialization controls for attribution to training.

Measurement and random-initialization control. For each predictor, we measure the relative error range across basis choices: $\Delta = (\max_W E_W - \min_W E_W)/|\operatorname{mean}_W E_W|$, where E_W is the coefficient-space L^2 prediction error. This error is $O(d)$-invariant for feature dimension d, so the test needs no reference solver. We compare shared and per-coordinate kernel selection, separately from the rank-5 labelled-block comparison in Appendix E.3. The results at $d = 100$ or 200 are not pooled with it.

Each configuration is evaluated with trained weights and 30 independent random initializations of the same architecture. All use the same 32 basis choices: the identity and 31 Haar draws. A discriminating result requires $\Delta_{\text{shared}} < 10^{-10} < \Delta_{\text{coord}}$. A random-initialization control is *inconclusive* when it also satisfies these inequalities and its per-coordinate error range agrees with the trained value within half a decade: the trained-to-random ratio lies between $10^{-0.5}$ and $10^{0.5}$ (approximately $[0.32, 3.16]$). The two-sided criterion tests agreement in magnitude. Agreement makes attribution to training inconclusive.

Table 12: The effective-readout study compares 56 configurations across three PDE families and a range of basis dimensions. "shared" reports the largest relative error range Δ_{shared}. Ratio is $\Delta^{\text{trained}}_{\text{coord}}/\Delta^{\text{random}}_{\text{coord}}$, reported as the range of configuration-level medians over 30 initializations. The last column is the fraction of inconclusive random-initialization controls; N_{cfg} counts configurations. The study includes mixture of experts (MoE) models. The plain PINN regresses over space. The other models regress over parameters. Their ratios are not directly comparable.

model	family	d	N_{cfg}	shared maximum	ratio (median range)	inconclusive
plain PINN, Fourier trunk	Graetz	16–128	13	2.0×10^{-15}	1.5–8.3	7–77%
parameter-conditioned PINN	cavity	128	6	3.2×10^{-15}	4.8–8.4	0%
branch–trunk, log-linear inputs	cavity	200	4	1.5×10^{-15}	317–504	0%
branch–trunk, inverse inputs	cavity	200	3	3.6×10^{-15}	1.9–2.7×10^3	0%
branch–trunk, linear inputs	cavity	200	5	3.2×10^{-14}	0.25–2.7	20–100%
branch–trunk, MoE	cavity	210	6	2.2×10^{-15}	528–1.1×10^3	0%
branch–trunk, MoE	Graetz	100	8	1.3×10^{-14}	10^{10}–10^{13}	0%
low-rank operator	Burgers	4	11	4.6×10^{-15}	1.8–2.6	57–90%

Shared tuning and training-dependent effects. The shared predictor's error range stays at rounding level in every configuration (Table 12), consistent with Observation 1. The per-coordinate predictor's relative error range is 1.5×10^{-4} to 5.3×10^{-1} in the discriminating configurations. Five model groups have no inconclusive controls among their random initializations (Table 12). Their ratios support attribution to training within each architecture. For the randomly initialized Graetz mixture of experts, the denominator is at rounding precision, so its ratio indicates only an order of magnitude.

Limits of attribution to training. Three model groups have inconclusive controls. For the branch–trunk cavity models at $d = 200$, the trunk, width, activation and basis dimension are fixed. The branch input transformation and resulting regression grid differ: log-linear on $[10, 1300]$, inverse-linear on $[90, 1100]$, and linear on $[90, 1100]$. Across these twelve configurations, the first two groups have ratios of at least $317\times$ on every draw. The linear-input group has inconclusive rates of 20–100% and ratios on both sides of 1. So the architecture, the input transformation and the grid all affect the control result.

For the plain PINN, twelve fully trained configurations (60,000 steps, four widths, three seeds) and one 500-step configuration give inconclusive rates of 7–77%, with median 37%. Fourier-feature networks have spatial variation even at initialization. These controls do not establish that training produced the measured basis dependence. The random-initialization control is inconclusive for these

rank-4 Burgers models as well. Rates are 57–90%, and $\Delta_{\text{shared}} \leq 4.6\times10^{-15}$. These measurements do not isolate rank as the cause. In all three groups, the effective-readout identity and Observation 1 still hold. The dependence itself needs no training. Per-coordinate fitting is not equivariant under a rotation of its inputs (Lemma 2). It depends on the basis of any feature space it is fitted on, trained or not. The control measures only the part of that dependence that training adds.

G AUDIT OF PUBLIC CHECKPOINTS

This appendix gives the checkpoint-level evidence for Section 3.1. The probes read public weights and use synthetic inputs. They require no training, benchmark data or reference solver. The five families are FNO, TFNO, U-Net, ConvNeXt U-Net and branch–trunk operators (Table 13). The deviation of Equation (3) uses the largest absolute entry in every probe except the ConvNeXt one, which was written separately for the held-out family and uses the Euclidean norm. Appendix F separately studies effective readout bases in our own models. Those activation rotations are excluded from the stored-weight certificate count. The discrete groups S_d and B_d denote permutations and signed permutations on d coordinates.

Table 13: The audit covers five families of public checkpoints. Each certificate pairs a reparameterization of the stored weights with a discriminating control. Of the certificates, 140 concern feature representations. The rest concern internal Tucker factors or an unread bias.

family	checkpoints	certificates	audited operations
FNO	15	45	Fourier blocks, projection head
TFNO	15	60	Fourier blocks, head, Tucker factors
U-Net	14	28	encoder and decoder blocks
DeepONet	4	8	Hadamard product, contraction
ConvNeXt U-Net	14	28	channel normalization, unused bias
total	62	169	ten transformation types

Scope of the certified groups. Maximality is proved within the declared reparameterization classes for 52 of the 169 pairs. The other 117 certify subgroups. These classes differ. Branch–trunk contractions constrain invertible linear head reparameterizations, and the last Fourier block admits every invertible affine channel map. The 14 unread-bias pairs are maximal within additive reparameterizations. Translations reach every bias value, and larger groups also preserve the output. The collapsed-feature diagnostic measures approximate invariance at a stated tolerance.

The full Tucker freedom and the conditioning of the sampled maps. The Tucker row certifies a subgroup of the known parameterization freedom. For a factor matrix $\mathbf{F}$ and core tensor $\mathcal{C}$, the compensated transformation is $\mathbf{F} \mapsto \mathbf{FA}$ and $\mathcal{C} \mapsto \mathbf{A}^{-1} \cdot \mathcal{C}$. The dot denotes multiplication on the core index associated with that factor. This transformation preserves the reconstructed weight for every invertible $\mathbf{A}$, giving $GL(r)$.

Restricting it to orthogonal maps would count 65,024 degrees of freedom in the channel factors. The general linear group has 131,072, a factor of 2.02 more. With this compensation the counted channel-factor probes stay at rounding level, and their uncompensated orthogonal controls change the output by order one (Table 14). The norm-preserving control isolates missing compensation. An ill-conditioned uncompensated map can amplify the output by orders of magnitude. The counted diagnostic uses orthogonal controls. The channel-factor subgroup leaves mode-axis freedom unaudited. The same algebra applies to every factor and raises 131,072 to 132,420 and 133,444 across the fifteen checkpoints. The multilinear rank matches the stored ranks on all fifteen. Full-rank Tucker factorizations are related by invertible factor maps over their scalar field. The stored factors are complex. The probes use real maps. The quoted dimensions count real reparameterization parameters. Complex invertible maps are additional algebraic freedoms outside this numerical audit.

The counted row stops at the channel factors because the all-factor real reparameterization, sampled like the other rows with maps of condition number 50, misses the 10^{-12} tolerance. That tolerance was fixed before the audit and we keep it. The miss comes from the conditioning of the sampled maps. Each checkpoint chains twenty compensations, and every one inverts a sampled map. We repeated the all-factor reparameterization with better-conditioned maps on the fifteen checkpoints (Table 15). With orthogonal maps, and with general linear maps of condition number 5, it clears the

Table 14: Ten transformation types yield paired certificates on 62 public checkpoints. $\epsilon_{\mathrm{reparam}}$ is the largest relative output deviation across the row's checkpoints; $\epsilon_{\mathrm{control}}$ gives the range of control deviations. Only reparameterizations of the stored weights count as certificates. Rows in the first group test transformations of feature representations, which leave latent axes ambiguous. Rows in the second group concern parameter non-identifiability. The Tucker factors are compensated inside their own weight, and the ConvNeXt bias is never read by its branch, so its reparameterization needs no compensation. The separate readout-basis tests rotate extracted activations. The Fourier-mode tests reject a transposition. The collapsed-feature diagnostic reports approximate invariance at a stated tolerance. Maximality is relative to the declared reparameterization class, which is additive for the unread bias.

readout location	certified transformation	named control	count	$\epsilon_{\mathrm{reparam}}$	$\epsilon_{\mathrm{control}}$	analytical status
feature representations						
FNO/TFNO last block	affine $GL(128) \ltimes \mathbb{R}^{128}$	the same reparameterization one block earlier, where GELU applies	30	3.9×10^{-14}	3.02–23.14	**maximal**
FNO/TFNO block before it	unsigned S_{128}	a rotation in $O(128)$	30	2.5×10^{-15}	0.120–3.16	subgroup
FNO/TFNO projection head	unsigned S_{256}	signed B_{256}, since GELU is not odd; a rotation in $O(256)$, drawn in Figure 2	30	3.2×10^{-15}	0.0905–3.87 0.054–6.59	subgroup
U-Net `decoder1`	signed B_{48}	generic $O(48)$ on the unfrozen block	14	2.2×10^{-15}	0.0292–2.45	subgroup
U-Net `encoder1`	unsigned S_{48}	signed B_{48}, the group `decoder1` admits	14	5.2×10^{-14}	0.0276–3.66	subgroup
ConvNeXt decoder stage	unsigned S_{84}	the same permutation carrying signs, which channel-mean subtraction blocks	14	1.6×10^{-15}	0.622–7.46	subgroup
DeepONet stage-1 Hadamard	monomial, independent diagonals	generic $O(256)$	4	4.9×10^{-14}	1.44–2.49	**maximal**
DeepONet stage-2 contraction	$GL(256)^2$ as $(W, W^{-\top})$	generic $O(512)$	4	1.2×10^{-14}	0.851–1.15	**maximal**
feature-representation pairs			140			
parameter non-identifiability						
TFNO Tucker factors	$GL(128)$ in each channel factor	the factor rotated without compensating the core	15	7.6×10^{-14}	0.834–2.26	subgroup
ConvNeXt resample norm bias, unread	translation $\mathbb{R}^{84}$	the same module's weight, which that branch reads	14	0	0.0571–0.84	**maximal** within additive reparameterizations
parameter pairs			29			
weight-space pairs			**169**	$\leq 7.6 \times 10^{-14}$	≥ 0.027	
readout basis, FNO/TFNO and U-Net	one global $O(r)$	–	44	4.8×10^{-15}	0.818–9.03	not counted: identity
FNO/TFNO Fourier mode axis	transposition rejected	–	30	–	0.160–0.76	not counted: rejected
U-Net collapsed features	tolerance-dependent invariance	–	1	5.5×10^{-15}	2.22–2.22	tolerance-dependent

tolerance on every checkpoint by about two orders of magnitude. The same maps with the wrong compensation change the output by order one or more. So the larger group is certified numerically at these draws, and the counted row stays as registered.

Table 15: All-factor Tucker reparameterization on the fifteen TFNO checkpoints: largest relative output deviation of the compensated reparameterization, number of checkpoints under the 10^{-12} tolerance, and smallest deviation of the same maps with the wrong compensation.

condition number of the sampled maps	reparameterization	under the tolerance	wrong compensation
1 (orthogonal)	1.4×10^{-14}	15 of 15	≥ 0.92
5	1.9×10^{-14}	15 of 15	$\geq 6.9 \times 10^{13}$
50 (as in the audit)	1.4×10^{-11}	1 of 15	$\geq 4.2 \times 10^{40}$

The eight branch–trunk certificates have been replayed with the original code's 10^{-12} reparameterization and 10^{-3} control thresholds explicitly recorded, together with inputs, precision, random draws and checkpoint digests. The earlier records, which omitted tolerances, remain archived as historical evidence. The replay records the settings used for the subsequent verification.

Normalization in the held-out ConvNeXt family. The public ConvNeXt implementation uses one `LayerNorm` class for two behaviours, and only its `channels_last` branch subtracts the mean over channels. Mean subtraction blocks a sign flip. The decoder block admits unsigned permutations. The resample admits signs, and the `channels_first` branch never reads its stored bias. So the forward computation does not depend on this bias. Additive reparameterizations $b \mapsto b + d$ reach every equivalent bias value. The reported translation group is scoped to that reparameterization class. The larger affine action $b \mapsto Ab + d$ for any $A \in GL(84)$ also preserves the output. The numerical pairs establish the reported translations and their controls. The complete non-identifiability of the bias value follows from the unread tensor in the forward computation.

Derivation for the public branch–trunk operators. The four public DeepONet checkpoints come from a flow-past-object benchmark (Rabeh et al., 2025; 2026; BGLab, 2025). Their forward code supports the architectural and downstream checks in stages 1 and 3 of Table 2. We do not perform physical-response validation on these checkpoints.

The first head pair produces arrays g, ℓ of shape $(256, 128)$ and combines them as $g \odot \ell$ without summing over the mode index. A shared permutation of the modes can be compensated in the downstream input columns. Independent diagonal maps also pass through the product: $(\mathbf{D}_1 g) \odot (\mathbf{D}_2 \ell)$ scales it by $\mathbf{D}_1\mathbf{D}_2$, which the next stage absorbs. These reparameterizations give a monomial group with 512 continuous degrees of freedom. A shared sign flip on both heads tests only a discrete subgroup.

The second head pair has shape $(512, 128)$ and contracts over the mode index while retaining output channels. For branch and trunk vectors b, t, the action $(b, t) \mapsto (\boldsymbol{W}b, \boldsymbol{W}^{-\top}t)$ preserves channel c with coordinate projector $\boldsymbol{P}_c$ precisely when $\boldsymbol{W}^\top \boldsymbol{P}_c \boldsymbol{W}^{-\top} = \boldsymbol{P}_c$. Thus $\boldsymbol{W}^\top$ must commute with every channel projector and preserve each channel block. Both heads are unnormalized, giving $GL(256)^2$ with dimension 131,072. Compensated scaling by $(2\boldsymbol{I}, \frac{1}{2}\boldsymbol{I})$ is one example. A dimension-based hypothesis would assign $O(512)$ with dimension 130,816. The dimensions differ by less than 0.2%, but each group contains transformations absent from the other. The paired controls distinguish them.

For stage 3, the benchmark evaluates field error. This procedure reads the decoded output and is the same in every basis by Observation 1.

Each reported reparameterization requires a discriminating control. For orthogonal $\boldsymbol{W}$, $(\boldsymbol{W}b) \cdot (\boldsymbol{W}t) = b \cdot t$ is an algebraic identity for *any* b and t. Applying $\boldsymbol{W}$ to stored weights and re-evaluating the readout preserves the output within float64 rounding error on the trained checkpoints (1.9 to 3.5×10^{-15}), on randomly initialized weights of the same shapes (3.2×10^{-15}) and on all-ones matrices (8.8×10^{-16}). These measurements verify the implemented contraction identity within floating-point error. They motivate the random-initialization comparison in Appendix F. To test the architectural group here, we pair each compensated reparameterization with its specified control. Across all four checkpoints, the stage-1 monomial reparameterization and the stage-2 $GL(256)^2$ reparameterization preserve the field to rounding level, and generic orthogonal maps on the same heads change it by order one (Table 14). The two measurements jointly test the proposed group, with the outcomes of Algorithm 1. Degeneracy, a weak control or a larger admissible group can each make a pair inconclusive, and negligible deviation alone does not establish maximality.

Extension to Fourier and U-Net families. The four branch–trunk checkpoints above establish the group on one family, which depends on its specific forward computation. We repeat the derivation on 58 further public checkpoints, 15 FNOs, 15 TFNOs, 14 U-Nets and, held out until the audit code for the other families was frozen, 14 ConvNeXt U-Nets, all trained by others on a physics-simulation benchmark (Ohana et al., 2024; Polymathic AI, 2025). Across all 62 checkpoints this yields 169 discriminating pairs, counting only those whose candidate transformation is a reparameterization of that checkpoint's weights. Each reparameterization gives deviation below 10^{-13}. Its control changes the decoded field by at least 0.027 in relative error. The deviations differ by more than eleven orders of magnitude.

Those 169 pairs are ten distinct certified transformations replayed across checkpoints. Of them, 102 assert a discrete group of dimension zero, 52 continuous parameter transformations outside the Tucker factors, and 15 the continuous Tucker factor freedom. The discrete rows distinguish admissible permutations from rejected rotations. The continuous rows include both representation

transformations and parameter non-identifiability, as the table note specifies. These tests repeat the ten group identifications across checkpoints and architectures.

We report reparameterization deviations as bounds because rounding errors vary with the thread count of the Basic Linear Algebra Subprograms (BLAS) library. The closest reparameterization and control deviations differ by eleven orders of magnitude. Tested thread-count changes do not alter any certificate status. We exclude 44 basis rows from the count because their transformation is not a reparameterization of the checkpoint's weights. It rotates the readout's input activations and compensates in a matrix constructed for the probe, $\boldsymbol{W}_1\boldsymbol{W}^\top\boldsymbol{W}H = \boldsymbol{W}_1H$, leaving every stored tensor untouched. Every counted transformation is an architectural identity of the forward computation. Repeating it with random matrices checks that identity independently of training. The basis rows establish that the checkpoint's readout shares one tensor across grid points. A further 30 rows reject the tested Fourier-mode transposition. The collapsed-feature diagnostic below is reported separately. The probes replay the relevant forward computations without the original training pipelines. DeepONet and ConvNeXt probes use PyTorch. Reimplementation exposed two library behaviours that affect the tests. A factorized-tensor view is rebuilt on every attribute access, so in-place writes were silently discarded. A fixed single-precision output limited numerical agreement to 10^{-7}.

Checkpoint enumeration and provenance. We include every published checkpoint of the four audited families at the physics benchmark's pinned revision: 15 FNOs, 15 TFNOs, 14 classic U-Nets and 14 ConvNeXt U-Nets. The classic U-Net release omits one dataset. The second source supplies four time-dependent branch–trunk checkpoints with 1, 4, 8 or 16 branch inputs. Together these give 62 checkpoints. None was excluded based on its audit result.

We held out ConvNeXt until the audit code for the other families was frozen at a recorded commit. This test required an additional rule (§3.1). The complete audit ran as one process on one CPU node in about 135 minutes. Median per-checkpoint times were 0.5 seconds for ConvNeXt, 3.9 for classic U-Nets and 47.4 for the Fourier family, whose times ranged from 21 to 981 seconds. The four DeepONet checkpoints averaged 3.3 seconds. These times exclude the analyst's derivation of each transformation.

Contrasting operations within one checkpoint. FNO blocks 2 and 3 have tensors of identical shape and certified groups of dimensions 0 and $16{,}512$. The public model omits the optional per-block multilayer perceptron (MLP). Its forward condition leaves the last block without an activation. That block sums three unnormalized linear paths, which can realize any invertible channel map. The projection head compensates the map and absorbs translations through its bias. The resulting affine group $GL(128) \ltimes \mathbb{R}^{128}$ gives output deviations below 10^{-13}. Its orthogonal subgroup has dimension $8{,}128$.

At block 2, unsigned permutations give deviations below 10^{-14}, rotations give 0.12–3.16, and signed permutations give 0.14–2.89. Diagonal scaling also changes the output. On two of 30 checkpoints, however, GELU is nearly scale-equivariant at the observed activations: the deviation remains near 10^{-3} even for a scale factor of four. The scaling control is inconclusive on those two checkpoints. Both remain in the comparison.

Activation parity and downstream constraints. Elementwise GELU commutes with permutations. Sign flips fail this condition. Odd $\tanh$ admits sign flips, provided later operations can also absorb them. The U-Net calls the same `_block` function at nine locations with different consumers. At `decoder1`, the linear readout absorbs signed permutations exactly. At `encoder1`, max-pooling permits unsigned permutations and rejects sign flips because $\max(-a, -b) = -\min(a, b)$. The output also feeds a skip concatenation. Compensating a permutation in only one of these consumers changes the field by at least 0.037. The reparameterization includes the batch normalization buffers: bias and running mean carry the sign. Weight and running variance retain their original signs. The signed-permutation control at `encoder1` has the smallest measured deviation, 0.027 on one checkpoint. Max-pooling causes this failure. At `decoder1`, the generic-rotation control reaches 0.029 on one checkpoint. Near-linear $\tanh$ on the scale supplied by batch normalization can help explain this small rotation deviation. A separate positive-diagonal control on the unfrozen decoder block changes the output on that checkpoint by 0.36 and never less than 0.27 on any of the fourteen checkpoints.

Approximate invariance from numerically inactive features. In `UNetClassic-gray_scott_reaction_diffusion`, 28 of the 48 `decoder1` features have collapsed: their batch normalization scales lie between 10^{-33} and 10^{-13}. Their running variances are of order one hundred. These channels are numerically inactive. Their 28 readout rows admit large changes with negligible output deviation: filling every one with noise, compensating only a constant shift of the output bias, changes the field by at most 5.5×10^{-15}. The same replacement on active rows gives deviation 2.2. The feature scales vary continuously, so the number classified as inactive depends on the threshold. Increasing the inactive-feature threshold above 10^{-13} makes the reparameterization exceed the reported bound. At 10^{-11} it is rejected. So the measured freedom depends on the numerical tolerance. The generic-rotation control requires inversion of the batch-normalization scale and becomes inconclusive on nearly inactive features. The active–inactive comparison distinguishes this degeneracy from the tested architectural symmetry (Figure 6).

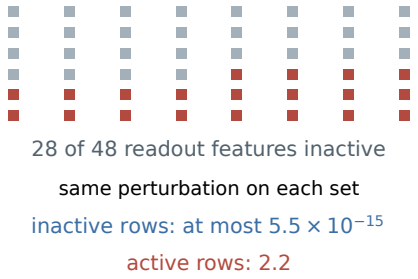


Figure 6: The inactive-feature diagnostic on one public U-Net. The same readout perturbation leaves the output unchanged on the numerically inactive features (grey) and changes it on the active ones (red). It is reported apart from the certificates because its flat arm reflects the feature scales of this checkpoint.

Channel importance on a real snapshot. The same checkpoint was run on a validation trajectory of The Well after its input format was checked against wrong-format controls. The readout-weight score of a `decoder1` channel is the Euclidean norm of its row in the 1×1 readout. The activity score multiplies that norm by the standard deviation of the channel's feature map over the snapshot. Features are inactive at the normalization-scale threshold of 10^{-13} used above. Across 16 certified reparameterizations, 8 redraws of the inactive readout rows with the output bias compensated and 8 signed permutations of the features, the prediction moves by at most 1.3×10^{-14}. After each redraw the ten channels ranked first by readout weight are inactive and their feature maps are blank, while the activity scores change by at most 1.7×10^{-14} (Figure 7).

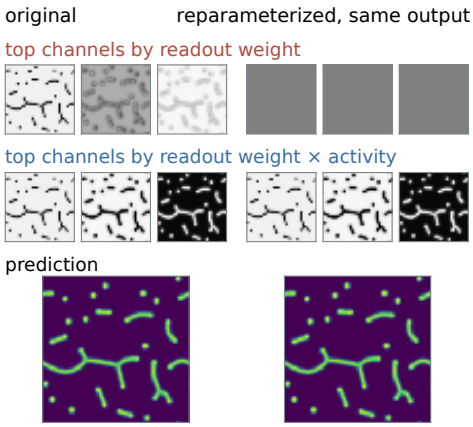


Figure 7: Channel importance on a public U-Net and a real snapshot. A ranking by readout weight depends on rows whose changes leave the prediction unchanged within the tested tolerance. A ranking weighted by feature activity is unchanged to the same tolerance.

Fourier-mode tests and the scope of the external audit. The tested Fourier-mode transposition changes the output well above tolerance (Table 14). In the contraction `'bixy,ioxy->boxy'`, the mode index is retained and each mode has its own channel-mixing matrix. Analysis and synthesis use the parameter-free `rfftn` and `irfftn` operations. So the tested reparameterization cannot compensate a mode transposition through stored weights on those operations. The Tucker variant admits one further certified transformation, a general invertible map inside a channel factor compensated in the core (Table 14). This parameterization freedom leaves every activation unchanged in exact arithmetic. The numerical deviations are reported above. A method reading those activations is therefore invariant in exact arithmetic. Stage 3 applies to all 62 checkpoints. The Well's `validation_metric_suite` evaluates root mean squared error, normalized and variance-scaled errors, maximum absolute error, binned spectral error and Pearson correlation. Each reads only the decoded field pair, so Observation 1 guarantees invariance. The external study covers stages 1 and 3. Physical identification and validation are assessed separately in our controlled families.

H Reference-solution structure and decoded-block alignment

We evaluate physical labels against reference solutions for the cavity and exact Cole–Hopf solutions for Burgers. Functional ANOVA defines parameter responses in field space, independently of latent-axis labels. Local tangent analysis then tests derivative alignment, and an operator truncation and a reference-POD rank sweep examine the dependence on the retained dimension. All field quantities use the stated lift subtraction, spatial metric and empirical parameter measure.

H.1 Reference components and decoded-block recall

We apply the functional decomposition of Equation (5) to each reference family, substituting that family's two parameters for Reynolds and asymmetry. Its components are orthogonal in the product of parameter and field spaces, verified to 10^{-16}. Table 16 reports their variance shares and effective ranks. The field-L^2 interaction share is small in both families (Table 16). The cavity asymmetry component has effective rank 3 at 99% energy, with leading energy shares $0.94/0.05/0.01$. The Reynolds component has effective rank 2. A two-dimensional asymmetry block therefore omits only about one percent of that component's energy if optimally aligned.

H.1.1 Component recall

We compute decoded-block recall $R_{g\leftarrow h}$ using Equation (6) for the model-only blocks used by the predictor. Each block is orthonormalized separately when forming its projector, which preserves its span. Table 17 gives the full recall matrices of s1 for varimax applied to the stored coefficients and to field-orthonormal coordinates. In the stored coefficients the Reynolds and interaction blocks capture much of the asymmetry component besides the asymmetry block. A component's recalls over the blocks then exceed its capture by the span. In field-orthonormal coordinates the blocks are orthogonal in the decoded field, so each component is divided among the blocks. The asymmetry component lies in the asymmetry block, and the interaction block holds more of the interaction component. Table 29 gives the ranges over the seven retained runs for both rules and metrics, and the basis-source comparison. Decoder Gram matrices and projected response covariances suffice to reconstruct every recall value. The experimental setup links this verification.

The same metric improves local tangent capture. On the 99 reference nodes of the retained runs s1 and s42, the asymmetry-labelled field-varimax block captures nearly all of the asymmetry tangents that the operator span captures, close to rank-matched snapshot POD. The Reynolds-labelled block captures more of the Reynolds tangents than the POD Reynolds block.

The endpoint response maps of Figure 3(a) project the lift-subtracted reference asymmetry component at $\kappa = +1$ onto each block and report its captured energy share (§2.5). Component recall aggregates the same projection over every value of κ. Over the seven retained runs, the share captured by the Reynolds and interaction blocks is at most about one percent in field-orthonormal coordinates and between a sixth and three quarters in the stored coefficients. The asymmetry block keeps the same share in both. For the Reynolds response at $Re = 600$, the field-varimax Reynolds block

captures the same share on every run, and more than the stored-coefficient Reynolds block on five of the seven.

Table 16: Field-level ANOVA gives reference variance shares in the field-L^2 metric and component SVD ranks at 99% energy. Burgers parameters are viscosity ν and initial amplitude a.

family	main-1	main-2	int.	ranks
cavity	5.3%	91.9%	2.7%	$2/3/\geq 2$
Burgers	4.0%	95.4%	0.6%	$2/1/\geq 2$

Table 17: Model-only decoded-block recall for s1, varimax applied to the stored coefficients and to field-orthonormal coordinates. Rows are blocks and columns reference components. The last row is energy outside the span.

	stored coefficients			field-orthonormal		
block \ component	Re	κ	int	Re	κ	int
Re	78%	57%	17%	74%	0%	11%
κ	24%	97%	16%	20%	97%	16%
int	19%	72%	19%	0%	1%	41%
outside	6%	2%	32%	6%	2%	32%

H.1.2 SVD AXES AS A SECOND POST-HOC RULE

SVD mode extraction orients a learned basis by energy (Venturi & Casey, 2023; Peyvan et al., 2024), and it can be applied after training. The SVD of an operator's centred correction fields over the parameter grid equals the SVD of its centred field-orthonormal table. So SVD and field varimax orient the same span, and we label the SVD axes by the same rule (Table 18). In field-orthonormal coordinates the SVD blocks are orthogonal in the decoded field and agree across runs as closely as the field-varimax blocks. The leading SVD axis carries most of the energy and nearly coincides with the field-varimax asymmetry subspace. Two of the other axes mix the two parameters, so the Reynolds-labelled SVD block recalls less of the Reynolds component than the field-varimax block. On the stored coefficients the SVD changes its allocation and its blocks between runs. The audit and the identification apply unchanged to a model whose modes were orthonormalized by an SVD during training.

Table 18: SVD axes of the stored coefficients and of field-orthonormal coordinates beside field varimax in the model-only basis, on the seven retained runs and with the registered labelling rule. Recall is the component recall of the block that carries the component's label. Double counting is the largest excess of a component's recall, summed over blocks, over its capture by the span. Between-run angles are mean principal angles between blocks with the same label in two runs, where their sizes match. The last two rows compare SVD blocks with the field-varimax blocks of the same run.

	SVD, stored	SVD, field	field varimax
allocation (Re/κ/int)	$1/2/2$ or $2/2/1$	$2/2/1$	$2/2/1$
Reynolds recall (%)	7–73	58–61	73–74
asymmetry recall (%)	95–96	95–96	97
interaction recall (%)	17–28	42–43	41–42
double counting (points)	152	0	0
between runs, Reynolds (deg)	26.7–64.7	0.8–2.5	0.8–2.2
between runs, asymmetry (deg)	0.8–43.6	0.4–1.4	0.3–0.8
between runs, interaction (deg)	7.0–20.3	0.8–2.0	1.0–2.2
leading SVD axis to varimax asymmetry block (deg)	–	within 0.6	–
SVD Reynolds block to varimax Reynolds block (deg)	–	33–35	–

H.1.3 WHOLE-REPRESENTATION EMBEDDINGS

Table 19 summarizes the embeddings of §2.5 on the seven retained operators. A t-SNE layout depends on the implementation, so the embedding study records its own environment: scikit-learn 1.8.0, NumPy 2.3.5 and SciPy 1.16.3. t-SNE uses perplexity 30, automatic learning rate, 1000 iterations and Euclidean distances. Points enter in a fixed order (Re outer, κ inner), the displayed embeddings use principal component analysis (PCA) initialization with random state 0, and they are shown without alignment. Rotation overlap compares the embedding in the field-varimax basis with the embedding in a Haar-rotated basis that decodes the same fields (median over eight rotations per run, drawn with seed 20260917 plus the run index). Rerun overlap compares PCA initialization with random initializations 1 and 2. Both use the 10-neighbour Jaccard index on the 41×33

grid. Reference-neighbour recall and the Spearman correlation between embedding and reference-solution distances use the 99 reference nodes and average three initializations. A layout residual is the largest coordinate difference between two embeddings after orthogonal Procrustes alignment, relative to the root-mean-square radius. We compute it between the two bases over all rotations, and between PCA and random initialization.

Recomputing the displayed embeddings of Figure 4 from their stored inputs with the same settings reproduces them with a largest absolute difference of 0. The markers of that figure score each of the 99 reference nodes on the displayed embedding. The standardized embedding misplaces 58 nodes in the varimax basis and 2 in the rotated basis, and the field-orthonormal embedding misplaces none. We compare neighbourhoods and draw no conclusion about the dimension or topology of the coefficient set. Under an input-basis rotation, field-metric distances are identical to a relative difference below 10^{-8}. For stored-coefficient and field-metric inputs the largest rotation layout residual lies below the smallest rerun layout residual on seven of seven runs. We include the standardized input because it is the usual preprocessing before an embedding, and the comparison shows what that habit costs. Standardized field-varimax coordinates change their neighbourhoods under rotation far more than between t-SNE runs. They still keep only a median 0.76–0.80 of their neighbours under rotations inside the identified blocks. A t-SNE study of a public U-Net's features showed no such benefit.

The field-metric embedding is laid out by κ, which carries most of the field energy. The Reynolds number still varies smoothly along each strand, and a nearest-neighbour regression on the layout recovers it on every run. No direction of the layout is shared by all strands, so its share of the layout as a whole is close to zero (Table 19). To show both parameters across the layout we also embed *block-normalized* coordinates, in which each identified block of field-varimax coordinates is scaled to unit mean square. Their distances depend on the identified subspaces and not on the axes inside them. This input gives the Reynolds number a visible share of the layout. It also stays consistent when we rotate the basis and identify the blocks again, which per-axis standardization does not. It loses the neighbours of the reference solutions, as standardization does, because its distances are no longer distances between fields.

Table 19: Embedding statistics over the seven retained cavity operators (ranges over runs). Overlaps and layout residuals under rotation and between initializations assess consistency. Recall and correlation assess agreement with the reference solutions. The last two columns give the main-effect share of the Reynolds number in the layout and the R^2 of a nearest-neighbour regression of it on the layout. For the block-normalized input the blocks are identified again after each rotation.

	neighbour overlap		layout residual		reference	distance	Re in the layout	
input	rotation	rerun	rotation	rerun	recall	correlation	share	R^2
stored coefficients	0.99	0.95–0.96	0.00–0.05	0.06–0.34	0.59–0.64	0.84–0.87	0.15–0.17	1.00
standardized, varimax basis	0.40–0.48	0.92–0.98	0.54–1.86	0.05–1.69	0.50–0.51	0.81	0.20–0.28	1.00
field-orthonormal	0.98–0.99	0.98	0.03–0.24	0.11–0.57	0.94	0.75	0.00	0.98
block-normalized	0.98–0.99	0.96–0.98	0.02–0.11	0.05–0.27	0.49	0.79	0.34	1.00

Input distances before any embedding. Reference-neighbour recall of an embedding mixes the accuracy of the model, the match between the input's distances and the fields, and the neighbours that t-SNE keeps. Table 1 isolates the second by scoring each input's own distance matrix at the reference nodes. The field input recalls the reference neighbours about as well before the embedding as after it, and its distances follow the reference distances more closely than any other input's. The lower correlation of its t-SNE layout in Table 19 is an effect of the embedding, and it disappears with PCA, with metric MDS and at a larger perplexity (Table 21). The table also explains the standardized input. In the field-varimax basis the per-axis variances differ by more than two orders of magnitude. Standardization gives the low-energy Reynolds and interaction axes the weight of the asymmetry axes, and the distances stop following the fields. A generic rotation mixes the axes and brings their variances closer together, so the same standardization changes the distances less. In every rotated basis of every run, the standardized input recalls more reference neighbours than in the varimax basis. Its recall still varies widely from one rotation to the next. The eleven Burgers operators of Appendix I, scored against their exact reference solutions, show the same ordering and the same mechanism (Table 20); their block-normalized input uses the cavity labelling threshold.

Table 20: The inputs of Table 1 on the eleven Burgers operators, scored against their exact references (ranges over operators; columns as in that table).

input	input recall	input correlation	variance ratio	embedding recall
stored coefficients	0.46–0.57	0.79–0.86		0.45–0.55
standardized, varimax basis	0.40–0.44	0.60–0.62	1699–1998	0.40–0.45
standardized, rotated bases	0.51–0.92		2–1687	0.50–0.91
field-orthonormal	0.90–0.93	0.99–1.00		0.90–0.93
block-normalized	0.44–0.45	0.64–0.68		0.44

Table 21: The ordering of the inputs does not depend on the embedding method. Reference recall of three inputs and median neighbour overlap between the varimax basis and the eight rotated bases, and Spearman correlation of layout distances with reference-solution distances, for PCA to two dimensions, metric MDS and t-SNE at three perplexities (ranges over runs). The field input has the highest recall and the higher overlap for every method on every retained cavity run and every Burgers operator.

		reference recall			rotation overlap		rank correlation of distances	
family	method	field	stored coefficients	standardized, varimax basis	field	standardized, varimax basis	field	stored coefficients
cavity	PCA	0.92–0.93	0.51–0.57	0.33–0.34	1.00	0.17–0.26	0.94	0.74–0.79
	metric MDS	0.93–0.94	0.55–0.57	0.43–0.44	1.00	0.31–0.39	0.94	0.77–0.79
	t-SNE, perplexity 10	0.91–0.92	0.75–0.79	0.59–0.60	0.96–0.99	0.39–0.46	0.52–0.57	0.75–0.76
	t-SNE, perplexity 30	0.94	0.59–0.64	0.50–0.51	0.98	0.40–0.48	0.75	0.84–0.87
	t-SNE, perplexity 50	0.94	0.58–0.62	0.50–0.51	0.98–0.99	0.40–0.49	0.90	0.84–0.87
Burgers	PCA	0.90–0.92	0.32–0.66	0.19–0.37	0.99–1.00	0.16–0.32	0.99–1.00	0.79–0.89
	metric MDS	0.90–0.93	0.49–0.60	0.37–0.41	1.00	0.27–0.40	0.99–1.00	0.81–0.87
	t-SNE, perplexity 10	0.94	0.50–0.60	0.43–0.47	0.98–1.00	0.30–0.45	0.57–0.62	0.67–0.81
	t-SNE, perplexity 30	0.90–0.93	0.45–0.55	0.40–0.45	0.95–0.98	0.29–0.46	0.82–0.90	0.81–0.87
	t-SNE, perplexity 50	0.88–0.92	0.45–0.55	0.40–0.45	0.97–0.99	0.29–0.45	0.94–0.97	0.81–0.87

H.2 Reference-derivative alignment and selection objectives

Two objectives use parameter derivatives in field-orthonormal coordinates and differ in their response source (Table 22). Denote their returned rotations by $\boldsymbol{W}_{\text{phys-diff}}$ (reference tangents) and $\boldsymbol{W}^{\text{op}}_{\text{coef-diff}}$ (the trained model's tangents). The allocation-free measure uses operator-independent local tangents $t_j(q) = \partial u/\partial p_j$ (FEM central differences) and the *physical* factor subspaces S_j (top-k_j SVD of the ANOVA components, with k_j chosen to capture 99% of component energy), with no fitted basis. For any field subspace S, P_S denotes its orthogonal projector under the field metric. The physical conditional attribution is $A_j^{\text{phys}} = \|P_{S_j} t_j\|^2 / \|P_{S_1+S_2} t_j\|^2$. It averages 90% for the Reynolds tangent and 100% for the asymmetry tangent, with minima of 80% and 97%. Each value is a share of its own tangent, so the two need not add to one. The tangent directions vary across the box. Their effective ranks are 7 for Reynolds and 5 for asymmetry, so the Reynolds tangent family exceeds the rank-5 representation.

Table 22: The derivative-based objectives share the field metric and differ in their response source.

objective	response source	metric
physical-response	reference solver	decoded field
operator-response	trained model	decoded field

For a fitted full span S and its identified subspaces S_j^{fit}, tangent alignment $T_j = \|P_{S_j^{\text{fit}}} t_j\|^2 / \|t_j\|^2$ factors as span capture $C_j = \|P_S t_j\|^2 / \|t_j\|^2$ times conditional attribution $A_j = T_j / C_j$ when $C_j > 0$. Means over points and physical parameters are denoted C and A; $C \approx 87\%$ for the rank-5 operator span, and a rank-8 reference-POD span captures 96–99% of the same tangents. We optimize each objective directly over $O(5)$ (Powell over the $\mathfrak{so}(5)$ exponential map, where $\mathfrak{so}(r)$ is the space of real skew-symmetric matrices, best of 8–12 restarts). The variance-based purity statistic is the mean, over coordinates, of the largest Reynolds, κ or interaction ANOVA share of field energy. Optimizing it gives $\boldsymbol{W}_{\text{agg}}$ with the highest purity and low derivative alignment, and $\boldsymbol{W}_{\text{phys-diff}}$ reaches $A = 90\%$ at lower purity. Field varimax lies near both optima (Figure 8). The weighted objective $\arg\max_{\boldsymbol{W}} [\omega\,\text{purity} + (1-\omega)A]$, with $\omega \in [0, 1]$, interpolates smoothly between the two optima. The variance and derivative objectives select different decompositions

of largely separable physical responses, and the difference vanishes on the separable Allen–Cahn control (Appendix H.3). Proposition 1 establishes that such a difference is possible.

Null distributions and held-out validation. Scoring random rotations against varimax's block labels gives a median attribution of 39–40% over 2000 draws per run. Letting each random rotation assign labels by the varimax rule raises the median to 59–63%, with a 95th percentile of 73–77%. Field varimax lies at the 99th to 100th percentile of both nulls on all eight runs. Fitting $\boldsymbol{W}_{\text{phys-diff}}$ on half the interior grid and evaluating it on the remaining tangents over 20 random $50/50$ splits gives $89 \pm 1\%$ held-out and $90 \pm 1\%$ fitted attribution. Field varimax reaches $84 \pm 1\%$ on held-out tangents, against $39 \pm 3\%$ and $62 \pm 2\%$ for the two random controls. Allen–Cahn gives 98% in every case.

Local consistency across the box. The attributions above average over the parameter box. A basis with a high-frequency cross-parameter component can have high variance-based purity and a nonzero cross-parameter derivative (Proposition 1). Per grid point on s1, field varimax attributes a mean of 70.5% of the Reynolds tangent to its Reynolds block (fifth percentile 44.7%) and 97.6% of the asymmetry tangent to its asymmetry block (92.6%); $\boldsymbol{W}_{\text{phys-diff}}$ gives 86.0% (64.0%) and 93.8% (87.3%). An objective that maximizes average alignment places no lower bound on alignment at individual parameter values.

Mesh convergence. We recompute the 11×9 (Re, κ) reference family on a refined 240×240 Taylor–Hood (P_2–P_1) mesh and compare it with the 160×160 references. The two solves agree to 5.3×10^{-6} in relative ℓ_2 velocity over all 99 parameter points, and every main result share and attribution is unchanged to three decimals.

Decoded rank and conditioning. The decoded Gram matrix is positive definite on all eight runs. Across the seven retained runs, its smallest eigenvalue is 0.0047–0.0069 times its largest, and its spectral condition number $\mathrm{cond}(\boldsymbol{G})$ is 145–215; $\mathrm{cond}(\boldsymbol{G}^{1/2})$ ranges from 12.1 to 14.6. All retained decoded spans are five-dimensional. Coefficient conditioning and decoded-field conditioning differ: the screened run s7 has the best-conditioned decoded Gram matrix, with $\mathrm{cond}(\boldsymbol{G}) = 69$.

Specification. $\boldsymbol{W}_{\text{phys-diff}}$ maximizes, over $\boldsymbol{W} \in O(r)$, the mean $A_j = \|E_j \boldsymbol{W} \gamma_j\|^2 / \|\gamma_j\|^2$ over physical axes and fitting points, with $\gamma_j = \boldsymbol{Q}^\top \boldsymbol{M} t_j$ the reference tangent coordinates for the field-orthonormal decoder $\boldsymbol{Q} = \boldsymbol{\Phi} \boldsymbol{G}^{-1/2}$. The coordinate projectors E_j select the labelled blocks ($2Re + 2\kappa$, or the two Allen–Cahn parameters). We exclude interaction modes, weight the two parameters equally and estimate tangents by central differences. The objective is invariant under within-block $O(2) \times O(2)$ rotations, so it can identify at most the identified subspaces. Held-out splits use four Powell restarts and 24 QR-generated random rotations per split.

Response source. $\boldsymbol{W}^{\text{op}}_{\text{coef-diff}}$ reaches 92% alignment with the model's own derivatives, and $\boldsymbol{W}_{\text{phys-diff}}$ captures 80% of the model tangents on the retained runs of the three-run diagnostic (s1, s42), with worst principal angles of 24–27° to $\boldsymbol{W}^{\text{op}}_{\text{coef-diff}}$. On the screened run s7 it captures 70%.

Robustness to the rank allocation. We re-optimize $\boldsymbol{W}_{\text{phys-diff}}$ under allocations $2Re/2\kappa/1\text{int}$, $1Re/3\kappa/1\text{int}$ and $2Re/3\kappa/0\text{int}$. Variance-based purity is allocation-independent, so $\boldsymbol{W}_{\text{agg}}$ is unchanged. On the two retained runs, derivative alignment is 81–90% and purity stays 11–26 points below its optimum under every allocation. Appendix K.2 measures how far the selected blocks move when the parameter measure or the response source changes.

H.3 A SEPARABLE CONTROL: ALLEN–CAHN

We use parametric Allen–Cahn as a separable *reference-family control* for the selection objectives of Appendix H.2. This control removes fluid dynamics and parameter-dependent boundary lifting. The equation is $u_t = \epsilon^2 u_{xx} + u - u^3 + h$ on $[-1, 1]$ (time t; subscripts denote partial derivatives) with homogeneous Neumann boundaries, where $\epsilon \in [0.05, 0.15]$ sets the interface width and $h \in [-0.20, 0.20]$ tilts the double well. Both parameters enter through the residual. The initial condition is $u(x, 0) = \tanh\big(x/(\sqrt{2}\,\epsilon_0)\big)$ with $\epsilon_0 = 0.10$ for every parameter value. We integrate to $t = 2$ on

513 grid points with 4000 steps, treating diffusion implicitly and the reaction explicitly. A snapshot is the space–time field on 129 points and 24 equally spaced times, minus the initial profile, and the parameter grid is 9×9. A mean-centred rank-4 POD of these snapshots isolates the geometry of the two objectives from operator training error. The reference solutions have variance shares of 19% for ϵ, 80% for h and 1.1% for interaction. Coefficient purity is 96–100%, and all 200 varimax starts agree. Physical-factor attribution reaches 99% for both parameters. Varimax gives 97–98% derivative alignment. One decomposition attains both optima together (Table 23, Figure 8).

Purity alone still leaves orientation underdetermined. Other purity optima reach the same purity with low attribution. In this example, width and position affect the field almost independently, and each response fits within its assigned subspace.

The two charts differ in kind. The cavity chart is the span of a trained operator, and the Allen–Cahn chart is a POD of reference solutions. So we repeated the cavity measurement on a mean-centred rank-5 POD of the cavity reference solutions, which captures 99.1% of their energy (Table 23). The two optima stay apart on that chart as well, and the alignment optimum gives up about half as much purity as on the operator's span. The difference between the variance and derivative objectives is a property of the cavity response family, and it is larger on the trained operator's span.

Table 23: Variance purity and derivative alignment, in percent, at varimax and at the two optima, for three charts. Each entry is purity / alignment.

chart	varimax	purity optimum	alignment optimum
cavity, span of operator s1	89 / 84	97 / 52	72 / 90
cavity, POD of reference solutions	93 / 79	98 / 26	85 / 84
Allen–Cahn, POD of reference snapshots	99 / 98	99 / 35	99 / 98

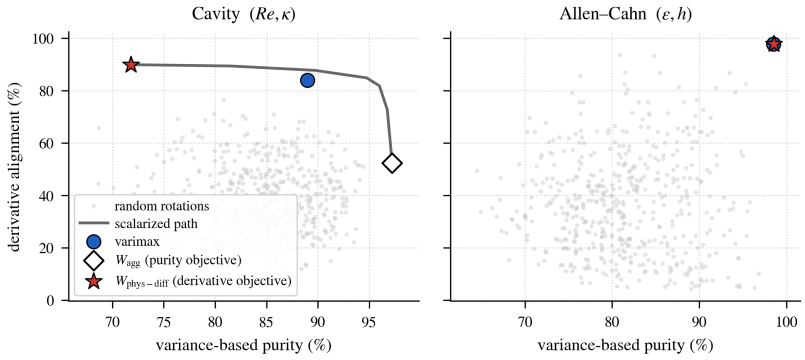


Figure 8: The selection objectives give different variance-based purity and field-L^2 derivative alignment. The plot includes random rotations and the path obtained by optimizing weighted sums of the objectives. In the cavity, variance-based selection and physical-response selection select different factor subspaces, and field varimax lies near both. In separable Allen–Cahn, the objectives are compatible. The physical-response selection is cross-validated out of sample (Appendix H.2).

H.4 The factor structure persists across rank

We examine rank dependence with two analyses that require no training. First, truncating the rank-5 operator's field-orthonormal coordinates to ranks 3–5 gives main-effect shares of field energy of 5% for Reynolds and 93–94% for asymmetry on s1, s42 and s7, matching the reference solutions.

Second, a nested POD sweep of the FEM family extends to rank 20. At ranks 2, 5, 8 and 20, the Reynolds shares are 2, 5, 5 and 5%, and the asymmetry shares are 97, 93, 92 and 92%. Varimax allocations are 0/2/0, 1/3/1, 2/3/3 and 2/4/14, respectively. Above rank 8, the shares remain close to the reference values of 5% Reynolds, 92% asymmetry and 3% interaction (Appendix H.1). Modes 6–20 contribute 0.9% of total variance. Modes above rank 8 contribute 0.15%.

In the POD sweep, matched subspace directions also converge. The rank-5 POD Reynolds subspace is within 3.5° of its rank-20 counterpart, and the rank-16 subspace is within 0.4°. The asymmetry

comparison is within approximately $4°$. Because dimensions can change with rank, these angles describe alignment of matched directions. Additional modes refine responses to the same two physical parameters.

Both analyses reuse the trained rank-5 operators or the reference solutions and involve no retraining. A rank sweep that retrains Burgers operators is in the repository (Table 5). The effective reference-component ranks, 2 for Reynolds and 3 for asymmetry, depend on the field metric and energy threshold. They are not defined by the imposed operator rank.

H.5 Basis dependence of disentanglement scores

We evaluate adapted DCI (Eastwood & Williams, 2018), MIG (Chen et al., 2018) and SAP (Kumar et al., 2018) against the factors $(1/Re, \kappa)$ on the 11×9 grid. DCI uses random-forest feature importances on the full coefficient surface and cross-validated informativeness. The reported scalar is its disentanglement component. MIG uses continuous mutual-information regression normalized by binned factor entropy. SAP uses in-sample gaps in the coefficient of determination R^2. These diagnostic conventions differ from parts of the original evaluation protocols.

For seed s1, field varimax gives DCI, MIG and SAP scores of 0.96, 0.44 and 0.82. The field-orthonormal coordinates before rotation give 0.79, 0.40 and 0.16, and a random rotation of them gives 0.79, 0.27 and 0.26. All three bases decode the same field, and SAP moves by more than 0.5 between them. These values characterize the chosen coordinates. They mix two effects. One is the basis, which the three rows above isolate. The other is the form of the scores. MIG and SAP reward one axis per factor, and here each parameter needs a block of two axes. An aligned basis then splits a parameter over two coordinates and loses part of the score.

Figure 9 illustrates the decoded patterns and field-coordinate surfaces after field varimax. Reference responses provide evidence for their interpretation (§3.2), and the downstream procedure requires its own equivariance check (§2.6).

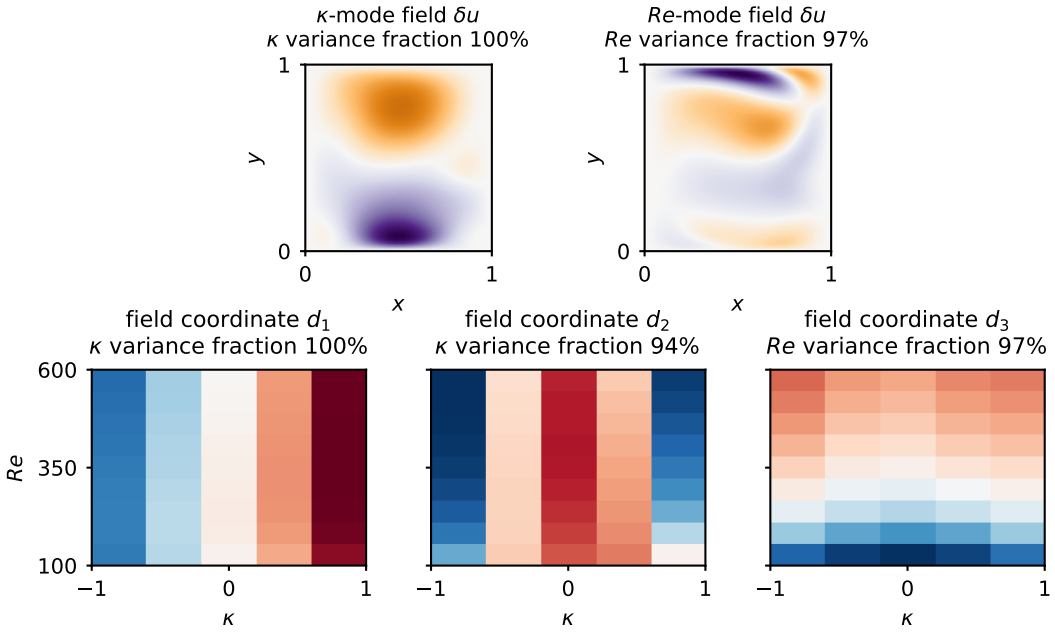


Figure 9: Field varimax concentrates main-effect variance while preserving the decoded field. Decoded mode perturbations appear alongside their field-coordinate surfaces. The rotation changes the latent vector $\boldsymbol{z}$ by less than 10^{-16} in maximum absolute entry. Here δu denotes a decoded mode perturbation and d_i the ith field-orthonormal coordinate. Percentages are ANOVA shares of field energy. Each panel uses its own symmetric colour scale.

Dependence on the parameter measure. The reported variance shares use the uniform product measure on the parameter lattice, under which functional ANOVA is an orthogonal decomposition.

For correlated inputs, the diagnostic allocates the explained variance of a weighted linear regression on normalized $1/Re$ and κ. The two-factor allocation of Lindeman, Merenda and Gold (LMG) (Grömping, 2006), the Shapley value of R^2, averages each predictor's increase in R^2 over both entry orders. The reported concentration is the larger allocation divided by their sum, averaged over coefficients. This regression-based statistic agrees with the uniform-measure ANOVA purity within 0.0007 at zero correlation in the tested data. It measures concentration of linearly explained variance.

In the Gaussian-copula sweep, the maximum absolute change from zero correlation ranges from 0.447 to 0.466 for this concentration across the eight runs. The corresponding changes in average derivative alignment range from 0.031 to 0.061, and the per-run ratio of these changes is 7–15. Relative to their zero-correlation values, the changes in concentration, MIG and SAP average 47%, 77% and 99% across runs, compared with 7% for derivative alignment. This measure sweep uses weighted histogram estimates for MIG and importance resampling for SAP. Its metric values are compared within the sweep. Variance-based claims require the parameter measure, the basis and the labelling rule. A pointwise attribution formula is independent of parameter weights. Its average and an optimized selector can depend on them.

I A SECOND PDE FAMILY: VISCOUS BURGERS

We repeat the low-rank construction on one-dimensional viscous Burgers, using an exact analytic reference. As in the cavity, a lifting function and vanishing factor impose boundary conditions, a shared trunk supplies spatial modes, and a low-rank latent represents the parameter dependence. Here the ansatz also imposes the initial condition. The coefficient table is bilinear on the parameter grid, with QR-orthonormal latent modes. All eleven seeds $\{1, 7, 11, 23, 42, 99, 314, 1729, 2026, 5150, 31337\}$ are trained from physics residuals. We evaluate the latent spectrum, coefficient ANOVA, varimax coordinates and physical-response capture. The field-metric analysis of these operators ran after the field metric and field varimax had been fixed as the primary rule on the cavity results. The commit that fixed the rule is an ancestor of the one the Burgers analysis ran from, so the analysis is a confirmatory use of the rule on a second family.

I.1 TIME-DEPENDENT BURGERS: SPAN RECOVERY AND A VISCOSITY AXIS

We solve

$$u_t + u\,u_x = \nu\,u_{xx}, \qquad (x,t) \in [-1,1] \times [0,1], \qquad u(x,0) = -a\sin(\pi x), \qquad u(\pm 1, t) = 0, \tag{16}$$

over the two-parameter box $\nu \in [0.0175, 0.05]$, $a \in [0.5, 1.5]$: ν sets the width of the interior shock and a its amplitude and formation time. The field is written

$$u(x,t;\,\nu,a) \;=\; \underbrace{-a\sin(\pi x)}_{\text{initial condition}} \;+\; \underbrace{t\,(1-x^2)}_{\text{boundary factor}}\;\langle\, \boldsymbol{z}(\nu,a),\, T(x,t)\,\rangle, \tag{17}$$

so the ansatz satisfies the initial condition at $t = 0$ and the boundary conditions for all t. The cavity uses the same boundary-lifting construction (Appendix C.1). The trunk is a Fourier-feature MLP in (x,t) and the latent rank is 4. The reference is the exact solution given by the Cole–Hopf transformation (Hopf, 1950; Cole, 1951), written as a Fourier–Bessel series,

$$u = 4\pi\nu\,\frac{\sum_{n\ge 1} n\,(-1)^n I_n\!\left(\frac{a}{2\pi\nu}\right) e^{-n^2\pi^2\nu t}\sin(n\pi x)}{I_0\!\left(\frac{a}{2\pi\nu}\right) + 2\sum_{n\ge 1}(-1)^n I_n\!\left(\frac{a}{2\pi\nu}\right) e^{-n^2\pi^2\nu t}\cos(n\pi x)}, \tag{18}$$

where I_n is the modified Bessel function of the first kind of integer order n. The series is evaluated with scaled Bessel functions and verified against an independent pseudo-spectral solver (relative L^2 agreement 1.5×10^{-7} to 7×10^{-6} across the box and across t) and against 50-digit arithmetic at the worst corner (2×10^{-6}). The lower limit on ν keeps the series above float64 cancellation. Table 24 summarizes the seeds. The construction is the cavity's: field-orthonormal coordinates and varimax from the identity. The labelling threshold is specific to the family. A Burgers mode receives a parameter label when that parameter's main effect accounts for more than 60% of its field energy, a value fixed with the first Burgers study, before any field-metric analysis, and it is labelled mixed otherwise. The cavity uses 50%. With the cavity threshold the viscosity axis, its share of the exact

viscosity response and its percentile are unchanged on all eleven operators. One mixed axis becomes an amplitude axis on ten of them, and the margin over the amplitude block is 79–84% in place of 79–89%.

Table 24: The time-dependent Burgers audit. Field varimax gives every operator one viscosity-labelled axis. Ranges include every basis in each row. S is the fraction of the exact response $\partial u/\partial \nu$ captured by the rank-4 span. The shares partition the captured response among the labelled field-orthonormal axes, the remainder lying in mixed modes.

bases	count	u_{rel} on-node	off-node	field-varimax labels	span capture S	ν-share	a-share
all operators	11	0.0283–0.0435	0.0164–0.0298	one ν, one a, two mixed	99.3–99.6%	80.9–91.2%	1.8–1.9%

In field-orthonormal coordinates, the ANOVA shares of the learned coefficient surfaces are 1.5–2.7% for viscosity, 96.4–97.8% for amplitude and 0.8–0.9% for interaction on all eleven operators. In the stored coefficients the interaction share is 21.6–28.5%. Field error is concentrated near the parameter-box boundary. Interior-node errors reach 0.014–0.021, and the lowest-viscosity edge has the largest errors. Off-node averages are lower because they include a different proportion of boundary cells.

Exact viscosity-response attribution. Viscosity does not enter the lifting function. We audit the operator's rank-4 span against its exact viscosity response $\partial u/\partial \nu$ (Cole–Hopf central difference). The rank-4 span captures 99.3–99.6% of the exact response on all eleven trained models. The field-varimax viscosity axis receives most of the captured response on every model and exceeds the amplitude axis by 79 to 89 points. Its share lies at the 96th to 99th percentile of a random-direction null (Table 24). Span-level recovery and recovery of a parameter-labelled axis therefore coincide on every tested Burgers operator.

Comparison with the cavity. Both families have low-rank latent representations, decreasing mode energies and varimax bases with parameter-labelled modes. Boundary or initial forcing accounts for most field variance: approximately 92% for cavity asymmetry and 95% for Burgers amplitude, compared with 5% for Reynolds and 4% for viscosity. Interaction contributes 2.7% and 0.6% of field energy, respectively. Its larger share in the learned coefficient surfaces reflects the coefficient metric and rank allocation. These comparisons describe the tested parameter domains and measures.

Matched block comparisons. On the eleven rank-4 Burgers operators, field-orthonormal coordinates improve every block predictor of the matched comparison. The labelled blocks, label-shuffled blocks of the same dimensions and the shared-kernel span each reach about half of their error in the stored coefficients or less. Labelled blocks have lower error than the label-shuffled ones on 10 of 11 operators (one-sided sign test $p = 0.0059$). Under the cavity's block-shared Gaussian predictor, the difference between field varimax and operator-response selection shrinks to about a tenth of its registered size (Table 30), as on the cavity.

J Stability of identified subspaces and varimax coordinates

J.1 Identified blocks across training runs

We compare the identified subspaces of the model-only basis across all 21 pairs of retained runs. Each block is orthonormalized separately, preserving its decoded column span. Given decoded matrices A_i, A_j and block transforms T_i, T_j, the singular values of

$$(T_i^\top A_i^\top A_i T_i)^{-1/2} T_i^\top A_i^\top A_j T_j (T_j^\top A_j^\top A_j T_j)^{-1/2}$$

are the cosines of the principal angles. The full spans of two runs agree to within a few degrees (Table 25). So does every field-varimax block, the asymmetry block most closely (Figure 3(b), Table 18). Operator-response selection also agrees across runs to within a few degrees (Table 29).

J.2 Varimax coordinates in the field metric

Antisymmetry of the two κ modes. Appendix H.5 (Figure 9) interprets the field-varimax κ modes using the symmetric fraction $s(f) = \|(f + f \circ \mathcal{R}_y)/2\|/\|f\|$. Here f is the horizontal velocity δu of

a decoded mode on the uniform grid. The map $\mathcal{R}_y : y \mapsto 1-y$ reflects the cavity about its horizontal midline, and $\circ$ denotes composition ($s = 0$ purely antisymmetric, $s = 1$ purely symmetric). The reflection maps the velocity (u, v) to $(u, -v)$ and the streamfunction to its negative, so it acts on u through the grid alone. The symmetric and antisymmetric parts of f are orthogonal on this grid, so s^2 is the share of the mode's energy in its symmetric part. Every run contains one nearly antisymmetric mode and one mostly symmetric mode, consistent with two-sided lid forcing, and the antisymmetric mode carries the larger coordinate amplitude on all eight runs. Across the seven retained runs the antisymmetric mode has $s = 0.229$–0.233, so about 95% of its energy is antisymmetric. The symmetric mode has $s = 0.808$–0.812, about two thirds of its energy.

A per-mode label depends on the orientation within its two-dimensional block, so we compare varimax with the full family of internal rotations, whose members span the same block. The decode is linear in the latent perturbation, so a rotated mode $\cos\varphi\, \boldsymbol{v}_a + \sin\varphi\, \boldsymbol{v}_b$ decodes to $\cos\varphi\, g_a + \sin\varphi\, g_b$, where φ is the rotation angle and g_a, g_b are the decoded mode fields. The scale-invariant ratio s removes mode amplitudes, so $s(\varphi)^2$ is a ratio of two quadratic forms built from the 2×2 Gram matrices of $\{g_a, g_b\}$ and their symmetric parts. This formula agrees with direct decoding at $\varphi = 0$ and $\pi/2$ to 10^{-9} on every run. Over that family s ranges from 0.13 to 0.85, and varimax returns the same orientation on every retained run: its antisymmetric-mode values span less than 1% of that range. The coordinate-level reading therefore depends on the identification rule and reproduces across runs of one architecture.

Optimization, sampling and training variation. All 300 random orthogonal starts of field varimax converge to the same solution up to signed permutation on s1. Bootstrapping the Reynolds and asymmetry grid levels with replacement 300 times gives asymmetry-mode purities of 0.99 $[0.95, 1.00]$ and 0.95 $[0.81, 0.99]$ and a median largest principal angle of 6.1° for the two-dimensional asymmetry block. Table 25 compares the identified subspaces of the reference-initialized basis across training runs, and Table 18 and Figure 3(b) compare the model-only basis. Inside one run the two bases differ in the asymmetry block by far more than either basis varies across runs (Table 29). Retained runs agree to a few degrees in the full span, the asymmetry block and the leading Reynolds axis. The screened run s7 differs by tens of degrees in its span.

Table 25: Reference-initialized field-varimax subspaces agree across training runs. Values are principal angles in degrees. The full-span column reports the largest angle and the other columns the mean angle. e_5 as in Table 7.

pair	e_5	full span (largest)	κ block (dimension 2)	leading Re axis (dimension 1)
s1 versus s5150	0.051	3.0	0.8	0.6
all 21 retained pairs		1.1–3.7	0.3–0.9	0.4–1.3
s1 versus s7 (screened)	0.009	66.3	7.2	12.8

J.3 Operators trained after the protocol was fixed

We trained the four confirmatory operators with the training command of the eight development operators, on seeds no earlier cavity run had used. We committed the analysis script, its control on the seven retained runs and the criteria before submitting the runs (`design/PREREG_CONFIRMATORY_SEEDS_20260921.md`). The script ran once, after all four runs had finished. A confirmatory run's resolved configuration differs from that of s1 in the seed, the output paths and an unused precision block. Every run keeps its last epoch, as the development operators do. Table 26 reports the four runs. All four pass screening and meet every preregistered criterion, so the identification result is replicated on operators that did not exist when the protocol was chosen.

Table 26: The four confirmatory operators replicate the identification result. e_5 as in Table 7. Recall is component recall of the field-varimax block with the component's label. Double counting is that of varimax on the stored coefficients, in percentage points, the contrast of Figure 3. In field-orthonormal coordinates it is zero on every run. Angles are the largest mean principal angle of a block against the seven retained runs and against the other confirmatory runs. The restart column is the largest such angle over the uniform restarts of varimax, all of which return the same allocation. The last row gives the preregistered bar of each column. Table 29 gives the same quantities for the seven retained runs.

			recall (%)			double counting	largest block angle (deg)		
seed	e_5	allocation	Re	κ	int	(stored, points)	retained	confirmatory	restarts
s11	0.070	$2/2/1$	73	97	41	126	2.3	2.4	0.03
s23	0.047	$2/2/1$	73	97	41	140	1.8	2.1	0.01
s99	0.058	$2/2/1$	74	97	41	115	2.3	2.4	0.02
s65537	0.055	$2/2/1$	74	97	41	105	2.2	2.0	0.01
preregistered bar	≥ 0.02	$2/2/1$	≥ 95				≤ 5		≤ 5

K Reproducibility and sensitivity of subspace identification

K.1 Restart stability under a fixed rule

We test numerical uniqueness within each checkpoint separately from cross-seed stability. The protocol and threshold were fixed before evaluation. The registered study maximizes the operator-response objective from starts near the varimax basis, which are exponentials of Gaussian skew matrices of scale 0.3. It keeps the solutions within 1% of the best objective and compares their labelled projectors. Every start qualifies on every run, and the 95th-percentile angle stays below the pre-specified 5° threshold (Table 27).

Those starts perturb one basis. In a later study we drew the starts uniformly from the rotation group $SO(5)$, for both rules and all eight checkpoints. Field varimax returned the same allocation and the same blocks from every start. For operator-response selection all but two solutions qualified, and the 95th-percentile angle fell far below the threshold. A few qualifying solutions lie far from the best one, the farthest on the screened run s7. Both searches reproduce the projector tuple for a fixed rule. They do not prove global uniqueness, and a projector tuple fixes no basis inside its blocks.

Table 27: Restarts of the two identification rules. A start qualifies when its objective is within 1% of the best. For field varimax the row counts the starts that return the identified allocation. Angles are between the labelled projectors of a solution and those of the best solution.

	registered starts	starts uniform on $SO(5)$	
	operator response	field varimax	operator response
starts per checkpoint	200	300	200
median rotation of a start (deg)	49	154–157	
qualifying starts	all	2400 of 2400	1598 of 1600
largest 95th-percentile angle (deg)	3.7	–	0.06
largest angle (deg)	–	0.03	23 retained, 67 on s7

K.2 Robustness to measure, allocation and response source

At fixed block dimensions, the registered statistic is the largest principal angle over the labelled Reynolds and asymmetry blocks of operator-response selection. We perturb the objective in two ways on the seven retained runs. One tilts the parameter measure $3\times$ toward high Re and $|\kappa|$. The other replaces the model responses with reference responses (Table 22). Both move the tuple beyond the 15° threshold, so the selected tuple depends on these choices (Table 28). Split by block, the tilted measure moves the Reynolds and interaction blocks and leaves the asymmetry block nearly in place. Reference responses move all three blocks by similar amounts. Field varimax uses no responses, so the second perturbation does not apply to it. Moving one κ axis to the Reynolds block ($2/2/1$ to $3/1/1$) keeps the original Reynolds plane inside the enlarged block.

Table 28: Sensitivity of operator-response selection to its own choices. Entries are medians over the seven retained runs of the largest principal angle, in degrees, between the perturbed and the unperturbed solution. The registered statistic takes the larger of the Reynolds and asymmetry angles. For the reallocation it measures how far the original Reynolds plane leaves the enlarged block.

perturbation	registered statistic	Reynolds	asymmetry	interaction
parameter measure tilted 3×	18.7	18.7	4.8	18.3
reference responses	26.4	26.4	24.4	26.1
one κ axis moved to the Reynolds block	1.2	–	–	–

L THE REGISTERED COEFFICIENT-METRIC PROTOCOL

The rule we registered in the repository before running identified subspaces by operator-response selection in the stored coefficients. For parameter j on the interior parameter grid $\mathcal{Q}$, the model response is the lift-subtracted field derivative expressed in coordinates, $g_j(q) = \boldsymbol{\Phi}^\dagger \partial_{p_j}(u_\theta - u_{\mathrm{fixed}})(q)$ in the stored coefficients and $g_j(q) = (\boldsymbol{\Phi}\boldsymbol{G}^{-1/2})^\top \partial_{p_j}(u_\theta - u_{\mathrm{fixed}})(q)$ in field-orthonormal coordinates, where u_θ is the model field, u_{fixed} its lift-plus-mean contribution and $\dagger$ the Moore–Penrose inverse. With E_j the diagonal projector onto the axes assigned to parameter j, the rule rotates the varimax axes by the maximizer of

$$J_{\mathrm{op}}(\boldsymbol{W}) = \frac{1}{2|\mathcal{Q}|} \sum_{q \in \mathcal{Q}} \sum_{j \in \{Re, \kappa\}} \frac{\|E_j \boldsymbol{W} g_j(q)\|_2^2}{\|g_j(q)\|_2^2 + \varepsilon}, \qquad \boldsymbol{W} \in O(5), \tag{19}$$

where $|\mathcal{Q}|$ counts grid points and $\varepsilon = 10^{-30}$ prevents division by zero. Internal rotations leave J_{op} unchanged, so the rule fixes subspaces and leaves their axes free (Proposition 2). We maximize J_{op} with Powell's method over $\boldsymbol{W} = \exp(\boldsymbol{S})$ with $\boldsymbol{S}$ skew-symmetric. We run ten fixed-seed starts near the varimax basis and keep the best solution. Appendix K tests other starts.

In the stored coefficients this objective scores model derivatives with coefficient norms, and physical validation measures decoded-field energy. A diagnostic of the registered results in both metrics led us to analyse every study in field-orthonormal coordinates. We registered that change before any field-metric production run and made field varimax the primary rule after reading the field-metric results. Each field-metric study repeats its registered counterpart with the same checkpoints, grids, draws, candidate kernels and tolerances, and the package records both changes. Tables 29 and 30 place the registered results beside the field-metric ones.

Selection recalls less of the Reynolds component than varimax on six of the seven retained runs in the stored coefficients and on all seven in field-orthonormal coordinates (Table 29). The change of metric does not explain this decrease. It comes with the objective. J_{op} rewards the share of the model's own derivatives that falls in a block, and component recall measures the energy of the reference components. Under both rules, the field metric eliminates double counting and substantially reduces differences between training runs.

Table 29: Model-only identification in the two metrics, seven retained runs. Completion error is the mean relative error of Table 10. Recall is component recall of the block with the component's label. Double counting is the largest excess of a component's recall summed over blocks over its capture by the span. Angles are the largest mean principal angle of a block between two runs, and the largest angle between model-only and reference-initialized blocks of one run.

	stored coefficients (registered)		field-orthonormal	
	varimax	selection	varimax	selection
completion error	0.1887	0.1565	0.1381	0.1378
Reynolds recall (%)	57–84	59–72	73–74	65–66
asymmetry recall (%)	97	96	97	94
interaction recall (%)	18–25	23–43	41–42	27–30
double counting (points)	143	59	0	0
between runs, Reynolds / interaction (deg)	23 / 12	18 / 37	2.2 / 2.2	2.6 / 3.7
model-only against reference-initialized (deg)	54	0.7	19	0.8

In the stored coefficients, selection corrected blocks that varimax left overlapping and improved cavity completion under every tested learner. On Burgers the registered difference is negative, so its

Table 30: Reference-initialized comparisons of varimax and operator-response selection in the two metrics. Entries are mean varimax-minus-selection completion error, so positive values favour selection, with the number of runs on which the difference is positive; n counts observed off-cross fields. The percentile row places varimax among 500 random subspaces of the same dimensions. The branched-map and second-architecture rows summarize studies whose accounts are in the repository (Table 5).

comparison	stored coefficients (registered)	field-orthonormal
Gaussian kernel, $n = 0$ (7 runs)	$+0.0525$, $7/7$	$+0.0059$, $7/7$
Gaussian kernel, $n = 1$	$+0.036$, $7/7$	-0.000, $1/7$
Gaussian kernel, $n = 2$	$+0.025$, $7/7$	$+0.000$, $3/7$
Gaussian kernel, $n = 4$	$+0.014$, $7/7$	-0.001, $0/7$
varimax percentile among random subspaces	72–83	1
kernel chosen by target error	$+0.0060$, $7/7$	-0.0050, $0/7$
cubic polynomial ridge	$+0.0049$, $7/7$	$+0.0000$, $3/7$
branched coefficient map	$+0.041$, $7/7$	$+0.003$, $4/7$
second architecture (5 runs)	$+0.049$, $5/5$	$+0.005$, $5/5$
Burgers, rank 4 (11 runs)	-0.022, $1/11$	-0.002, $1/11$

sign depends on the PDE family. In field-orthonormal coordinates varimax already returns orthogonal blocks that agree across runs. The difference between the two rules shrinks to about a tenth of its registered size or less, and the kernel chosen by target error favours varimax. Both rules stay far ahead of random subspaces of the same dimensions. The prediction invariance under internal rotations holds in both metrics.