

Too Symmetric to See: Diagnosing Over-Invariance in Finite-Symmetry Geometric Learning

D Yang Eng*

Technical University of Munich (TUM), Arcisstraße 21, 80333 Munich, Germany
Singapore Institute of Technology, 1 Punggol Coast Road, Singapore 828608, Singapore

Abstract

Invariant representations can discard target-relevant information while passing every symmetry test. We diagnose this over-invariance through conditional variance on the fibres collapsed by a representation. A synthetic projective benchmark gives the exact magnitude-only error floor, and a local H_3^+ control separates a persistent radial deficit from optimization error. Projective character pooling (PCP) retains selected phase information without orbit enumeration. On the Dwork quintic, matched-capacity, charge and intervention controls connect this information to lower determinant-ratio flatness error. On a bicubic threefold, intersection-volume checks validate the residue sampler before representation comparison. Exact orbit averaging matches or surpasses PCP at longer tested budgets. A canonicalization control exposes a separate failure: exact action invariance and low local loss coexist with discontinuous boundary gluing. The study provides a workflow for distinguishing information loss, optimization error and geometric validity in finite-symmetry learning.

1 Introduction

Symmetry constrains scientific learning, but an invariant representation can discard distinctions that its target requires. Permutation pooling, frame averaging and canonicalization encode known group actions [1–3]. Action-error tests verify that predictions are unchanged along those actions. They do not reveal information lost when the representation identifies additional, inequivalent inputs.

This problem is known as excessive invariance or over-invariance [4, 5]. We study the case in which a representation imposes a strict supergroup of the target symmetry. For complex projective coordinates, squared magnitudes remove a continuous phase torus, whereas a target with finite root-phase symmetry can retain relative-phase dependence. Learned invariances, approximate equivariance and invariant learning under distribution shift offer complementary ways to calibrate symmetry [5–8]. Our question is how to measure the information removed by a specified representation. It connects orbit separation in invariant theory [9] with conditional sufficiency in probabilistic descriptions of invariant learning [10].

We use conditional variance to quantify the irreducible squared error of predictors restricted to that representation. A synthetic projective benchmark gives this error in closed form. A local H_3^+ control then distinguishes a persistent radial error plateau from the training-dependent error of exact orbit averaging. Projective character pooling (PCP) supplies a comparison representation: it retains selected phase information through powers of block-normalized coordinates and charge-cancelling

*Author to whom correspondence should be addressed: dyang.eng@tum.de. ORCID: 0000-0002-0726-1698.

contractions. The construction supports diagonal root-phase actions and simultaneous coordinate permutations without enumerating their orbits.

Calabi–Yau metric learning tests this diagnosis under geometric constraints. Calabi’s conjecture and Yau’s theorem establish existence and uniqueness in a fixed Kähler class [11, 12]; explicit approximation still requires numerical methods. Balanced and energy-functional approaches provide classical references [13–16]. Learning-based approaches use direct metric parameterizations, global potentials, finite symmetries and section-space optimization [17–23]. An earlier fixed- ψ quintic study identified the importance of projective section structure [24]; the present work asks which phase distinctions a projective representation must preserve. Analytic approximation and symbolic distillation address the complementary task of constructing compact geometric surrogates [25–29].

The Dwork quintic and a symmetric bicubic provide two tests with different ambient spaces and finite groups. Matched-capacity, charge and checkpoint controls connect phase access to the determinant-ratio diagnostic. The bicubic sampler is checked against exact intersection volumes before training comparisons. A canonicalization boundary audit exposes a second failure: exact invariance and low local loss can coexist with discontinuous global gluing. Together, the experiments distinguish representation loss, optimization error and geometric validity within one evaluation workflow.

2 Controlled diagnosis of over-invariance

When a representation R identifies inputs with different target values T , every predictor restricted to R incurs the squared-error floor

$$\inf_f \mathbb{E}[(T - f \circ R)^2] = \mathbb{E}[\text{Var}(T \mid R)]. \quad (1)$$

This identity measures target variation on collapsed representation fibers. We first compute the floor analytically, then distinguish information loss from optimization error on electronic-structure labels.

2.1 A synthetic phase-content benchmark has an analytic error floor

Fix a single coordinate block of size $n \geq 2$ and an integer degree $d \geq 1$. The declared action is $z_i \mapsto \zeta^{a_i} z_i$, with $\zeta^d = 1$, together with coordinate permutations. For any nonzero representative define $\hat{z} = z/\|z\|$ and $u_i = |\hat{z}_i|^2$, so $\sum_i u_i = 1$. Sampling $\hat{z}$ uniformly on the unit sphere in $\mathbb{C}^n$ induces Fubini–Study probability measure on $\mathbb{CP}^{n-1}$. Throughout this benchmark the coefficients of all degree- d terms are exactly one:

$$f_\star(z) = (1 - \gamma)A(u) + \gamma C(\hat{z}), \quad A(u) = p_2(u) + \tfrac{1}{2}p_3(u), \quad C(\hat{z}) = \left| \sum_{i=1}^n \hat{z}_i^d \right|^2, \quad (2)$$

where $p_j(u) = \sum_i u_i^j$. No factor $1/\sqrt{n}$ is used in C . Both terms are projectively invariant and respect the declared root-phase and permutation actions. The coefficient $\gamma \in [0, 1]$ sets the phase contribution, not a variance-explained fraction.

For every magnitude-only predictor, Eq. (1) reduces to

$$\mathcal{E}_{\text{moduli}}^2 = \gamma^2 \mathbb{E}_u [\text{Var}(C(\hat{z}) \mid u)] = \gamma^2 n(n-1) \frac{(n-1)!(d!)^2}{(n+2d-1)!}. \quad (3)$$

The result is an exact Bayes error under the stated distribution. Appendix B.3 derives the phase cancellation and Dirichlet moments explicitly, and gives the coefficient-dependent formula for general c_i .

Table 1: Depth- and parameter-matched synthetic benchmark at $n = d = 5$. Held-out MSE against $f_\star$ in Eq. (2), in units of 10^{-6} , reported as mean $\pm$ population standard deviation over three seeds. The analytic floor applies to every magnitude-only predictor.

Arm ($\times 10^{-6}$)	$\gamma = 0$	$\gamma = \frac{1}{4}$	$\gamma = \frac{1}{2}$	$\gamma = \frac{3}{4}$	$\gamma = 1$
Analytic moduli-only floor	0	4.96	19.8	44.6	79.3
Moduli-only Deep Sets	9.87 ± 0.29	12.58 ± 0.19	25.95 ± 0.40	49.91 ± 1.39	84.69 ± 1.96
Character pooling	8.38 ± 0.54	5.89 ± 0.26	4.90 ± 0.25	4.31 ± 0.57	3.96 ± 0.90

Table 2: Local H_3^+ normal-mode energy convergence. Held-out RHF/STO-3G energy MSE in Hartree squared, mean $\pm$ population standard deviation over three seeds. All arms have 4,417 parameters.

Optimizer updates	Radial ($O(2)$)	Character pooling	Exact D_3 orbit average
900	$(2.56 \pm 0.16) \times 10^{-8}$	$(5.10 \pm 0.30) \times 10^{-11}$	$(1.84 \pm 1.08) \times 10^{-8}$
3,000	$(2.56 \pm 0.16) \times 10^{-8}$	$(5.04 \pm 0.31) \times 10^{-11}$	$(9.70 \pm 4.40) \times 10^{-10}$
9,000	$(2.56 \pm 0.16) \times 10^{-8}$	$(4.73 \pm 0.36) \times 10^{-11}$	$(6.10 \pm 2.46) \times 10^{-11}$

2.2 Matched models recover the predicted separation

We compared parameter-matched character pooling and moduli-only Deep Sets at $n = d = 5$ across $\gamma \in \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$. Both arms used 20,000 training points, 10,000 held-out points, 8,000 optimizer updates and three seeds. The architectures differed only in access to a phase statistic with zero phase-torus mean at fixed moduli; the magnitude-only arm received a zero vector of the same shape.

Table 1 and Figure 1 separate the two regimes. At $\gamma = 0$, the matched arms had similar errors. With phase content, the magnitude-only error approached its analytic floor, reaching within 7% at $\gamma = 1$, while character pooling remained below 6×10^{-6} MSE.

2.3 A local D_3 molecular control separates information loss from optimization

The degenerate in-plane normal mode of equilateral H_3^+ carries a D_3 action. A radial representation is invariant under the strict supergroup $O(2)$, whereas the local energy surface may contain the D_3 -allowed angular character $\text{Re}(q_x + iq_y)^3$. We generated 2,000 RHF/STO-3G energies around the equilateral geometry with PySCF [30]. Four independently sampled D_3 orbits agreed to 4.0×10^{-15} Hartree. These labels provide a reproducible symmetry control rather than a chemical-accuracy benchmark.

The radial, character, and exact orbit-average arms shared the data split, minibatches, schedule and 4,417 parameters within each seed. They received, respectively, $|q_x + iq_y|^2$, $(|q_x + iq_y|^2, \text{Re}(q_x + iq_y)^3)$, and all six D_3 images of a shared MLP. Table 2 separates two effects. The radial error remained at $2.56 \times 10^{-8} \text{ Ha}^2$ from 900 through 9,000 updates. By contrast, exact orbit averaging improved from 1.84×10^{-8} to $6.10 \times 10^{-11} \text{ Ha}^2$ and overlapped the character model at the longest budget. The radial plateau is therefore consistent with information excluded by the $O(2)$ representation, while the initial orbit-average deficit was optimization-limited under the tested schedule.

The synthetic benchmark quantifies discarded information exactly; the molecular control distinguishes a persistent radial plateau from an optimization-limited orbit average. We next apply this diagnosis to projective geometry.

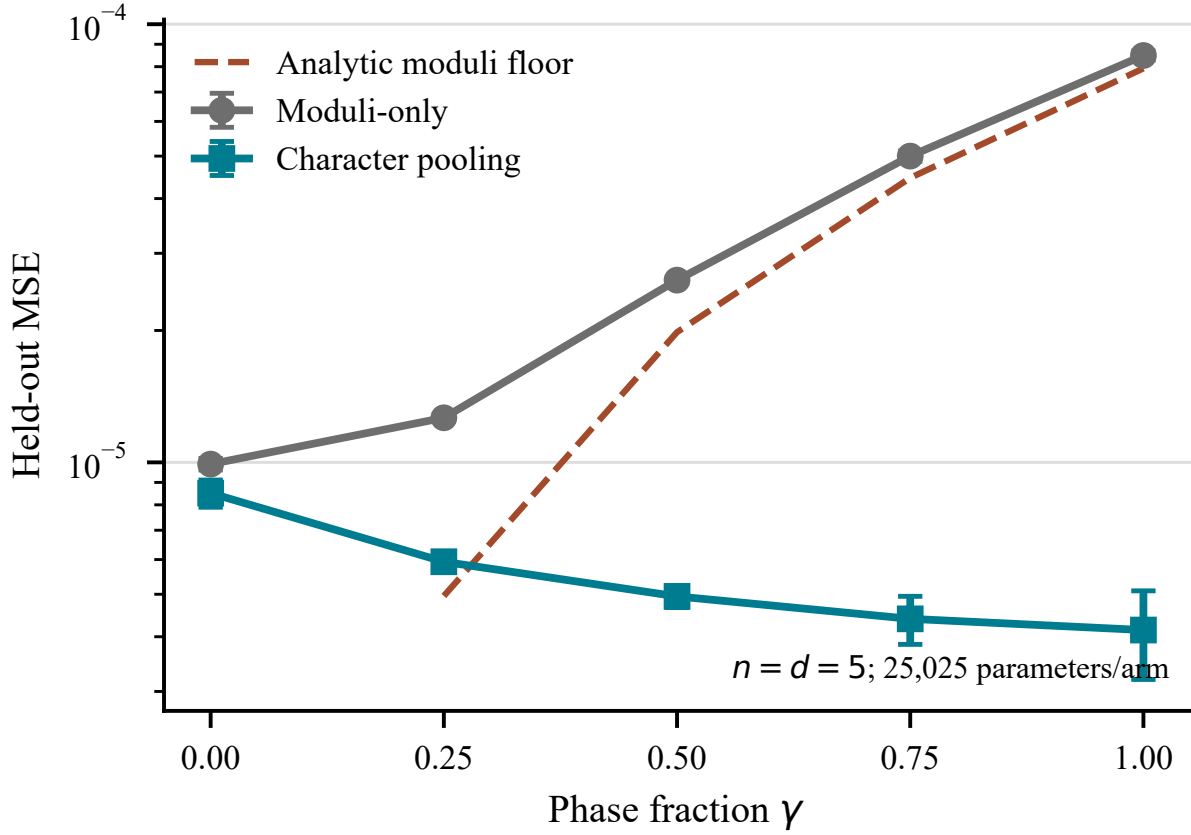


Figure 1: Controlled phase-content experiment at matched parameter count. The dashed curve is the analytic Bayes MSE of every magnitude-only predictor. Points show means and population standard deviations over three seeds. At nonzero γ , the magnitude-only model approaches its predicted floor, whereas character pooling remains below 6×10^{-6} MSE.

3 Geometric applications

The Dwork quintic and symmetric bicubic test the diagnosis across different ambient spaces, finite groups and samplers. Within each comparison, the arms share their base potential, sampler, objective, optimizer, schedule, seeds and evaluation rule; only the residual representation changes. Methods defines the projective metric and flatness diagnostic.

3.1 Dwork quintic

3.1.1 Geometry and symmetry

The one-parameter Dwork family is

$$X_\psi = \left\{ [z_0 : \cdots : z_4] \in \mathbb{CP}^4 : \sum_i z_i^5 - 5\psi \prod_i z_i = 0 \right\}. \quad (4)$$

At $\psi = 0$ the root-phase and permutation group is $(\mathbb{Z}_5)^4 \rtimes S_5$. For generic $\psi \neq 0$, the product term requires fifth-root charges to sum to zero modulo five, reducing the group to $(\mathbb{Z}_5)^3 \rtimes S_5$ of order 15,000. Our nonzero- ψ scan keeps this symmetry class fixed while varying the geometry. Points

Table 3: Dwork modulus scan. Late-training validation $\sigma(\eta)$, mean $\pm$ population standard deviation over three seeds; each seed contributes the mean of its final five evaluations. Percentages are reductions relative to the moduli-only baseline. All three symmetry-structured arms had zero observed positivity violations on the 128-point diagnostic subset.

ψ	Moduli-only Deep Sets	Root-phase-aware	Dwork-adapted $m = 1$
0.3	0.0479 ± 0.0009	0.0277 ± 0.0009 (42%)	0.0278 ± 0.0016 (42%)
0.5	0.0696 ± 0.0038	0.0316 ± 0.0025 (55%)	0.0309 ± 0.0015 (56%)
0.8	0.1031 ± 0.0096	0.0408 ± 0.0013 (60%)	0.0393 ± 0.0014 (62%)

Table 4: Controls at $\psi = 0.5$. The projector MLP is globally projective but has no declared permutation or character structure. The chart-fixed MLP is evaluated in the common canonical chart and is a capacity and structure control, not a globally projective model. Positivity violations are averaged over the final five diagnostic evaluations.

Residual	Parameters	$\sigma(\eta)$	Positivity viol. (%)
Moduli-only Deep Sets	20,929	0.0696 ± 0.0038	0.00
Projector MLP control	24,949	0.0683 ± 0.0007	0.00
Root-phase-aware	29,121	0.0316 ± 0.0025	0.00
Dwork-adapted $m = 1$	25,217	0.0309 ± 0.0015	0.00
Dwork-adapted $m = 2$ (neg. control)	25,217	0.0655 ± 0.0014	0.00
Dwork-adapted $m = (1, 2)$	29,505	0.0306 ± 0.0013	0.00
Chart-fixed MLP	25,741	0.1249 ± 0.0124	21.88

are drawn on $\mathbb{CP}^4$, projected onto X_ψ by Newton–Raphson iteration, retained subject to a residual tolerance, and weighted by Shiffman–Zelditch-style factors. The scan uses 20,000 training and 2,000 validation points, batch size 1024, 200 epochs, and a cosine schedule from 10^{-3} to 10^{-5} .

3.1.2 Phase-aware residuals improve the Dwork flatness diagnostic

Both phase-aware families reduce the held-out flatness diagnostic at every tested modulus, by 42%–62% relative to moduli-only Deep Sets (Table 3). The root-phase and Dwork-adapted variants have similar scores. As specified in Methods, all reductions summarize three-seed means.

3.1.3 Controls

Table 4 collects the pre-declared controls at $\psi = 0.5$, each targeting one alternative explanation.

Neither extra capacity nor longer training closes the gap. Widening moduli-only Deep Sets to 24,991 parameters, close to the Dwork arm’s 25,217, gives 0.0653 ± 0.0017 ; training the original baseline for 800 epochs gives 0.0634 ± 0.0009 .

Generic phase access is also insufficient in this protocol. The projector control uses $\Pi(z) = zz^\dagger/z^\dagger z$, which retains the full projective point but imposes no permutation or character structure, and remains at baseline level. The $m = 2$ arm likewise retains the structured pipeline and parameter count but imposes an additional phase invariance through its even-power channels. Its baseline-level score is consistent with the over-invariance diagnosis. Together, the controls identify the chosen character structure, rather than parameter count or projective well-definedness alone, as the effective change.

At $\psi = 0.9$ and 0.95 , the Dwork-adapted arm reaches 0.0443 ± 0.0002 and 0.0602 ± 0.0099 , against 0.1143 ± 0.0076 and 0.1183 ± 0.0057 for Deep Sets, with positivity-violation rates of 0.09% and 0.15%. A joint data-and-compute sweep at $\psi = 0.5$ retains the advantage from 5,000 points (0.0583 ± 0.0008 against 0.0779 ± 0.0019) to 80,000 points (0.0254 ± 0.0017 against 0.0655 ± 0.0011). This fixed-epoch sweep increases both data and optimizer updates.

3.1.4 Exact action invariance does not guarantee smooth gluing

Canonicalization exposes a distinction between pointwise invariance and global geometric validity. The trained residual passes held-out Dwork-action and projective-rescaling audits to floating-point precision away from fundamental-domain boundaries and achieves a low local flatness score. A global Kähler-potential approximation also requires smooth gluing across those boundaries.

We therefore constructed tangent curves crossing every one of the 25 magnitude-order and phase-sector boundary classes and evaluated paired points approaching each boundary at four scales from 10^{-2} to 3×10^{-4} . For two independently trained checkpoints, all boundary classes were covered. At the smallest scale, total-potential differences remained between 0.047 and 0.090, residual-gradient differences remained between 1.82 and 1.87, and fixed-chart induced-metric discrepancies remained between 13% and 17%. These quantities did not decay with the approach scale, which is incompatible with smooth gluing. The sampled induced metrics remained positive on both sides of every tested pair, but positivity alone does not repair non-vanishing jumps in the potential, its gradient, or its metric.

The canonical model therefore fails as a global Kähler-potential approximation despite its low local score. Action invariance and local flatness alone miss this failure; smoothness must be established separately, as in the construction proof in Appendix D.3.

3.1.5 Interventions and a representational floor

Zeroing phase-feature inputs and recomputing $\sigma(\eta)$ degrades the successful arms by factors of 3.84–4.46, while leaving the ineffective $m = 2$ arm unchanged (0.999 ± 0.007). The trained models thus use the restored phase pathway.

An independent check evaluates degree- $k = 10$ H-matrix determinant observables along numerical fixed-moduli phase continuations on X_ψ . With 256 retained fibres of 16 points each, mean within-fibre variances are 1.78×10^{-3} , 2.28×10^{-3} and 2.38×10^{-3} across the three moduli; their 95% percentile-bootstrap intervals exclude zero. Under the equal-fibre empirical measure, each value is the exact minimum squared error for predictors restricted to the assigned fibre labels. Calling it a moduli-only floor additionally requires exact common fingerprints: floating-point tolerance alone cannot restrict arbitrary measurable predictors. Appendix B proves the bound and gives the continuum rank-and-derivative criterion, which remains uncertified for $T_{\det}$.

The scalar-prediction bound is distinct from a flatness bound for moduli-only potentials, whose induced determinant observables can depend on phase through tangent restriction and the polynomial derivative.

3.2 A residue sampler for product-projective hypersurfaces, validated against exact volumes

A geometric flatness diagnostic depends on its integration measure. We derive the bicubic residue sampler and check its absolute normalization against intersection-theoretic volumes before comparing representations.

Table 5: Absolute validation of the bicubic residue sampler against exact intersection-theoretic volumes, 200,000 points per estimate. The expected values are the closed form (6).

(t_x, t_y)	Estimate	Expected	Relative error
(1, 1)	2.9913 ± 0.0213	3	-0.29%
(2, 1)	8.9788 ± 0.0432	9	-0.24%
(2, 2)	23.9300 ± 0.1701	24	-0.29%
(3, 1)	17.9625 ± 0.0661	18	-0.21%

The geometry is the bicubic threefold

$$F(x, y) = 10 \sum_a x_a^3 y_a^3 + 3 \sum_{a \neq b} x_a^3 y_b^3 = 0 \quad \subset \quad \mathbb{CP}_x^2 \times \mathbb{CP}_y^2. \quad (5)$$

In the chart $x_0 = y_2 = 1$ the bicubic is diagonal in the cubic powers of y , so one y coordinate can be eliminated analytically with three branches. Sampling x from the Fubini–Study measure, the free y coordinate from a complex-Cauchy proposal, and a branch uniformly gives an importance sampler for the residue-form measure with weight $|\partial F / \partial y_\delta|^{-2} / q$. Single-channel proposals of this kind have heavy empirical tails. Their estimated Hill tail index is near 1.5 [31], which is consistent with, but does not analytically establish, an infinite-variance regime. The observed effective-sample-size fraction declines as the sample size increases, so we do not use pointwise bootstrap errors from these proposals as an acceptance test. We therefore combine the y_0 - and y_1 -elimination channels with the balance heuristic [32], whose chart-independent weight $2 / (w_0^{-1} + w_1^{-1})$ yields a materially larger effective-sample-size fraction that is stable across our independent replicate audits.

The acceptance test is absolute rather than symmetry-based. For a hypersurface of bidegree $(3, 3)$ in $\mathbb{CP}^2 \times \mathbb{CP}^2$, intersection theory fixes the volume in the class $t_x \omega_x + t_y \omega_y$ exactly:

$$V_{\text{norm}}(X; t_x, t_y) := \pi^{-3} \int_X \frac{\omega_{\text{FS}}^3}{3!} = \frac{3}{2} t_x t_y (t_x + t_y), \quad (6)$$

where the code uses $\phi_{\text{FS}} = t_x \log \|x\|^2 + t_y \log \|y\|^2$ and $\omega_{\text{FS}} = (i/2) \partial \bar{\partial} \phi_{\text{FS}}$. In this convention a projective line has area π ; hence the explicitly normalized volume above. Define $\eta_{\text{FS}} = \det(g_{X, \text{FS}}) |\partial F_{\text{aff}} / \partial w_\delta|^2$ at these same Kähler parameters. For the ideal equal-mixture proposal with raw residue importance weight W , $\mathbb{E}[W \eta_{\text{FS}}] / \pi^3 = V_{\text{norm}}$; Appendix C derives every factor. Table 5 agrees with the exact values to within 0.3%. The checks constrain the absolute normalization and jointly test tangent restriction and projective factors. They are correlated numerical checks on a shared sample, rather than a proof of the screened sampler’s exact law. Two symmetry-related proposals can agree despite a shared normalization error, so symmetry checks alone cannot replace this test.

3.3 Representation comparison on the symmetric bicubic

The bicubic adds a product-projective ambient space, two Kähler parameters and a tractable exact orbit average. The effective symmetry of Eq. (5) is $((\mathbb{Z}_3)_x^2 \times (\mathbb{Z}_3)_y^2) \rtimes (S_3 \times \mathbb{Z}_2)$, of order 972; all experiments use $t_x = t_y = 1$. Whereas Lee and Lukas [26] fit analytic Kähler potentials to numerical bicubic metrics, we compare residual representations at fixed measure and objective.

The block-exchange $\mathbb{Z}_2$ is implemented separately by symmetrizing slot features and character blocks. The declarative specification covers the diagonal subgroup and simultaneous S_3 , as detailed in Appendix D.4.

Table 6: Bicubic principal comparison at 5,000 training points. Late-training validation $\sigma(\eta)$ is the mean $\pm$ population standard deviation over three seeds, each contributing the mean of its final five evaluations. The projector MLP receives complete rank-one projectors and is globally projective, but has no declared bicubic phase or permutation symmetry. Positivity was monitored on randomized 512-point subsets.

Residual	Parameters	$\sigma(\eta)$	Positivity violations (%)
Moduli-only	29,093	0.2320 ± 0.0091	0.00
Projective MLP control	29,083	0.1649 ± 0.0163	0.00
Character phase-aware	29,249	0.0230 ± 0.0023	0.00

Table 7: Bicubic fixed-step data-size and generic-phase-control check. Each arm receives exactly 1,000 optimizer updates; values are mean $\pm$ population standard deviation over three seeds, each contributing the mean of its final five evaluations. The 5,000-point row restates the principal condition in Table 6; the additional rows test data-size sensitivity. The moduli-only, projective-MLP, and character models have 29,093, 29,083, and 29,249 parameters, respectively. The projective MLP receives rank-one projectors but no declared bicubic characters or permutation pooling. No positivity violation was observed in the sampled diagnostics.

Training points	Moduli-only $\sigma(\eta)$	Projective MLP $\sigma(\eta)$	Character phase-aware $\sigma(\eta)$
5,000	0.2320 ± 0.0091	0.1649 ± 0.0163	0.0230 ± 0.0023
12,000	0.2327 ± 0.0091	0.1583 ± 0.0224	0.0228 ± 0.0014
80,000	0.2310 ± 0.0100	0.1770 ± 0.0357	0.0240 ± 0.0016

3.3.1 Principal comparison

The principal comparison uses 5,000 training points, 2,000 validation points, 100 epochs and three seeds, with relative residual filtering, mixture-channel shuffling and randomized 512-point positivity subsets. Character pooling improves on both moduli-only pooling and the complete projective-MLP control (Table 6). The full-orbit comparison follows in Table 8.

3.3.2 Fixed-step data-size and generic-phase-control check

We repeated all three arms at 12,000 and 80,000 training points while fixing the budget at 1,000 updates and holding the schedule, validation size and seeds constant. The character advantage persists across data sizes (Table 7): late $\sigma(\eta)$ is about 90% lower than moduli-only pooling and 86% lower than the projective MLP, with little change across the sixteenfold data increase.

3.3.3 Full-orbit reference at matched update budgets

Because the declared bicubic group has order 972, exact orbit averaging is tractable at a small batch size. We therefore compare character pooling with a shared projector MLP averaged over all 972 actions. The models have 29,249 and 29,083 parameters, respectively. Each comparison uses 2,000 training points, batch size 8, identical seeds, schedule and update budget. Full-orbit validation is evaluated in fixed point chunks for memory control, followed by the exact weighted reduction over the complete held-out set.

Character pooling has the lower mean diagnostic at 30 updates, but exact orbit averaging is lower at 200 and 1,000 updates (Table 8). The ordering depends on the optimization budget. As in

Table 8: Matched-update bicubic comparison with an exact 972-element orbit average. Each row uses 2,000 training points, batch size 8, three seeds and near-matched parameter counts. Validation size is matched between arms within each row. Values are the late-training held-out $\sigma(\eta)$, mean $\pm$ population standard deviation; for the 1,000-update row, each seed contributes its final three evaluations. No positivity violation was observed on the stated diagnostic subsets.

Updates	Validation points	Character pooling	Exact orbit average
30	128	0.2671 ± 0.0270	0.4835 ± 0.0302
200	512	0.1033 ± 0.0086	0.0896 ± 0.0052
1,000	512	0.0629 ± 0.0013	0.0347 ± 0.0008

the H_3^+ control, an early character-model advantage does not establish superiority at longer training budgets.

4 Discussion

The study makes over-invariance measurable. Conditional variance identifies the target variation discarded by a representation, and the synthetic benchmark supplies an exact error floor. The molecular convergence control shows why a persistent representation deficit and a short-budget optimization deficit require different diagnoses. These results give an operational test for the established concern about excessive invariance [4, 5].

The geometric applications add two independent checks. Absolute volume tests assess the integration measure, while boundary tests assess global gluing. The canonicalization control passes the action-error test and achieves low local loss, yet fails the gluing audit. Symmetry, local accuracy and global regularity therefore require separate validation. Smooth invariant residuals address the gluing requirement by construction, with their assumptions and proofs stated in Appendix D.

The orbit comparisons establish a budget-dependent trade-off. Character pooling reaches low error earlier per update; exact averaging closes the gap on H_3^+ and overtakes it on the bicubic at longer budgets. PCP avoids explicit enumeration, so its cost depends on the supplied coordinates and channels. A direct matched-cost training comparison for the 15,000-element Dwork group remains the test of its practical large-group advantage.

The geometric measurements establish determinant-ratio flatness improvements with sampled positivity checks. Three-seed comparisons remain descriptive, and certified metric accuracy and downstream observables require further validation. The finite-teacher fibre witness quantifies variation on its declared support; the continuum criterion specifies the additional rank and derivative certificate needed to extend it. Full fifth-power orbit separation is classical background, whereas completeness of the finite PCP channels is unresolved.

The resulting workflow is reusable: measure information loss, check optimization at several budgets, validate the integration measure, and test global gluing independently of action error. These steps identify which constraint limits a learned geometric field and which improvement the evidence supports.

5 Methods

5.1 Projective potential and symmetry specification

PCP represents a scalar residual on B aligned projective blocks, each with n coordinates. For fixed positive Kähler weights t_k , set

$$\phi_\theta = \phi_{\text{base}} + \alpha r_\theta, \quad \phi_{\text{base}}(z) = \sum_{k=1}^B t_k \log \|z^{(k)}\|^2. \quad (7)$$

The residual is invariant under independent block rescalings. The base carries the projective transformation law

$$\phi_\theta(\lambda_1 z^{(1)}, \dots, \lambda_B z^{(B)}) = \phi_\theta(z) + \sum_{k=1}^B t_k \log |\lambda_k|^2. \quad (8)$$

Local potentials therefore differ by a function with zero complex Hessian, so their induced forms agree on chart overlaps.

A specification supplies degrees d_k , a diagonal subgroup of the coordinatewise d_k th-root rotations, and a slot-permutation subgroup $\Pi \leq S_n$ acting simultaneously on all blocks. The permutations must normalize the diagonal subgroup. We denote its effective projective action by $\mathcal{A}$, quotienting common block rotations when they act trivially. For the Dwork quintic,

$$\mathcal{A}_{\text{Dw}} = \left\{ a \in (\mathbb{Z}_5)^5 : \sum_i a_i = 0 \right\} / \langle (1, \dots, 1) \rangle \quad (9)$$

has order 125; adding S_5 gives 15,000 actions. The bicubic has independent projective $(\mathbb{Z}_3)^2$ factors and simultaneous S_3 permutations. Its block exchange is implemented separately. Table 11 records the supported action class.

5.2 Character pooling

Normalize each nonzero block by $\hat{z}^{(k)} = z^{(k)} / \|z^{(k)}\|$ and define $u_i^{(k)} = |\hat{z}_i^{(k)}|^2$. One shared real encoder maps each aligned slot $\mathbf{u}_i = (u_i^{(1)}, \dots, u_i^{(B)})$ to $h_i = E_\theta(\mathbf{u}_i) \in \mathbb{R}^L$. Sum and mean pooling give $\text{pool}(h) = [\sum_i h_i, n^{-1} \sum_i h_i]$.

For an integer p , define the signed monomial

$$\mathcal{M}_p(v) = \begin{cases} v^p, & p > 0, \\ 1, & p = 0, \\ \bar{v}^{-p}, & p < 0. \end{cases} \quad (10)$$

For each supplied charge vector $\mathbf{p}^{(t)} \in \mathbb{Z}^B$, form

$$\varphi_i^{(t)} = \prod_{k=1}^B \mathcal{M}_{p_k^{(t)}}((\hat{z}_i^{(k)})^{d_k}), \quad (11)$$

then compute

$$S_c^{(t)} = \sum_i h_{i,c} \varphi_i^{(t)}, \quad \Phi_c^{(t)} = |S_c^{(t)}|^2. \quad (12)$$

The degree- d_k powers remove the declared root rotations while preserving relative-phase information in the contractions. Taking absolute squares removes the common phase of each block.

An optional global term (β, m, s) , with $m \geq 0$ an integer, adds

$$P_\beta = \prod_i \hat{z}_i^{(\beta)}, \quad Q^{(s)} = \sum_c S_c^{(s)}, \quad D = P_\beta^m \overline{Q^{(s)}}. \quad (13)$$

Projective invariance requires cancellation in every block:

$$nm \delta_{k\beta} = d_k p_k^{(s)} \quad \text{for all } k. \quad (14)$$

Finite-group invariance additionally requires

$$m \sum_i a_i^{(\beta)} = 0 \pmod{d_\beta} \quad \text{for every lifted diagonal action } a. \quad (15)$$

The decoder receives $\text{pool}(h)$, all $\Phi_c^{(t)}$, and $(\text{Re } D, \text{Im } D, |D|^2)$ for each global term. Its final layer is initialized at zero, so training begins at the base potential. All reported PCP models use tanh activations.

The declarative Dwork specification uses $B = 1$, $n = d_1 = 5$, $\mathbf{p}^{(1)} = (1)$ and $(\beta, m, s) = (1, 1, 1)$. The bicubic specification uses $B = 2$, $n = d_1 = d_2 = 3$, $\mathbf{p}^{(t)} \in \{(1, 0), (0, 1), (1, -1)\}$ and no global term.

5.3 Invariance and smoothness

Proposition 1 (Conditional invariance). *For every parameter setting, the residual above is invariant under independent projective block rescalings when every global term satisfies Eq. (14). It is invariant under the declared diagonal root-phase subgroup and simultaneous slot permutations when every global term also satisfies Eq. (15).*

Proposition 2 (Smooth projective descent). *With smooth encoder and decoder maps, the invariant residual descends to a unique smooth real function on the product projective space. Its complex Hessian defines a global real closed $(1, 1)$ -form.*

Appendix D gives the full proofs, the separate bicubic exchange construction and numerical implementation details. The classical Dwork quotient identifies full projective fifth-power data that separate generic finite orbits [33, 34]; Theorem 3 and its proof are included there. The completeness of the finite PCP feature set is an open question.

5.4 Metric diagnostic and training

We use the convention

$$g_{a\bar{b}} = \partial_{w_a} \partial_{\bar{w}_b} \phi_\theta, \quad \omega_\phi = \frac{i}{2} \sum_{a,b} g_{a\bar{b}} dw_a \wedge d\bar{w}_b. \quad (16)$$

In each projective factor, the largest coordinate selects the affine chart. The largest defining-polynomial derivative selects the eliminated coordinate δ . Both discrete selections are held fixed during automatic differentiation. Restricting the ambient complex Hessian to the hypersurface tangent space gives g_X . If $Q_{\text{aff}} = 0$ is the local equation, the determinant-to-residue ratio is

$$\eta = \frac{dV_{g_X}}{d\nu_\Omega} = \det g_X \left| \frac{\partial Q_{\text{aff}}}{\partial w_\delta} \right|^2. \quad (17)$$

For a positive metric on a compact connected Calabi–Yau manifold, η is constant exactly when the metric is Ricci flat. Appendix D.7 derives the tangent restriction, residue measure and chart independence.

For positive importance weights w_ℓ , define

$$\langle h \rangle_w = \frac{\sum_\ell w_\ell h_\ell}{\sum_\ell w_\ell}, \quad \sigma(\eta) = \left\langle \left(\frac{\eta}{\langle \eta \rangle_w} - 1 \right)^2 \right\rangle_w^{1/2}. \quad (18)$$

Training instead uses $\ell = \log \eta$, $\bar{\ell}_w = \langle \ell \rangle_w$ and

$$\mathcal{L}_{\text{Dwork}} = \langle (\ell - \bar{\ell}_w)^2 \rangle_w + 0.1 \bar{\ell}_w^2, \quad \mathcal{L}_{\text{bicubic}} = \langle (\ell - \bar{\ell}_w)^2 \rangle_w. \quad (19)$$

Each reported seed contributes its final five evaluations unless the table states another window. Results show the mean and population standard deviation across three seeds; percentage reductions describe ratios of those means. Positivity is monitored through held-out eigenvalue audits. Appendices C and D give the normalization, perturbation criterion and numerical stabilizers.

5.5 Molecular control and computational cost

The molecular dataset contains 2,000 RHF/STO-3G energies on the two-dimensional degenerate normal mode of equilateral H_3^+ . The reference bond length is 1.65 Bohr and the maximum displacement is 0.18 Bohr. Within each seed, all arms share the data split, minibatches, Adam schedule and 4,417 parameters. Batch size 512 gives three updates per epoch. Energies are standardized on the training set and errors converted back to Hartree squared. Appendix D.8 gives the representations and exact-invariance argument.

PCP evaluates the supplied encoder, character channels and decoder without enumerating group actions. Its contractions cost $O(nqL)$ for n slots, q terms and width L ; Appendix D.9 includes all encoder, power and decoder costs. Direct exact orbit averaging evaluates a base network $|G|$ times. The action-error measurements in Appendix A.2 test the numerical invariance of the supplied representations.

A Secondary representation checks

A.1 Bidirectional within-family transfer

For each architecture and seed, we initialized a target- ψ Dwork run either from the Fubini–Study base potential or from the best compatible source- ψ checkpoint. Both target runs used newly sampled target data, a fresh optimizer state, the same 20,000/2,000 train/validation split sizes, batch size 1024, learning-rate schedule, and 50-epoch budget. This tests within-family initialized fine-tuning.

Fine-tuning improves both architectures, with larger reductions for the phase-aware residual in both directions (Table 9). All final positivity diagnostics reported zero observed violations.

A.2 Single-block recompilation and sampled-orbit action exactness

Recompiling the hand-written Dwork residual from its declarative specification ($B = 1$, $n = 5$, $d_1 = 5$, one character term and one global product) gives 0.0324 ± 0.0017 , against 0.0309 ± 0.0015 at the same 25,217 parameters. This checks implementation transfer for the single-block specification.

On batches of 128 random Dwork homogeneous coordinates, parameter-matched sampled-orbit residuals approach invariance as their orbit budget increases but retain nonzero held-out action errors at $K = 512$. Character pooling is invariant to floating-point precision (Table 10).

Table 9: Bidirectional Dwork initialized fine-tuning. Late-training target validation $\sigma(\eta)$, mean $\pm$ population standard deviation over three seeds; each seed contributes the mean of its final five evaluations. Scratch and transfer arms have identical target data sizes, schedules and 50-epoch update budgets. Reductions are relative to scratch.

Source $\rightarrow$ target	Residual	Scratch	Transfer	Reduction
0.3 $\rightarrow$ 0.8	Moduli-only Deep Sets	0.1221 ± 0.0019	0.1124 ± 0.0065	8%
0.3 $\rightarrow$ 0.8	Dwork-adapted $m = 1$	0.1072 ± 0.0072	0.0580 ± 0.0058	46%
0.8 $\rightarrow$ 0.3	Moduli-only Deep Sets	0.0535 ± 0.0006	0.0451 ± 0.0006	16%
0.8 $\rightarrow$ 0.3	Dwork-adapted $m = 1$	0.0436 ± 0.0012	0.0301 ± 0.0017	31%

Table 10: Dwork representation exactness benchmark at matched parameter counts. The action error is the maximum absolute held-out difference over the stated action audit. Sampled orbit averages use independently sampled, non-closed subsets and are therefore only approximately invariant.

Latent	Method	K	Max. action error
32	Character pooling	–	5.55×10^{-16}
32	Sampled orbit average	1	0.657
32	Sampled orbit average	8	0.184
32	Sampled orbit average	64	0.112
32	Sampled orbit average	512	0.0223
64	Character pooling	–	6.66×10^{-16}
64	Sampled orbit average	1	0.859
64	Sampled orbit average	8	0.379
64	Sampled orbit average	64	0.0985
64	Sampled orbit average	512	0.0758

B Representation floors and their continuum criterion

B.1 Conditional expectation gives the optimal representation predictor

Let $(\mathcal{P}, \mathcal{F}, \mu)$ be a probability space, let $T \in L^2(\mu)$ be real-valued, and let $\pi : \mathcal{P} \rightarrow \mathcal{U}$ be a measurable map to a standard Borel space. Write $\mathcal{G} = \sigma(\pi)$ and $m = \mathbb{E}[T \mid \mathcal{G}]$. Conditional Jensen gives $\mathbb{E}[m^2] \leq \mathbb{E}[T^2]$, so $m \in L^2(\mathcal{G})$. For a bounded $\mathcal{G}$ -measurable h , the defining property of conditional expectation gives

$$\mathbb{E}[(T - m)h] = \mathbb{E}[h \mathbb{E}[T - m \mid \mathcal{G}]] = 0. \quad (20)$$

For general $h \in L^2(\mathcal{G})$, apply this to the truncations $h_N = \max(-N, \min(h, N))$. They converge to h in L^2 , so Cauchy–Schwarz allows passage to the limit. Thus $T - m$ is orthogonal to every element of $L^2(\mathcal{G})$. For any $g \in L^2(\mathcal{G})$, expansion of the square therefore gives

$$\|T - g\|_{L^2(\mu)}^2 = \|T - m\|_{L^2(\mu)}^2 + \|m - g\|_{L^2(\mu)}^2 + 2\mathbb{E}[(T - m)(m - g)] \quad (21)$$

$$= \|T - m\|_{L^2(\mu)}^2 + \|m - g\|_{L^2(\mu)}^2. \quad (22)$$

The minimum is attained by $g = m$. By the definition of conditional variance and the tower property,

$$\mathbb{E}[\text{Var}(T \mid \pi)] = \mathbb{E}[\mathbb{E}[(T - m)^2 \mid \mathcal{G}]] = \|T - m\|_{L^2(\mu)}^2. \quad (23)$$

Consequently,

$$\inf_{g \in L^2(\mathcal{P}, \mathcal{G}, \mu)} \|T - g\|_{L^2(\mu)}^2 = \mathbb{E}[\text{Var}(T \mid \pi)]. \quad (24)$$

The measurable-factorization lemma writes $m = f_* \circ \pi$ almost everywhere; $\|f_*\|_{L^2(\pi_*\mu)} = \|m\|_{L^2(\mu)}$. This gives the equivalent infimum over $f \in L^2(\mathcal{U}, \pi_*\mu)$ used in the main text. This projection identity is a standard fact of conditional expectation.

B.2 Exact finite-fibre specialization

Suppose the support consists of J fibres $F_j = \{z_{j1}, \dots, z_{jm}\}$, with $\pi(z_{j\ell}) = u_j$ and pairwise distinct u_j , and assigns mass $1/(Jm)$ to each point. Put $\bar{T}_j = m^{-1} \sum_{\ell} T(z_{j\ell})$. Since $\sum_{\ell} (T(z_{j\ell}) - \bar{T}_j) = 0$, direct expansion gives

$$\frac{1}{m} \sum_{\ell=1}^m (T(z_{j\ell}) - f(u_j))^2 = \frac{1}{m} \sum_{\ell=1}^m (T(z_{j\ell}) - \bar{T}_j)^2 + (\bar{T}_j - f(u_j))^2. \quad (25)$$

Distinct u_j permit choosing $f(u_j) = \bar{T}_j$ independently. Averaging over j yields

$$\inf_f \frac{1}{Jm} \sum_{j,\ell} (T(z_{j\ell}) - f(u_j))^2 = \frac{1}{J} \sum_{j=1}^J \text{Var}_{\ell=1,\dots,m}[T(z_{j\ell})], \quad (26)$$

where the finite variance uses denominator m , not $m-1$. This is exact for exact common fingerprints. A tolerance test on floating-point fingerprints is evidence of numerical agreement; it does not by itself prove exact equality for arbitrary measurable predictors. If the numerical fingerprints are all distinct, an unrestricted predictor can interpolate each point. The empirical floor must therefore be understood on the declared ideal equal-fibre support, or on explicitly identified fibre labels, with moduli drift and hypersurface residuals reported separately. Positive variance on that support does not itself establish a continuum result.

B.3 Derivation of the synthetic floor

To fix the sampling and normalization, start with independent circular complex Gaussian variables $g_1, \dots, g_n$ and set $\hat{z} = g/\|g\|$. Rotational invariance makes $\hat{z}$ uniform on the complex unit sphere. Writing $\hat{z}_i = \sqrt{u_i} e^{i\theta_i}$, the phases are independent uniform variables on $[0, 2\pi)$ and independent of u . The squared Gaussian magnitudes are independent exponential variables with a common scale, so normalizing their sum gives $u \sim \text{Dirichlet}(1, \dots, 1)$. Using the unit sphere retains an independent common phase in the lift; projectively invariant targets do not depend on that common phase.

For arbitrary fixed complex coefficients c_i and integer $d \geq 1$, write

$$C_c = \left| \sum_i c_i \hat{z}_i^d \right|^2 = \sum_i |c_i|^2 u_i^d + R, \quad (27)$$

$$R = \sum_{i \neq j} c_i \bar{c}_j u_i^{d/2} u_j^{d/2} e^{id(\theta_i - \theta_j)}. \quad (28)$$

For an integer $k \neq 0$, $(2\pi)^{-1} \int_0^{2\pi} e^{ik\theta} d\theta = 0$. Applying this independently to each phase gives

$$\mathbb{E}[C_c \mid u] = \sum_i |c_i|^2 u_i^d, \quad \mathbb{E}[R \mid u] = 0. \quad (29)$$

The characters $e^{id(\theta_i - \theta_j)}$ for ordered pairs $i \neq j$ are orthonormal: the expectation of one times the conjugate of another vanishes unless $e_i - e_j = e_p - e_q$, which for such ordered pairs requires

$i = p, j = q$. Here e_i is the i th standard basis vector of $\mathbb{Z}^n$. Since R is real (its summands occur in conjugate pairs), $R^2 = |R|^2$. Expanding that absolute square and using orthogonality gives

$$\text{Var}(C_c \mid u) = \mathbb{E}[|R|^2 \mid u] = \sum_{i \neq j} |c_i|^2 |c_j|^2 u_i^d u_j^d. \quad (30)$$

For $u \sim \text{Dirichlet}(1, \dots, 1)$, its density on the simplex is $\Gamma(n)$. The simplex beta integral therefore gives, for $i \neq j$,

$$\mathbb{E}[u_i^d u_j^d] = \Gamma(n) \frac{\Gamma(d+1)\Gamma(d+1) \prod_{k \neq i,j} \Gamma(1)}{\Gamma((d+1) + (d+1) + (n-2))} \quad (31)$$

$$= \frac{(n-1)!(d!)^2}{(n+2d-1)!}. \quad (32)$$

In $f_* = (1-\gamma)A(u) + \gamma C_c$, the first term is known given u , so the conditional variance is $\gamma^2 \text{Var}(C_c \mid u)$. Combining the preceding equations with the projection identity yields

$$\mathcal{E}_{\text{moduli}}^2 = \gamma^2 \frac{(n-1)!(d!)^2}{(n+2d-1)!} \sum_{i \neq j} |c_i|^2 |c_j|^2. \quad (33)$$

The main experiment uses $c_i = 1$, recovering Eq. (3). Equal coefficients $c_i = c$ instead multiply that expression by $|c|^4$. At $n = d = 5$ and $\gamma = 1$, the unscaled floor is $7.928579357 \times 10^{-5}$; choosing $c = 1/\sqrt{5}$ would divide it by 25. Unit coordinate phases $z_i/|z_i|$ would define a different target, lose the u_i^d factors, and be undefined at coordinate zeros; they are not used here.

The synthetic model also centres each learned phase channel conditionally on u . If a real encoder produces $h_{i,c} = h_{i,c}(u)$, the same phase integration gives

$$\mathbb{E} \left[\left| \sum_i h_{i,c}(u) \hat{z}_i^d \right|^2 \mid u \right] = \sum_i h_{i,c}(u)^2 u_i^d. \quad (34)$$

Subtracting this right-hand side gives the zero-conditional-mean phase statistic supplied to the synthetic decoder. The no-phase arm substitutes a zero vector of identical shape. This derivation explains the centring used in the code; it does not change the target or its analytic floor.

B.4 A quantitative sufficient criterion for a continuum floor

Theorem 1 (Local vertical variation implies a positive continuum floor). *Let M and U be finite-dimensional second-countable smooth manifolds and $\pi : M \rightarrow U$ a smooth map. Let μ be a probability measure, and let $T \in L^2(M, \mu)$ be real-valued. Suppose that, near a point $x \in M$, T is C^1 , μ has a continuous strictly positive density with respect to a smooth volume density, and π has constant rank r . If there is $v \in \ker(d\pi_x)$ such that $dT_x(v) \neq 0$, then*

$$\inf_{f \in L^2(U, \pi_* \mu)} \|T - f \circ \pi\|_{L^2(\mu)}^2 = \mathbb{E}_\mu[\text{Var}_\mu(T \mid \pi)] > 0. \quad (35)$$

Proof. The equality follows from the projection identity. We give a local lower bound valid for every measurable predictor f , which proves strict positivity without assumptions about other components of a fibre.

The constant-rank theorem provides coordinates $(y, t) \in \mathbb{R}^r \times \mathbb{R}^k$ near x , with $k = \dim M - r \geq 1$, and a target chart in which $\pi(y, t) = \iota(y) = (y, 0)$. The vector v has only t components. A linear

change of the t coordinates and a possible sign reversal therefore make $\partial T/\partial t_1(x) > 0$. Continuity gives a smaller coordinate box on which $\partial T/\partial t_1 \geq a > 0$.

At the base coordinate y_0 of x , choose two points (y_0, t^-) and (y_0, t^+) in this box, differing only by $t_1^+ - t_1^- = h > 0$. The fundamental theorem of calculus gives

$$T(y_0, t^+) - T(y_0, t^-) = \int_{t_1^-}^{t_1^+} \frac{\partial T}{\partial t_1}(y_0, s, t_2^-, \dots, t_k^-) ds \geq ah. \quad (36)$$

By continuity choose a bounded base box Y and disjoint vertical boxes B_-, B_+ of equal Lebesgue volume $b > 0$, with compact closures in the coordinate neighbourhood, such that

$$T(y, t_+) - T(y, t_-) \geq \Delta > 0 \quad (y \in Y, t_{\pm} \in B_{\pm}). \quad (37)$$

The coordinate density of μ includes the volume-density Jacobian. It is continuous and positive, so it has a lower bound $\rho_0 > 0$ on these compact boxes.

Fix y and set $c = f(\iota(y))$. For any real A, B, c ,

$$(A - c)^2 + (B - c)^2 = \frac{1}{2}(A - B)^2 + 2\left(c - \frac{A+B}{2}\right)^2 \geq \frac{1}{2}(A - B)^2. \quad (38)$$

Insert $A = T(y, t_+)$ and $B = T(y, t_-)$ and integrate over $B_+ \times B_-$. Using Eq. (37), the left-hand side is b times the sum of the two single-box error integrals, while the right-hand side is at least $b^2 \Delta^2/2$. Dividing by b gives

$$\int_{B_+ \cup B_-} (T(y, t) - f(\iota(y)))^2 dt \geq \frac{b \Delta^2}{2}. \quad (39)$$

Finally, restricting the global nonnegative error integral to these boxes and using the density lower bound yields

$$\|T - f \circ \pi\|_{L^2(\mu)}^2 \geq \rho_0 |Y| \frac{b \Delta^2}{2} > 0. \quad (40)$$

This bound is independent of f , so it survives the infimum. If $r = 0$, there is no base variable; the same proof uses $|Y| = 1$. $\square$

For the Dwork application, smoothness can be checked directly. At a singular point of $F_\psi = 0$, a zero coordinate would make $\partial_i F_\psi = 5z_i^4$ nonzero at any nonzero coordinate z_i , so every coordinate must be nonzero. The gradient equations then give $z_i^5 = \psi P$ for every i . Multiplying these five equalities gives $P^5 = \psi^5 P^5$, hence $\psi^5 = 1$. Conversely, when $\psi^5 = 1$, the point $[\psi^{-1} : 1 : 1 : 1 : 1]$ obeys the polynomial and all gradient equations. Thus $\psi^5 \neq 1$ is precisely the smoothness condition used below.

Corollary 2 (Dwork tangent certificate). *Fix a smooth Dwork quintic, so $\psi^5 \neq 1$, and let μ be a probability measure on X_ψ with smooth positive density. Let $T \in L^2(X_\psi, \mu)$ be real-valued and C^1 near $[z] \in X_\psi^\times$. Define $P = \prod_{i=0}^4 z_i$, $b_i = z_i^5 - \psi P$ and the real matrix*

$$B(z) = \begin{pmatrix} \operatorname{Re} b_0 & \cdots & \operatorname{Re} b_4 \\ \operatorname{Im} b_0 & \cdots & \operatorname{Im} b_4 \end{pmatrix}. \quad (41)$$

Suppose $\operatorname{rank}_{\mathbb{R}} B(z) = 2$ and there is $a \in \mathbb{R}^5$ such that

$$B(z)a = 0, \quad \text{equivalently} \quad \sum_i a_i z_i^5 - \psi \left(\sum_i a_i \right) P = 0, \quad (42)$$

with $dT_{[z]}([i(a_i z_i)_i]) \neq 0$. Then Eq. (35) is strictly positive for the projective moduli map $\pi([z]) = (|z_i|^2 / \sum_j |z_j|^2)_{i=0}^4$.

Proof. Work with a unit representative z . Differentiating $u_i = |z_i|^2/N$, $N = \sum_j |z_j|^2$, gives

$$\delta u_i = \frac{2 \operatorname{Re}(\bar{z}_i \delta z_i)}{N} - \frac{|z_i|^2}{N^2} 2 \sum_j \operatorname{Re}(\bar{z}_j \delta z_j). \quad (43)$$

If $\delta u_i = 0$ and all $z_i \neq 0$, division by $|z_i|^2$ shows that $\operatorname{Re}(\delta z_i/z_i)$ is the same real number for every i . Subtracting that common radial variation leaves $\delta z_i = i a_i z_i$ for real a_i . A common value of all a_i gives a projectively trivial common phase. Thus the ambient projective moduli kernel is precisely these phase directions modulo the common phase.

For $F_\psi(z) = \sum_i z_i^5 - 5\psi P$, the product rule gives $\delta P = iP \sum_i a_i$ and hence

$$dF_\psi[\delta z] = 5i \sum_i a_i z_i^5 - 5i\psi P \sum_i a_i = 5i \sum_i a_i b_i. \quad (44)$$

It vanishes exactly when $B(z)a = 0$. Moreover, $\sum_i b_i = F_\psi(z) = 0$, so $(1, \dots, 1)$ belongs to this real kernel. The rank-two assumption makes $\ker B(z)$ three-dimensional and leaves a two-dimensional vertical tangent space after quotienting the common phase. Since the smooth quintic has real dimension six, $\operatorname{rank} d\pi = 6 - 2 = 4$. A nonzero 2×2 minor stays nonzero near z ; four is the maximal rank because $\sum_i u_i = 1$. Thus π has constant rank four near $[z]$. The assumed nonzero target derivative is therefore exactly the hypothesis of Theorem 1. $\square$

Remark 1 (What remains to certify). The current H-matrix experiments establish numerical variation on declared finite fibres. They do not certify the rank and derivative hypotheses of Corollary 2. A rigorous application needs an exact or validated enclosure of a point on X_ψ , a nonzero rank minor, a tangent direction, a derivative interval excluding zero, and the stated regularity and L^2 properties of the chosen observable. Treating a stored matrix as exact binary data fixes the finite teacher; it does not validate the remaining geometric computations. Even a completed certificate would concern that finite teacher, rather than the exact Ricci-flat metric.

C Residue weights, volume conventions and positivity

C.1 The metric convention fixes the factor of π^3

For one affine coordinate w on $\mathbb{CP}^1$, the potential $\phi = \log(1 + |w|^2)$ gives

$$g_{w\bar{w}} = \partial_w \left(\frac{w}{1 + |w|^2} \right) = \frac{1}{(1 + |w|^2)^2}. \quad (45)$$

With the convention $\omega = (i/2)g_{w\bar{w}}dw \wedge d\bar{w}$ and $w = s + it$, $(i/2)dw \wedge d\bar{w} = ds \wedge dt$. In polar coordinates,

$$\int_{\mathbb{CP}^1} \omega = \int_0^{2\pi} \int_0^\infty \frac{r}{(1 + r^2)^2} dr d\theta = 2\pi \left[-\frac{1}{2(1 + r^2)} \right]_0^\infty = \pi. \quad (46)$$

Thus $[\omega/\pi]$ is the unit hyperplane class. Using $i\partial\bar{\partial}\phi$ instead would double the two-form and give 2π on $\mathbb{CP}^1$. Specifying the factor $1/2$ is therefore necessary for the absolute volume calculation.

For the product-projective bicubic use

$$\phi_{\text{FS}}(x, y; t_x, t_y) = t_x \log \|x\|^2 + t_y \log \|y\|^2, \quad t_x, t_y > 0. \quad (47)$$

Let $g_{X,\text{FS}}$ be its induced three-dimensional Hermitian matrix and put, in any regular elimination chart,

$$\eta_{\text{FS}}(x, y; t_x, t_y) = \det g_{X,\text{FS}} \left| \frac{\partial F_{\text{aff}}}{\partial w_\delta} \right|^2. \quad (48)$$

This is the same determinant-to-residue ratio as the learned diagnostic, evaluated at the stated product Fubini–Study potential. The Kähler parameters enter through this potential; the definition is not restricted to $(t_x, t_y) = (1, 1)$.

Write H_x, H_y for the unit hyperplane classes on the two factors. Their cohomology relations and normalization are

$$H_x^3 = H_y^3 = 0, \quad \int_{\mathbb{CP}^2 \times \mathbb{CP}^2} H_x^2 H_y^2 = 1. \quad (49)$$

A hypersurface of bidegree $(3, 3)$ has class $[X] = 3H_x + 3H_y$. The normalized Kähler class on X is $J = (t_x H_x + t_y H_y)|_X$. Expanding the cube gives

$$(t_x H_x + t_y H_y)^3 = 3t_x^2 t_y H_x^2 H_y + 3t_x t_y^2 H_x H_y^2, \quad (50)$$

because the pure cubic terms vanish. Multiplication by $[X]$ leaves

$$\int_X J^3 = \int_{\mathbb{CP}^2 \times \mathbb{CP}^2} (3H_x + 3H_y)(t_x H_x + t_y H_y)^3 \quad (51)$$

$$= 9t_x^2 t_y + 9t_x t_y^2. \quad (52)$$

The physical form used by the code has class πJ . Dividing its volume by π^3 therefore yields

$$\frac{1}{\pi^3} \int_X \frac{\omega_{\text{FS}}^3}{3!} = \frac{1}{6} \int_X J^3 = \frac{3}{2} t_x t_y (t_x + t_y). \quad (53)$$

Equation (6) denotes this normalized volume.

C.2 Single-channel proposals and the balance weight

The selected bicubic is smooth. To see this, write its coefficient matrix as $A = 7I + 3\mathbf{1}\mathbf{1}^T$. Every entry is nonzero, every 2×2 minor is 91 or ± 21 , and $\det A = 16 \cdot 7^2 = 784$. Thus every submatrix has full possible rank. If a projective point were singular, let I and J be the nonempty supports of x and y . The gradient equations at nonzero coordinates would give $A_{I,J}(y_j^3)_{j \in J} = 0$ and $A_{I,J}^T(x_i^3)_{i \in I} = 0$. Full row rank rules out the second equation when $|I| \leq |J|$, and full column rank rules out the first when $|I| \geq |J|$. This contradiction proves smoothness.

In the full-measure patch $x_0 = y_2 = 1$, set $x = (1, x_1, x_2)$ and write

$$F(x, y) = \sum_{j=0}^2 A_j(x) y_j^3, \quad A_j(x) = 3 \sum_{a=0}^2 x_a^3 + 7x_j^3. \quad (54)$$

For channel $\delta \in \{0, 1\}$ let $\epsilon = 1 - \delta$. Sample (x_1, x_2) and the free coordinate y_ϵ with normalized densities

$$q_x = \frac{2}{\pi^2(1 + |x_1|^2 + |x_2|^2)^3}, \quad q_y(y_\epsilon) = \frac{1}{\pi(1 + |y_\epsilon|^2)^2}. \quad (55)$$

For example, the second density integrates to one because $2 \int_0^\infty r(1 + r^2)^{-2} dr = 1$. For the first, the radial measure in $\mathbb{R}^4$ is $2\pi^2 r^3 dr$, so its integral is $4 \int_0^\infty r^3(1 + r^2)^{-3} dr = 1$.

The remaining coordinate satisfies

$$y_\delta^3 = -\frac{A_\epsilon(x) y_\epsilon^3 + A_2(x)}{A_\delta(x)}. \quad (56)$$

Away from $A_\delta = 0$ and multiple roots, choose each of the three roots with probability $1/3$. On each sheet, the proposal density relative to $d^6 \xi = d^2 x_1 d^2 x_2 d^2 y_\epsilon$ is $q_\delta = q_x q_y / 3$. The residue

coefficient in these coordinates is $1/F_{y_\delta}$, where $F_{y_\delta} = 3A_\delta y_\delta^2$. Therefore the importance weight for the unnormalized residue measure is

$$W_\delta = \frac{|F_{y_\delta}|^{-2}}{q_\delta}, \quad s_\delta = \frac{dP_\delta}{d\nu_\Omega} = q_\delta |F_{y_\delta}|^2 = W_\delta^{-1}. \quad (57)$$

Here P_δ is the ideal channel proposal on X . Expressing proposal density relative to the same ν_Ω makes s_0, s_1 comparable even though the two channels have different free coordinates. For $P_{\text{mix}} = \frac{1}{2}P_0 + \frac{1}{2}P_1$,

$$\frac{dP_{\text{mix}}}{d\nu_\Omega} = \frac{s_0 + s_1}{2}, \quad W_{\text{mix}} = \frac{d\nu_\Omega}{dP_{\text{mix}}} = \frac{2}{s_0 + s_1} = \frac{2}{W_0^{-1} + W_1^{-1}}. \quad (58)$$

At every sampled point both proposal densities must be evaluated. Using only the weight of the generating channel would be a different estimator.

For an integrable function h , cancellation of these Radon–Nikodym derivatives gives

$$\mathbb{E}_{P_{\text{mix}}}[W_{\text{mix}}h] = \int_X \frac{d\nu_\Omega}{dP_{\text{mix}}} h dP_{\text{mix}} = \int_X h d\nu_\Omega. \quad (59)$$

Taking $h = \eta_{\text{FS}}$ and using $\eta_{\text{FS}} d\nu_\Omega = dV_{g_{X,\text{FS}}}$ yields

$$\frac{\mathbb{E}_{P_{\text{mix}}}[W_{\text{mix}}\eta_{\text{FS}}]}{\pi^3} = \frac{3}{2}t_x t_y (t_x + t_y), \quad (60)$$

with no division by the total residue mass. Consequently absolute validation uses raw weights. Self-normalized weights instead estimate expectations under $\nu_\Omega/\nu_\Omega(X)$ and lose this absolute normalization.

These are identities for exact proposal sampling and exact local geometry. Equal deterministic counts from each independent proposal also satisfy Eq. (60) in expectation. Unequal channel fractions require their actual mixture coefficients. The finite implementation screens unstable denominators and polynomial residuals; a nonzero rejection probability conditions each proposal and may change its normalizing constant. The volume audit is an empirical check of the size of this numerical effect, not a proof that screened samples have the exact ideal law. The four volume checks use a common sample, so their errors are correlated; in particular, doubling both Kähler parameters multiplies the volume by $2^3 = 8$.

C.3 Uniform perturbation bound for positive definiteness

Proposition 3 (A sufficient positivity bound). *Let M be a compact complex manifold with a positive Hermitian metric g_0 whose form $\omega_0 = (i/2) \sum g_{0,a\bar{b}} dw_a \wedge d\bar{w}_b$ is Kähler. Let $r \in C^\infty(M, \mathbb{R})$ and $H_r = (\partial_a \partial_{\bar{b}} r)$. If*

$$\sup_{x \in M} \|g_0^{-1/2} H_r g_0^{-1/2}\|_{\text{op}} \leq c < 1, \quad (61)$$

then $g_0 + H_r$ is positive definite and $\omega_0 + (i/2)\partial\bar{\partial}r$ is Kähler.

Proof. In a g_0 -orthonormal frame, the perturbation is a Hermitian matrix A_x . The spectral theorem and Eq. (61) imply that every eigenvalue of A_x lies in $[-c, c]$. For any nonzero vector v ,

$$v^*(I + A_x)v \geq (1 - c)\|v\|^2 > 0. \quad (62)$$

Returning to arbitrary coordinates gives $g_0 + H_r \geq (1 - c)g_0$. The added form is real and closed because r is smooth and real, and $d\partial\bar{\partial}r = 0$. Thus positivity and closedness make the new form Kähler. With $r = \alpha r_0$, compactness makes $C = \sup \|g_0^{-1/2} H_{r_0} g_0^{-1/2}\|$ finite; any $|\alpha|C < 1$ suffices. This is a standard perturbation criterion. The sampled eigenvalue audits in this paper do not establish its uniform bound. $\square$

Table 11: Boundary of the implemented construction. Exactness refers to the conditional invariance statement of Proposition 1, not to completeness of the feature algebra or to universal approximation.

Action or requirement	Status	Treatment in this work
Diagonal root-phase rotations in the prescribed subgroup $\mathcal{A}$	Supported exactly	Degree- d_k monomials trivialize the declared characters; global products require trivial finite character on $\mathcal{A}$.
Simultaneous slot permutations in a subgroup Π normalizing $\mathcal{A}$	Supported exactly	Shared slot encoder followed by symmetric pooling.
Projective rescaling of each block	Supported exactly	Block normalization and per-block charge validation.
Specified charge-cancelling global contractions	Supported exactly when valid	Rejected at specification time unless the per-block condition in Eq. (14) holds.
Independent per-block permutations or block exchange	Not compiled	The bicubic exchange $\mathbb{Z}_2$ is supplied by a separate hand-written symmetrization.
Continuous groups, arbitrary unitary finite-group actions, or fermionic sign representations	Out of scope	Require representation-specific equivariant or antisymmetrization machinery not implemented here.
Exhaustive invariant-feature algebra or universal approximation	Not claimed	Evaluated empirically under matched budgets only.

D Construction proofs and numerical implementation

This appendix retains the full transformation and descent arguments. Methods states the construction and its guarantees; the derivations below establish each step and document numerical qualifications.

D.1 Boundary of the implemented action class

D.2 Projective and finite-group invariance

Proof of Proposition 1. Independent projective rescalings. Write $\lambda_k = r_k \omega_k$, where $r_k = |\lambda_k| > 0$ and $|\omega_k| = 1$. Directly from normalization,

$$\widehat{\lambda_k z^{(k)}} = \omega_k \hat{z}^{(k)}, \quad \frac{|\lambda_k z_i^{(k)}|^2}{\|\lambda_k z^{(k)}\|^2} = u_i^{(k)}. \quad (63)$$

Thus every $h_{i,c}$ and the moduli pool are unchanged. If $|\omega| = 1$, Eq. (10) gives $\mathcal{M}_p(\omega v) = \omega^p \mathcal{M}_p(v)$ for all integer p . For $p < 0$, this follows from $\bar{\omega} v^{-p} = \bar{\omega}^{-p} \bar{v}^{-p} = \omega^p \bar{v}^{-p}$. Consequently, with

$$\chi_t(\omega) = \prod_k \omega_k^{d_k p_k^{(t)}}, \quad (64)$$

one has

$$\varphi_i^{(t)} \mapsto \chi_t \varphi_i^{(t)}, \quad S_c^{(t)} \mapsto \chi_t S_c^{(t)}, \quad Q^{(t)} \mapsto \chi_t Q^{(t)}, \quad \Phi_c^{(t)} \mapsto |\chi_t|^2 \Phi_c^{(t)} = \Phi_c^{(t)}. \quad (65)$$

For one global term, $P_\beta^m \mapsto \omega_\beta^{nm} P_\beta^m$, so

$$D \mapsto \left[\prod_k \omega_k^{nm \delta_{k\beta} - d_k p_k^{(s)}} \right] D. \quad (66)$$

Every exponent vanishes under Eq. (14); hence the real, imaginary and absolute-square parts appended to the decoder are unchanged.

Finite root-phase actions. Such an action preserves each block norm and each modulus. For every slot,

$$(\zeta_k^{a_i^{(k)}} \hat{z}_i^{(k)})^{d_k} = \zeta_k^{d_k a_i^{(k)}} (\hat{z}_i^{(k)})^{d_k} = (\hat{z}_i^{(k)})^{d_k}. \quad (67)$$

Thus φ , S , Φ and Q are unchanged. The global product, which contains the original normalized coordinates rather than their degree- d_β powers, instead obeys

$$P_\beta^m \mapsto \left(\prod_i \zeta_\beta^{a_i^{(\beta)}} \right)^m P_\beta^m = \zeta_\beta^{m \sum_i a_i^{(\beta)}} P_\beta^m. \quad (68)$$

Equation (15) makes this factor one. There is no additional d_β in this exponent because P_β is a product of first powers.

Simultaneous permutations. A slot permutation τ sends $\mathbf{u}_i$, h_i and $\varphi_i^{(t)}$ to their respective values at $\tau(i)$. Since τ is a bijection,

$$\sum_i h_{\tau(i),c} \varphi_{\tau(i)}^{(t)} = \sum_j h_{j,c} \varphi_j^{(t)}, \quad \prod_i \hat{z}_{\tau(i)}^{(\beta)} = \prod_j \hat{z}_j^{(\beta)}. \quad (69)$$

All decoder inputs are therefore unchanged, which proves residual invariance. Finally, the total potential has the stated cocycle when the base has Eq. (8); finite-action invariance of the total potential also uses invariance of the base, which holds for Eq. (7). $\square$

D.3 Smooth projective descent

Proof of Proposition 2. Smoothness before quotienting. On $\mathbb{C}^n \setminus \{0\}$, $N(z) = \sum_i |z_i|^2$ is strictly positive and smooth as a real function. Thus $N^{-1/2}$, $z_i N^{-1/2}$ and $|z_i|^2/N$ are smooth even where an individual coordinate vanishes. Ordinary powers and conjugate powers in Eq. (10), and nonnegative powers of P_β , remain smooth there. A finite composition of these operations with smooth encoder and decoder maps is smooth on the product of punctured blocks.

Descent. Restrict this residual to the product of unit spheres $S = \prod_{k=1}^B S^{2n-1}$. The map $q : S \rightarrow \prod_k \mathbb{CP}^{n-1}$ is the product Hopf bundle with fibres $U(1)^B$. Proposition 1 makes the restricted residual constant on each fibre. Define $r(q(z)) = \phi_{\text{res},\theta}(z)$; this is well-defined and unique because q is surjective. On an affine patch, choose in each block the smooth unit representative whose patch coordinate is positive real. This gives a smooth local section s of q . The descended function on that patch is $r = \phi_{\text{res},\theta} \circ s$, hence is smooth. Two such sections differ by $U(1)^B$, so their local functions agree on overlaps.

Global differential form. Since r is a globally defined smooth real function, $i\partial\bar{\partial}r$ is globally defined. It is real because complex conjugation reverses both i and the order of the two differentials, and it is closed because $d = \partial + \bar{\partial}$, $\partial^2 = \bar{\partial}^2 = 0$ and $\partial\bar{\partial} = -\bar{\partial}\partial$. Its restriction to a smooth hypersurface remains a global real closed $(1,1)$ -form. These statements alone do not imply positive definiteness. $\square$

D.4 Bicubic exchange symmetry

The experimental bicubic model uses an additional construction for exchange $x \leftrightarrow y$. At slot i , its real encoder receives

$$(u_i^x + u_i^y, u_i^x u_i^y, (u_i^x - u_i^y)^2), \quad (70)$$

so exchange leaves each encoded value $h_{i,c}$ unchanged. Put $p_i = \hat{x}_i^3$ and $q_i = \hat{y}_i^3$, and form

$$S_{x,c} = \sum_i h_{i,c} p_i, \quad S_{y,c} = \sum_i h_{i,c} q_i, \quad S_{xy,c} = \sum_i h_{i,c} p_i \bar{q}_i. \quad (71)$$

Exchange sends $S_x \leftrightarrow S_y$ and $S_{xy,c} \mapsto \overline{S_{xy,c}}$, since $h_{i,c}$ is real. Thus the two phase-feature blocks $|S_{x,c}|^2 + |S_{y,c}|^2$ and $|S_{xy,c}|^2$ are unchanged. They are also invariant under independent block phases: S_x, S_y acquire their respective cubic phases, while S_{xy} acquires their ratio, all removed by the absolute squares. Cubic-root actions fix p_i, q_i pointwise, and simultaneous permutations merely reindex the sums. The decoder therefore preserves all declared bicubic symmetries. The base is exchange invariant only when $t_x = t_y$, as required in the experiments. This proves the hand-written symmetrization used here; it does not add exchange handling to the generic specification compiler.

D.5 Charge validation and numerical stabilizers

Remark 2 (Implementation note on charge validation). The reference implementation validates the per-block projective condition (14) at specification time. An aggregate equality such as $\sum_k d_k p_k^{(s_r)} = n m_r$ is insufficient: a source term with charge in another block can satisfy that weaker equality while producing a non-invariant contraction. Finite root-phase invariance additionally requires the geometric charge-subgroup condition (15). The code’s generic specification does not store or automatically enumerate $\mathcal{A}$; this finite-group condition must be verified mathematically for a supplied specification. The present Dwork specification obeys it because its charges sum to zero, while the bicubic declarative specification has no global product term. Numerical denominator stabilizers implement the ideal formulas only up to roundoff-scale corrections. The exact projective theorem applies to division by $\|z\|$, whereas a fixed addition such as $\sqrt{\|z\|^2 + 10^{-20}}$ is not exactly homogeneous for arbitrarily small representatives. Our action audits use well-scaled representatives and report numerical errors.

D.6 Generic Dwork orbit separation

The following elementary form of the classical Dwork quotient is independent of the neural parameterization. The quotient by fifth-root rotations and its fifth-power coordinates are described in the mirror-quintic construction [33, 34]. We include the proof to make the retained information explicit. Let

$$X_\psi^\times = \{[z] \in X_\psi : z_i \neq 0 \text{ for } i = 0, \dots, 4\}, \quad (72)$$

and define the projective fifth-power map

$$\mathcal{P}([z_0 : \dots : z_4]) = [z_0^5 : \dots : z_4^5]. \quad (73)$$

For $\psi \neq 0$, its fibres are exactly the diagonal Dwork symmetry orbits.

Theorem 3 (Generic Dwork orbit separation). *Fix $\psi \neq 0$. On $X_\psi^\times$, two points have equal images under Eq. (73) if and only if they differ by an element of $\mathcal{A}_{\text{Dw}}$. Consequently,*

$$X_\psi^\times / (\mathcal{A}_{\text{Dw}} \rtimes S_5) \longrightarrow \mathbb{CP}^4 / S_5, \quad (\mathcal{A}_{\text{Dw}} \rtimes S_5)[z] \longmapsto S_5 \mathcal{P}([z]) \quad (74)$$

is injective.

Proof. Well-definedness and forward implication. Replacing z by λz multiplies all entries of $\mathcal{P}([z])$ by λ^5 , which leaves the projective image unchanged. An element of the diagonal group is represented

by $(\alpha_i) \in \mu_5^5$ with $\prod_i \alpha_i = 1$, where μ_5 denotes the fifth roots of unity. It fixes each z_i^5 because $\alpha_i^5 = 1$. Thus every diagonal orbit lies in one fibre of $\mathcal{P}$.

Reverse implication. Equality of the projective images means that there is $c \in \mathbb{C}^*$ with $w_i^5 = cz_i^5$ for every i . Choose $\lambda \in \mathbb{C}^*$ satisfying $\lambda^5 = c$ and put $w' = \lambda^{-1}w$. Then $(w'_i)^5 = z_i^5$. Since $z_i \neq 0$, the ratios $\alpha_i = w'_i/z_i$ are well-defined and satisfy $\alpha_i^5 = 1$. Homogeneity of the Dwork polynomial implies $w' \in X_\psi$, and hence

$$0 = \sum_i (w'_i)^5 - 5\psi \prod_i w'_i \quad (75)$$

$$= \sum_i z_i^5 - 5\psi \left(\prod_i \alpha_i \right) \prod_i z_i. \quad (76)$$

Subtracting $\sum_i z_i^5 - 5\psi \prod_i z_i = 0$ gives

$$5\psi \left(1 - \prod_{i=0}^4 \alpha_i \right) \prod_{i=0}^4 z_i = 0. \quad (77)$$

Since $\psi \neq 0$ and every z_i is nonzero, $\prod_i \alpha_i = 1$. Thus (α_i) defines an element of $\mathcal{A}_{\text{Dw}}$ and $[w] = [w'] = [(\alpha_i z_i)_i]$. Changing λ by a fifth root ξ changes every α_i by ξ^{-1} . This is the projectively trivial common rotation already quotiented in Eq. (9). The charge-zero subgroup in $(\mathbb{Z}_5)^5$ has 5^4 elements, and its common-rotation subgroup has 5 elements, giving $|\mathcal{A}_{\text{Dw}}| = 5^3$.

Permutation quotient. The map $\mathcal{P}$ commutes with coordinate permutations. If two images in $\mathbb{CP}^4/S_5$ agree, there is a permutation τ such that $\mathcal{P}([w]) = \mathcal{P}(\tau[z])$. The reverse implication just proved supplies $a \in \mathcal{A}_{\text{Dw}}$ with $[w] = a\tau[z]$, so the two original points have the same full finite orbit. This proves injectivity of Eq. (74). $\square$

Corollary 4 (Continuous factorization on compact generic subsets). *Let $K \subset X_\psi^\times$ be compact and invariant under $G = \mathcal{A}_{\text{Dw}} \rtimes S_5$. Every continuous real-valued G -invariant target on K factors continuously and uniquely through the image of Eq. (74) restricted to K/G .*

Proof. Let $q : K \rightarrow K/G$ be the quotient map. Invariance defines $\bar{f}(q(x)) = f(x)$ independently of the representative x , and the quotient topology makes $\bar{f}$ continuous. Since the group is finite, K/G is compact and Hausdorff, as is the target $\mathbb{CP}^4/S_5$. The continuous injection $\bar{\mathcal{P}} : K/G \rightarrow \mathbb{CP}^4/S_5$ from Theorem 3 is therefore a homeomorphism onto its image: a closed subset of the compact domain has compact, hence closed, image in the Hausdorff target. Define $F = \bar{f} \circ \bar{\mathcal{P}}^{-1}$ on that image. Then $f = F \circ \bar{\mathcal{P}} \circ q$, and surjectivity onto the image gives uniqueness of F there. $\square$

Remark 3 (Scope of the separation theorem). The projective moduli are recoverable from the full fifth-power vector through

$$\frac{|z_i|^2}{\sum_j |z_j|^2} = \frac{|z_i^5|^{2/5}}{\sum_j |z_j^5|^{2/5}}. \quad (78)$$

The fractional-power expression is used only to recover moduli continuously from full fifth-power data; it is not a PCP feature. As a function of an arbitrary complex fifth-power coordinate it is not smooth at zero. Its pullback to the original coordinates equals the smooth quantity $|z_i|^2$. Thus full fifth-power data retain all projective moduli and the generic finite-orbit information. Theorem 3 does not assert that the finite pooled features in PCP separate these orbits, nor that they yield universal approximation. At $\psi = 0$ or on coordinate-zero strata, the product constraint used in the proof degenerates and the theorem does not apply.

D.7 Tangent restriction and the residue ratio

An adaptive affine chart is chosen by moving the largest-magnitude homogeneous coordinate into the patch position in each projective block. Its index is held constant during automatic differentiation. Write $Q_{\text{aff}}(w) = 0$ for the local hypersurface equation in four complex ambient coordinates. Choose an eliminated index δ with $Q_\delta = \partial Q_{\text{aff}}/\partial w_\delta \neq 0$ and let $\xi = (w_a)_{a \neq \delta}$. The implicit-function theorem gives

$$\frac{\partial w_\delta}{\partial \xi_a} = -\frac{Q_a}{Q_\delta}, \quad (g_X)_{a\bar{b}} = \sum_{j,k} \frac{\partial w_j}{\partial \xi_a} g_{j\bar{k}} \frac{\overline{\partial w_k}}{\partial \xi_b}. \quad (79)$$

The implementation chooses the largest $|Q_\delta|$ for conditioning and holds that index fixed in the local differentiation. With $w = x + iy$, the real Hessian gives

$$\phi_{w_a \bar{w}_b} = \frac{1}{4} (\phi_{x_a x_b} + \phi_{y_a y_b} + i\phi_{x_a y_b} - i\phi_{y_a x_b}), \quad (80)$$

which follows by multiplying $\partial_{w_a} = \frac{1}{2}(\partial_{x_a} - i\partial_{y_a})$ and $\partial_{\bar{w}_b} = \frac{1}{2}(\partial_{x_b} + i\partial_{y_b})$.

The Poincaré residue has local expression $\Omega = \pm d\xi_1 \wedge d\xi_2 \wedge d\xi_3/Q_\delta$. Define its positive measure and the metric volume in these coordinates by

$$d\nu_\Omega = |Q_\delta|^{-2} d^6 \xi, \quad \frac{\omega_\phi^3}{3!} = \det(g_X) d^6 \xi, \quad (81)$$

where $d^6 \xi = \prod_{a=1}^3 d\text{Re } \xi_a d\text{Im } \xi_a$. Their ratio is

$$\eta(z) = \frac{dV_{g_X}}{d\nu_\Omega} = \det g_X(z) \left| \frac{\partial Q_{\text{aff}}}{\partial w_\delta} \right|^2. \quad (82)$$

Under a change $\xi = \xi(\xi')$ with Jacobian J , both $\det g_X$ and the squared coefficient of Ω acquire $|\det J|^2$, so their ratio is unchanged. The quintic and bicubic have global residue forms by adjunction; consistent projective affine charts describe this same form up to sign. This establishes exact chart independence for the geometric expression, which we also test numerically. For a positive metric on a compact connected Calabi–Yau manifold, the Ricci form is $-i\partial\bar{\partial} \log \eta$. Thus constant η implies Ricci flatness; conversely, Ricci flatness makes the global real function $\log \eta$ pluriharmonic and hence constant by compactness.

D.8 Local D_3 molecular convergence control

For the external control, we parameterize the degenerate in-plane normal mode of equilateral H_3^+ by $q = (q_x, q_y)$. Writing $w = q_x + iq_y$, the six-element D_3 action is generated by $w \mapsto e^{2\pi i/3} w$ and $w \mapsto \bar{w}$. We generated 2,000 single-point RHF/STO-3G energies about the equilateral geometry with bond length 1.65 Bohr and maximum normal-mode displacement 0.18 Bohr. The data-generation script checks that the calculated energy is constant over four independently selected D_3 orbits; the maximum observed spread was 3.997×10^{-15} Hartree.

For every seed, the radial, character and orbit-average arms use the same randomized train/test split, 4,417 trainable parameters, batch size 512 and Adam schedule. Each epoch consists of three optimizer updates, so the reported budgets of 300, 1,000 and 3,000 epochs correspond to 900, 3,000 and 9,000 updates. Energies are standardized using the training-set mean and standard deviation, and held-out MSE is reported after converting predictions back to Hartree. The radial control receives $|w|^2$; the character model receives $(|w|^2, \text{Re}(w^3))$; and the orbit-average model applies one shared MLP to all six D_3 images before averaging. Thus the latter is exactly D_3 invariant, while the radial arm is invariant under the strict supergroup $O(2)$. This experiment measures representation

and finite-budget optimization on a local electronic-structure target. It is not a molecular dynamics, wavefunction-learning or chemical-accuracy benchmark.

For completeness, $|w|^2$ is unchanged under every rotation and reflection, whereas $\text{Re}(w^3)$ is fixed by $w \mapsto e^{2\pi i/3}w$ and $w \mapsto \bar{w}$. For any finite group G , the exact averaged model $\bar{f}(z) = |G|^{-1} \sum_{g \in G} f(gz)$ obeys $\bar{f}(hz) = |G|^{-1} \sum_{g \in G} f(ghz) = \bar{f}(z)$ because $g \mapsto gh$ is a bijection of G . This is the algebraic exactness used by the orbit references; it places no ordering on their learned errors.

D.9 Cost

Let C_E denote the cost of one shared-slot encoder evaluation, C_D the decoder cost, and C_{mon} the cost of forming all block monomials in Eq. (11). For n slots, q character terms, g global terms and latent width L , the encoder costs nC_E , moduli pooling costs $O(nL)$, and the character contractions cost $O(nqL)$. Forming each coordinate product costs $O(n)$ and each pooled source $O(L)$; bounded nonnegative global exponents then give $O(g(n+L))$. Including the decoder, the total forward cost is

$$O(nC_E + nL + C_{\text{mon}} + nqL + g(n+L) + C_D). \quad (83)$$

For fixed integer exponents, $C_{\text{mon}} = O(nBq)$; larger exponents introduce their power-evaluation costs. For an encoder with ℓ_E hidden layers of width W and output width L , $C_E = O(BW + (\ell_E - 1)W^2 + WL)$ when $\ell_E \geq 1$. These terms matter when width, depth, blocks or charges are varied. The construction removes explicit orbit enumeration; it does not establish a fixed cost for all finite groups, because the number of sufficient channels may depend on the action. Direct exact orbit averaging of a base network costs $O(|G|C_{\text{base}})$ including comparable per-action overhead; a sampled K -action average costs $O(KC_{\text{base}})$ and generally requires an empirical action-error check. For the tested Dwork action $|G| = 15,000$, whereas PCP evaluates the supplied coordinate and channel specification. Section A.2 reports its numerical action-exactness check.

Data availability

The source code, configuration files, generated datasets, trained checkpoints, training logs and figure-source data supporting this study are archived at Zenodo [35]: <https://doi.org/10.5281/zenodo.21722111>.

Acknowledgements

Open-access funding and article-processing charges were supported by Project DEAKL and the Technical University of Munich (TUM). Computations were performed using NVIDIA GPUs and Kaggle.

The author used ChatGPT (OpenAI), Claude (Anthropic), and Kimi 2 (Moonshot AI) to assist with Python-code debugging and manuscript language editing. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of this article.

References

- [1] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov, and A. Smola. Deep Sets. In *Advances in Neural Information Processing Systems*, volume 30, 2017.
- [2] O. Puny, M. Atzmon, H. Ben-Hamu, I. Misra, A. Grover, E. J. Smith, and Y. Lipman. Frame averaging for invariant and equivariant network design. In *International Conference on Learning Representations (ICLR)*, 2022.
- [3] S.-O. Kaba, A. K. Mondal, Y. Zhang, Y. Bengio, and S. Ravanbakhsh. Equivariance with learned canonicalization functions. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 15546–15566. PMLR, 2023.
- [4] J.-H. Jacobsen, J. Behrmann, R. S. Zemel, and M. Bethge. Excessive invariance causes adversarial vulnerability. In *International Conference on Learning Representations*, 2019. URL <https://openreview.net/forum?id=BJf2voC5FX>.
- [5] J. Wang, Y. Zhou, Z. Zhang, Q. Chen, Y. Chen, and J. Cheng. Divil: Unveiling and addressing over-invariance for out-of-distribution generalization. *Transactions on Machine Learning Research*, 2025-February:1–23, 2025. URL <https://openreview.net/forum?id=2Zan4ATYsh>.
- [6] G. Benton, M. Finzi, P. Izmailov, and A. G. Wilson. Learning invariances in neural networks. In *Advances in Neural Information Processing Systems*, volume 33, 2020.
- [7] M. Finzi, G. Benton, and A. G. Wilson. Residual pathway priors for soft equivariance constraints. In *Advances in Neural Information Processing Systems*, volume 34, pages 30037–30049, 2021.
- [8] R. Wang, R. Walters, and R. Yu. Approximately equivariant networks for imperfectly symmetric dynamics. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 23078–23091. PMLR, 2022.
- [9] P. J. Olver. *Classical Invariant Theory*, volume 44 of *London Mathematical Society Student Texts*. Cambridge University Press, 1999. ISBN 9780521558211. doi: 10.1017/CBO9780511623660.
- [10] B. Bloem-Reddy and Y. W. Teh. Probabilistic symmetries and invariant neural networks. *Journal of Machine Learning Research*, 21(90):1–61, 2020.
- [11] E. Calabi. On Kähler manifolds with vanishing canonical class. In R. H. Fox, D. C. Spencer, and A. W. Tucker, editors, *Algebraic Geometry and Topology: A Symposium in Honor of S. Lefschetz*, pages 78–89. Princeton University Press, Princeton, 1957.
- [12] S.-T. Yau. On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation, I. *Commun. Pure Appl. Math.*, 31(3):339–411, 1978. doi: 10.1002/cpa.3160310304.
- [13] S. K. Donaldson. Scalar curvature and projective embeddings, I. *J. Differential Geom.*, 59(3): 479–522, 2001.
- [14] S. K. Donaldson. Some numerical results in complex differential geometry. *Pure Appl. Math. Q.*, 5(2):571–618, 2009.
- [15] M. R. Douglas, R. L. Karp, S. Lukic, and R. Reinbacher. Numerical Calabi-Yau metrics. *J. Math. Phys.*, 49:032302, 2008. doi: 10.1063/1.2888403.

- [16] M. Headrick and A. Nassar. Energy functionals for Calabi-Yau metrics. *Adv. Theor. Math. Phys.*, 17(5):867–902, 2013. doi: 10.4310/ATMP.2013.v17.n5.a1.
- [17] A. Ashmore, Y.-H. He, and B. A. Ovrut. Machine learning Calabi-Yau metrics. *Fortsch. Phys.*, 68:2000068, 2020. doi: 10.1002/prop.202000068.
- [18] M. R. Douglas, S. Lakshminarasimhan, and Y. Qi. Numerical Calabi-Yau metrics from holomorphic networks. In *Proceedings of the 2nd Mathematical and Scientific Machine Learning Conference*, volume 145 of *Proceedings of Machine Learning Research*, pages 223–252. PMLR, 2022. URL <https://proceedings.mlr.press/v145/douglas22a.html>.
- [19] V. Jejjala, D. K. Mayorga Peña, and C. Mishra. Neural network approximations for Calabi-Yau metrics. *JHEP*, 2022(08):105, 2022. doi: 10.1007/JHEP08(2022)105.
- [20] Y. Hendi, M. Larfors, and M. Walden. Learning group invariant Calabi-Yau metrics by fundamental domain projections. *Mach. Learn. Sci. Technol.*, 6:015050, 2025. doi: 10.1088/2632-2153/adb4bb.
- [21] A. Rahman. GlobalCY I: a JAX framework for globally defined and symmetry-aware neural Kähler potentials, 2026. arXiv:2604.11404 [hep-th].
- [22] V. Mirjanić and C. Mishra. The sharp edges of Calabi–Yau manifolds: designing symmetric models for Ricci-flat metrics, 2026. arXiv:2606.26892 [hep-th].
- [23] C. H. Ek, O. Kim, and C. Mishra. Calabi-Yau Metrics through Grassmannian Learning and Donaldson’s Algorithm. In *Artificial Intelligence and Mathematics Research*, volume 835 of *Contemporary Mathematics*, pages 89–124. American Mathematical Society, 2026. doi: 10.1090/conm/835/16754.
- [24] D. Y. Eng. Geometric priors for single-patch neural Calabi–Yau metrics on the quintic: a conservative fixed- ψ ablation study. *Mach. Learn.: Sci. Technol.*, 7(3):035049, 2026. doi: 10.1088/2632-2153/ae7494.
- [25] V. Mirjanić and C. Mishra. Symbolic Approximations to Ricci-flat Metrics Via Extrinsic Symmetries of Calabi–Yau Hypersurfaces. *Mach. Learn. Sci. Technol.*, 6(3):035029, 2025. doi: 10.1088/2632-2153/adf68c.
- [26] S.-J. Lee and A. Lukas. Approximate Ricci-flat metrics for Calabi–Yau manifolds. *Fortschr. Phys.*, 74(7):e70129, 2026. doi: 10.1002/prop.70129.
- [27] A. Constantin, S.-J. Lee, A. Lukas, and L. A. Nutricati. Calabi–Yau metrics with full moduli dependence, 2026. arXiv:2606.28487 [hep-th].
- [28] A. Constantin, A. Lukas, and L. A. Nutricati. Calabi–Yau metrics with Kähler moduli dependence, 2026. arXiv:2603.12384 [hep-th].
- [29] D. Y. Eng. Interpretable analytic Calabi–Yau metrics via symbolic distillation, 2026. arXiv:2602.07834 [cs.LG].
- [30] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K.-L. Chan. PySCF: the python-based simulations of chemistry framework. *WIREs Computational Molecular Science*, 8(1):e1340, 2018. doi: 10.1002/wcms.1340.

- [31] B. M. Hill. A simple general approach to inference about the tail of a distribution. *Ann. Stat.*, 3(5):1163–1174, 1975. doi: 10.1214/aos/1176343247.
- [32] E. Veach and L. J. Guibas. Optimally combining sampling techniques for Monte Carlo rendering. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, pages 419–428. ACM Press, 1995. doi: 10.1145/218380.218498.
- [33] B. R. Greene and M. R. Plesser. Duality in Calabi-Yau moduli space. *Nucl. Phys. B*, 338: 15–37, 1990. doi: 10.1016/0550-3213(90)90622-K.
- [34] G. Bini, B. van Geemen, and T. L. Kelly. Mirror quintics, discrete symmetries and shioda maps, 2008. Section 1.1.
- [35] D. Y. Eng. Too Symmetric to See: Diagnosing Over-Invariance in Finite-Symmetry Geometric Learning, 2026. URL <https://doi.org/10.5281/zenodo.21722111>. Zenodo research software, data, and results archive.