WHAT EXPLAINS QUADRATIC FEATURE GAINS IN SEMI-SUPERVISED CLASSIFICATION?

Anonymous authors
Paper under double-blind review

ABSTRACT

Adding quadratic interactions can improve classification with sparse annotations. Does that gain require additional feature directions, or merely a different regularizer? We study this question for spectral graph features using exact controls, a classification theorem, and paired experiments. Projected products are exactly a low-dimensional linear component with an anisotropic penalty. More surprisingly, we construct connected graphs on which full and projected models make identical decisions, although fitting the complementary block separately achieves perfect classification. No positive quadratic reweighting of the fixed linear-feature space recovers that result; the error separation persists as the labeled fraction tends to zero. On three image graphs, a 52-budget study establishes a reference gain. A separate matched decomposition with a common free intercept shows that jointly fitting a separate complementary coefficient block helps on two graphs, while recoupling offsets part of that benefit. A theory-prescribed scalar control leaves further gains in some annotation regimes, while stronger model selection remains competitive. Finally, signed class-margin bounds identify decisions invariant along the entire coupling path. Their numerical coverage rises from 33.02% for a norm bound to 93.17% with signs and 98.20% after refinement. The resulting lesson is operational: separate reweighting, complementary directions, and their coupling before interpreting an enrichment gain or a near-zero ablation gap.

1 INTRODUCTION

A similarity graph lets a small annotated sample inform predictions for many unannotated examples. Spectral coordinates summarize that graph, and their pairwise products are a natural way to enrich a classifier. The practical question is straightforward: do the products improve classification? The scientific question is more demanding: **what does an improvement actually establish about the usefulness of additional feature directions?**

An enriched dictionary changes both its span and the penalty imposed on predictions. These effects are particularly easy to confuse in a model combining a feature predictor with a graph-regularized correction (Rasmussen & Williams, 2006; Li et al., 2019). The correction already permits every finite score field. Features change which fields are inexpensive, not which total fields are representable. Throughout this paper, *linear* and *quadratic* describe the parametric component; the graph correction is retained in both.

Our central finding is that **a near-zero ablation gap can hide useful complementary features**. We construct actual graph-derived classifiers for which Full and Parallel make identical decisions but Decoupled classifies perfectly, beyond every positive quadratic reweighting of the same linear-feature space. In a separate seven-budget CIFAR-10 experiment with a common free intercept, adding the complementary coefficient block improves accuracy by 0.57 point and restoring coupling removes 0.40. This partial cancellation survives equalizing the affine span. The theorem refutes agreement-as-redundancy; the image experiments examine the phenomenon without inheriting its universal obstruction.

The paper proceeds from exact linear-feature controls to matched block interventions and then to decision-level sensitivity. Ten-trajectory reference studies and three-trajectory targeted controls are kept distinct. We ask what a gain or an ablation gap establishes, rather than assume that quadratic features or decoupling always win.

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.

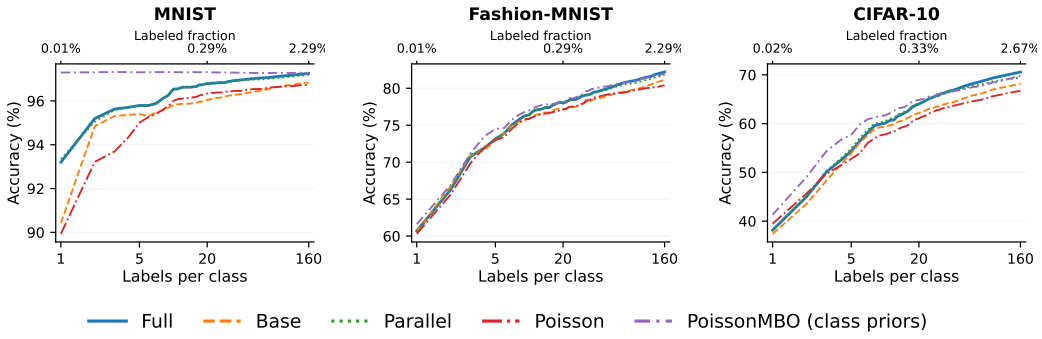


Figure 1: **Classification performance is the starting point.** Native accuracy over all 52 measured annotation budgets, averaged over ten paired trajectories on the same graph. Full and Parallel are often visually close here. Figure 2 separately tests attribution after giving every model a common free intercept. Poisson is a source-based reference; PoissonMBO additionally receives class proportions during fitting. Top axes show labeled fractions. No curve is extended beyond $b = 160$.

2 Related work: enrichment, regularization and sensitivity

Graph-based classification includes harmonic extension (Zhu et al., 2003), Poisson Learning and volume-informed PoissonMBO (Calder et al., 2020), and constrained partitioning such as CutSSL (Holtz et al., 2024). They motivate comparisons on the same graph with different propagation mechanisms and information budgets. Spectral coordinates and diffusion maps are classical (Belkin & Niyogi, 2003; Coifman & Lafon, 2006); no new image encoder is trained here.

Basis functions plus a correlated residual yield an additive covariance (Rasmussen & Williams, 2006, Sec. 2.7). Graph-penalized effects are established (Li et al., 2019; Ioannidis et al., 2018; Jia & Benson, 2020), as are positive-definite polynomial kernels on spheres (Schoenberg, 1942) and graph-kernel enrichment by Schur–Hadamard products (Erb, 2020). LapRLS penalizes the complete predicted function (Belkin et al., 2006); Correct and Smooth trains a predictor and then corrects and smooths its outputs, including with spectral features (Huang et al., 2021). These ingredients are not our novelty.

Harmful sharing appears in multioutput Gaussian processes (Li & Kontar, 2022), and suppressor variables can confound feature attribution (Wilming et al., 2022). Our construction proves a specific classification obstruction: identical quadratic and projected-model decisions coexist with a benefit of the separately parameterized complement that no positive quadratic penalty in the fixed linear-feature space recovers. The coordinates come from a connected graph and the error separation persists at vanishing label density.

For decision sensitivity, Stephenson et al. (2022) seek changes under qualitatively interchangeable GP kernels. We instead bound margins throughout one prescribed coupling path. Projection and inverse-matrix identities underlie the exact controls; the deterministic diffusion limit differs from random high-dimensional kernel linearization (El Karoui, 2010). Our contribution is the classification obstruction, its matched empirical examination, and a signed whole-path diagnostic, not the general observation that kernel choice matters.

3 A common classifier and paired experimental basis

Write $\mathbf{1}_n = (1, \dots, 1)^\top \in \mathbb{R}^n$ and I_n for the all-ones vector and identity matrix; omit dimensions only when clear. Thus $\mathbf{1}_N \mathbf{1}_N^\top$ is the all-ones matrix, not I_N. For vectors, $\|\cdot\|_2$ is Euclidean; for matrices it is spectral, and $\|\cdot\|_F$ is Frobenius. The symbols tr, col and diag denote trace, column space and diagonal formation; $\circ$ and $\odot$ are entrywise products. Matrix inequalities $\succeq 0$ and $\succ 0$ mean positive semidefinite and positive definite, respectively.

Let $W = W^\top \geq 0$ entrywise describe a connected N-vertex graph, with $L = \mathrm{diag}(W\mathbf{1}_N) - W$. An ordered labeled set S has m vertices; $E \in \mathbb{R}^{m \times N}$ restricts to them. For C classes, $Y \in \mathbb{R}^{m \times C}$ contains one-hot rows centered by their observed mean. For any matrix, subscripts select rows and

columns in this order; a colon selects all indices and $U_i = U_{i,:}$. Retain d orthonormal eigenpairs of L, with $0 = \lambda_1 < \lambda_2 \leq \cdots$ and $u_1 = \mathbf{1}_N/\sqrt{N}$. At time $t \geq 0$, define column coordinates

$$z_i = (e^{-t\lambda_1/2}u_1(i), \ldots, e^{-t\lambda_d/2}u_d(i))^\top, \qquad \phi_i = z_i/\left\|z_i\right\|_2, \quad \Phi_{i,:} = \phi_i^\top. \tag{1}$$

The reference G64 uses $(d, t) = (64, t_0)$, $t_0 = 4/\lambda_{10}$; G128 denotes $(128, t_0/2)$. These coordinates use no class labels. Define $G = \Phi\Phi^\top$ and $\Psi_{i,:} = \mathrm{svec}(\phi_i\phi_i^\top)^\top$, where svec stacks upper-triangular entries in a fixed order, weighting off-diagonal entries by $\sqrt{2}$. Then

$$J = \Psi\Psi^\top = G \circ G, \qquad q = d(d+1)/2, \qquad \operatorname{tr} J = \operatorname{tr} G = N. \tag{2}$$

BASE uses $Z = \Phi$; FULL uses $Z = [\Phi, \Psi]$. Both solve

$$U = Z\Theta + V + \mathbf{1}_N c^\top, \qquad \min_{\Theta,V,c} \frac{1}{2}\left\|EU - Y\right\|_F^2 + \frac{1}{2}\operatorname{tr}(\Theta^\top A\Theta) + \frac{1}{2}\operatorname{tr}(V^\top HV), \tag{3}$$

Here $U, V \in \mathbb{R}^{N\times C}$, $c \in \mathbb{R}^C$, and Θ has one row per column of Z. The correction uses $H = \tau L + \sigma I_N \succ 0$. Feature penalties are $A = \alpha_\ell I_d$ for Base and $A = \mathrm{diag}(\alpha_\ell I_d, \alpha_q I_q)$ for Full, with $\alpha_\ell, \alpha_q > 0$. The intercept c is zero or free as specified. For $m > 0$ the minimizer is unique; with $c = 0$,

$$K = H^{-1} + ZA^{-1}Z^\top, \qquad U = K_{:,S}(I_m + K_{SS})^{-1}Y. \tag{4}$$

Thus K^{-1} is the effective penalty on total scores. All variants retain the same correction when used as attribution controls. Computation uses sparse graph actions and support-sized solves, not a dense $N \times N$ kernel.

Data and comparison units. We use GraphLearning's frozen neighbor arrays from Calder et al. (2020): VAE for MNIST and Fashion-MNIST, and AET for CIFAR-10. The AET method is due to Zhang et al. (2019). The graphs contain 70,000, 70,000 and 60,000 examples. Accuracy is evaluated on the unannotated complement of the entire graph, not only the official test partition. With ten classes and b labels per class, $m = 10b$; even $b = 160$ labels at most 2.67% of a graph. Encoders and graphs remain fixed.

The reference panel has ten class-stratified, nested annotation trajectories over $b \in \{1, \ldots, 40\} \cup \{50, 60, \ldots, 160\}$. Its penalties are $(\alpha_\ell, \alpha_q, \tau, \sigma) = (.03, .03, 1, .01)$ and $c = 0$. The ordered attribution in Figure 2 instead uses all four common-free-intercept models on ten trajectories at $b \in \{1, 5, 10, 20, 40, 80, 160\}$, with these same fixed penalties and geometry. The exact-match and stronger-selection panels use three paired trajectories at $b = 5, 40, 160$. Attribution contrasts hold geometry, support, graph correction and intercept fixed; independently selected procedures answer a different question. Dense summaries integrate each paired trajectory over $\log b$, normalized by $\log 160$; compact summaries average their specified budgets equally. Ten-trajectory intervals are marginal paired bootstraps; three-trajectory displays show observed values and ranges. Argmax is primary, oracle counts are separate, and PoissonMBO receives population proportions during fitting. Appendix B gives all settings and effect sizes. These graphs participated in development; our statistical scope is conditional on them.

4 WHAT EXPLAINS THE CLASSIFICATION GAIN?

4.1 ESTABLISH THE GAIN, THEN TEST THE EXPLANATION

Figure 1 retains the lowest annotation counts and the propagation references. Full exceeds Base in mean accuracy at 52/52, 50/52 and 52/52 budgets on MNIST, Fashion-MNIST and CIFAR-10. The integrated gains are +0.68, +0.53 and +1.61 points (Table 5). They survive a common free intercept, a doubled-trace linear covariance and the tested penalty-selection controls. For example, common-intercept gains on the seven-budget panel are +0.82, +0.66 and +1.52 points. These comparisons establish a useful enrichment recipe, not yet the need for new parametric directions.

Full improves on Poisson Learning over the dense integrated criterion on all three graphs, but PoissonMBO remains stronger on MNIST and CIFAR-10, particularly at very low label counts. Complete C&S is also competitive: its CIFAR-10 mean is close to selected Full on the separate seven-budget protocol, and it leads at that protocol's endpoint. Appendix C.8 retains the absolute accuracies, paired comparisons and both decoding protocols.

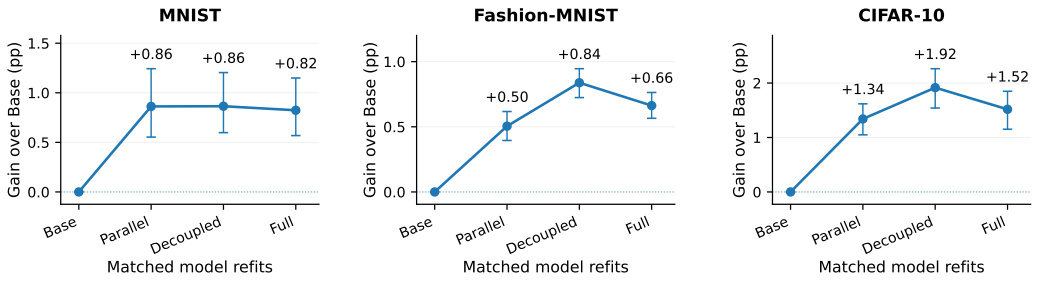


Figure 2: **Block addition and recoupling after controlling the constant.** Every model has the same free intercept, G64 geometry, fixed penalties and annotated support. The vertical axis is gain over Base, averaged equally over the seven compact budgets within each of ten seeds; whiskers are paired 95% bootstrap intervals against Base. The horizontal axis joins *model interventions*, not annotation budgets. Decoupled rises above Parallel on Fashion-MNIST and CIFAR-10, then Full gives back part of that gain. Increment-wise intervals are in Table 3.

4.2 Projected products reveal a linear-feature explanation

Let P be the orthogonal projector onto $\mathrm{col}([\mathbf{1}_N, \Phi])$ and $Q = I_N - P$, and set $\Psi_{\|} = P\Psi$, $\Psi_{\perp} = Q\Psi$. Parallel uses $[\Phi, \Psi_{\|}]$. Because $\sum_j \phi_{ij}^2 = 1$, the quadratic dictionary contains the constant, even when the normalized linear dictionary does not. With $c = 0$, Parallel may therefore add a constant direction missing from Base: its dense contrast is *not* a pure within-span penalty intervention. For attribution we give both models the same free intercept. Then

$$\mathrm{col}([\mathbf{1}, \Phi]) = \mathrm{col}([\mathbf{1}, \Phi, P\Psi]),$$

so the first contrast changes regularization within the same affine parametric span. The free intercept also absorbs the isotropic quadratic coefficient (Appendix A.2).

Proposition 1 (Exact linear-feature control)**.** *Let $Q_0 \in \mathbb{R}^{N \times r_0}$ have orthonormal columns spanning $\mathrm{col}([\mathbf{1}, \Phi])$, $B = Q_0^\top \Phi$ and $B_q = Q_0^\top \Psi$. For $M = BB^\top/\alpha_\ell + B_q B_q^\top/\alpha_q \succ 0$,* Parallel *is exactly Eq.* (3) *with parametric component $Q_0\beta$, $\beta \in \mathbb{R}^{r_0 \times C}$, and penalty $\mathrm{tr}(\beta^\top M^{-1}\beta)/2$, retaining the same intercept convention. If $M = C_M C_M^\top$, $Q_0 C_M$ implements it with unit coefficient penalty and $r_0 \le d + 1$ columns.*

At the reference rank, 2,080 projected products yield at most 65 coordinates with an explicit anisotropic penalty; scalar trace matching generally cannot reproduce it. With the common free intercept, Figure 2 tests this reweighting without enlarging the affine span. The zero-intercept reference is not given that interpretation.

4.3 A near-zero net effect can conceal useful complementary features

Decoupled fits $[\Phi, \Psi_{\|}, \Psi_{\perp}]$ with separate coefficients and equal penalties on the two quadratic blocks. Full uses their sum with common coefficients. On $T = S^c$, predict $\widehat{y}_i = \arg\max_k U_{ik}$ with a fixed label-independent tie order. Define percentage accuracy by $\mathcal{A}_T = 100|\{i \in T : \widehat{y}_i = y_i\}|/|T|$, where y_i is the true class; write $\mathcal{A}$ when T is common. With a *common free intercept, geometry and penalties*,

$$\begin{aligned}\mathcal{A}(\textsc{Full}) - \mathcal{A}(\textsc{Base}) = &\underbrace{\mathcal{A}(\textsc{Parallel}) - \mathcal{A}(\textsc{Base})}_{\text{same-affine-span reweighting}} + \underbrace{\mathcal{A}(\textsc{Decoupled}) - \mathcal{A}(\textsc{Parallel})}_{\text{add an independent complementary block}} \\ &+ \underbrace{\mathcal{A}(\textsc{Full}) - \mathcal{A}(\textsc{Decoupled})}_{\text{restore coupling}}.\end{aligned} \tag{5}$$

This is an exact accounting of three matched interventions, not an additive decomposition of accuracy by feature column. Each model is refitted, and the order specifies the comparison. It avoids interpreting a small Full–Parallel gap as evidence that the complementary block cannot be useful.

On the common-intercept CIFAR-10 panel, the three increments are $+1.341$, $+0.575$ and -0.398 points, giving a $+1.517$ Full–Base gain. On Fashion-MNIST they are $+0.505$, $+0.334$ and -0.176. In both graphs, the complementary-block benefit and recoupling loss occur on every seed-wise trajectory and their paired intervals exclude zero. Thus the opposing effects survive removal of the constant-span difference at the fixed recipe. On MNIST the complementary increment is only $+0.002$ point, with an interval spanning zero; the finding is not universal.

Appendix C retains the zero-intercept, 52-budget decomposition, whose first step can add a constant. Its CIFAR-10 Full–Parallel contrast changes sign with budget despite a small integrated mean. Those log-budget summaries are not subtracted from the seven-budget control: *a small average gap establishes neither redundancy at every budget nor robustness to tuning.*

4.4 A CLASSIFICATION THEOREM: AGREEMENT DOES NOT IMPLY REDUNDANCY

The preceding contrast could be dismissed as an accidental cancellation of averages. The next theorem shows a stronger obstruction, even when Full and Parallel agree *at every vertex*. Write $\mathrm{err}_T = 1 - \mathcal{A}_T/100$ for the error fraction on T.

Theorem 1 (Useful complementary features hidden by coupling)**.** *For every integer $r \geq 1$, there is a connected positive-weight graph with $128r$ vertices, two normalized diffusion coordinates, balanced binary classes and one observed label per class, for which*

$$\widehat{y}_{\textsc{Full}} = \widehat{y}_{\textsc{Parallel}} \text{ on all vertices},$$
$$\mathrm{err}_T(\textsc{Full}) = \mathrm{err}_T(\textsc{Parallel}) = \frac{2r}{128r-2}, \qquad \mathrm{err}_T(\textsc{Decoupled}) = 0. \tag{6}$$

The three models share positive feature penalties, H, the support and a free intercept; Full and Decoupled have the same feature span and weighted covariance trace. For every $F \succeq 0$ *with* $\mathrm{col}(F) \subseteq \mathrm{col}[\mathbf{1}, \Phi]$*, the free-intercept estimator with kernel $H^{-1} + F$ has error at least $r/(128r-2)$. The three displayed classifications persist under sufficiently small graph and parameter perturbations. Under independent uniform sampling of one label per class, the stated separation holds with probability at least $(63/64)^2$ on this constructed family.*

Why it happens. In the construction, $c, p \in \mathbb{R}^N$ are coordinate columns, not intercepts or class indices. With $\bar{c} = p^\top (c \odot p)/\|p\|_2^2$, the only odd product projects as

$$P(\sqrt{2}\, c \odot p) = \sqrt{2}\, \bar{c}\, p, \qquad Q(\sqrt{2}\, c \odot p) = \sqrt{2}(c - \bar{c}\mathbf{1}) \odot p.$$

Across the two types, the latter changes sign. Independent coefficients preserve its useful negative transmission, whereas raw-product coupling cancels it. A positive-semidefinite reweighting of the linear-feature space cannot reverse the relevant contrast order (Lemma 1). Appendix A.3 gives the graph, exact rational sign bounds and the full proof. The all-penalty obstruction is proved at the exact construction; local robustness concerns the three fixed models.

The error gap tends to $1/64$ while the labeled fraction tends to zero. Thus the obstruction is not confined to vanishingly few mistakes on a growing graph. Its weights, penalties and labels are constructed, not fitted to an image benchmark. The theorem refutes agreement-as-redundancy; the paired image effects above establish a corresponding cancellation in the tested recipes, not the impossibility of all other image classifiers.

5 WHAT SURVIVES SCALAR MATCHING AND STRONGER MODEL SELECTION?

5.1 A THEORY-PRESCRIBED SCALAR COMPARATOR

Set $a = \alpha_\ell^{-1}, \omega = \alpha_q^{-1}$. Unit row norms give the exact identity

$$J = 2G - \mathbf{1}_N \mathbf{1}_N^\top + R, \qquad R_{ij} = (1 - G_{ij})^2 = \tfrac{1}{4}\|\phi_i - \phi_j\|_2^4. \tag{7}$$

A free intercept absorbs the constant covariance. The natural scalar comparison is therefore between $K_F = H^{-1} + aG + \omega J$ and $K_L = H^{-1} + (a + 2\omega)G$, fixing H, geometry and annotations. Its penalty is

$$\alpha_{\mathrm{match}} = (\alpha_\ell^{-1} + 2\alpha_q^{-1})^{-1}. \tag{8}$$

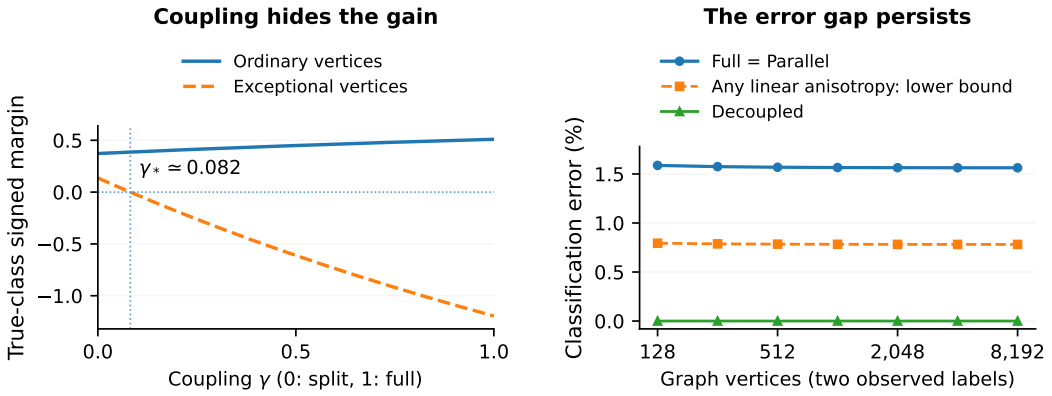


Figure 3: **The construction makes masking exact and persistent.** These are closed-form predictions of Theorem 1, not image-benchmark observations. Left: at $r = 1$, the exceptional true-class margin changes sign at $\gamma_* \approx 0.08186$ as coupling is restored; ordinary margins stay positive. Right: two labels suffice for a nonvanishing error separation as graph size grows. The dashed curve is a proved lower bound for every admissible linear-feature anisotropy, not the measured error of a selected model.

At the reference penalties this is 0.01, not the 0.015 of the doubled-trace control. It tests a specific scalar explanation, distinct from the anisotropic Parallel control and from selecting a best linear procedure.

For $m \geq 2$, let $T_S \in \mathbb{R}^{m\times(m-1)}$ have orthonormal columns spanning $\mathbf{1}_m^\perp = \{v : \mathbf{1}_m^\top v = 0\}$, and define

$$A_K = I_{m-1} + T_S^\top K_{SS} T_S, \qquad \mathcal{C}_i(K) = (K_{iS} - \mathbf{1}_m^\top K_{SS}/m) T_S. \tag{9}$$

With $\bar{Y} = \mathbf{1}_m^\top Y/m \in \mathbb{R}^{1\times C}$, the free-intercept fit is $U_i = \bar{Y} + \mathcal{C}_i(K) A_K^{-1} T_S^\top Y$.

Theorem 2 (Exact score remainder beyond scalar matching)**.** *Write U_F, U_L for the fits under K_F, K_L. With the compressed block $R_S = T_S^\top R_{SS} T_S$ and $D_i = \mathcal{C}_i(K_L) A_{K_L}^{-1}$,*

$$(U_F - U_L)_i = \omega[\mathcal{C}_i(R) - D_i R_S] A_{K_F}^{-1} T_S^\top Y. \tag{10}$$

For fixed graph, H, support, penalties and $d \geq 2$, the diffusion coordinates in Eq. (1) *give $U_F - U_L = O(e^{-2t\lambda_2})$ as $t \to \infty$.*

The finite identity determines what the paired experiment measures. The deterministic leading coefficient and the exact condition for universal scalar recovery are proved in Appendix A.4. The remainder is not assumed positive semidefinite and is not the projected block $Q\Psi$. It measures departure from the prescribed scalar covariance, not irreducibility to every anisotropic linear-feature penalty.

The prescribed experiment. The 54 comparisons cross three graphs, two geometries and three budgets on three common seeds. At G64, the mean Full–matched differences over $b = 5, 40, 160$ are $+0.232$, $+0.445$ and $+0.633$ points. At $b = 160$, all eighteen graph/geometry/seed differences are positive; the CIFAR-10 means are $+1.324$ and $+1.047$ points at G64 and G128. Nevertheless, Fashion-MNIST/G128/$b = 40$ has three negative differences. The scalar adjustment is informative but does not recover the entire empirical effect.

Equation (10) agrees with the independently fitted score differences to a maximum recorded absolute discrepancy of 2.58×10^{-8}. At CIFAR-10/$b = 160$, about 6.43% and 8.37% of decisions differ, substantially more than the net accuracy gains. The coarse uniform bound in Appendix A.4 is conservative at the tested times (median bound/change ratio 3.52×10^9); the finite response, not the asymptotic approximation, is the quantitative link used here.

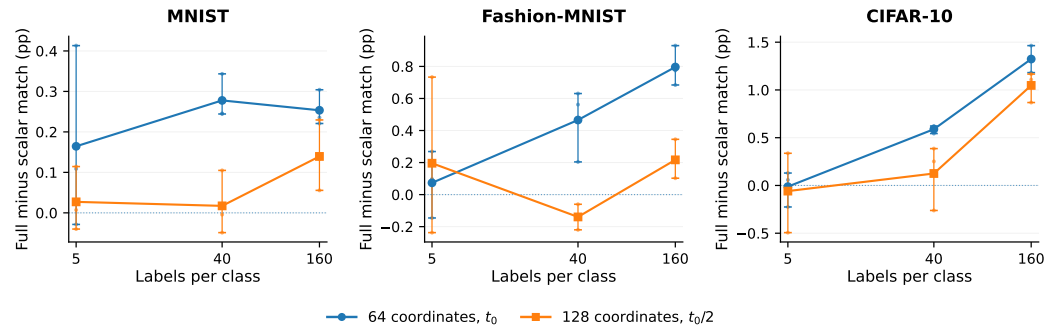


Figure 4: **The exact scalar adjustment leaves a measurable classification benefit.** Full minus the comparator in Eq. (8), with identical geometry, H, support and free intercept. Every seed-wise difference is displayed; markers and whiskers show three-seed means and observed ranges, not confidence intervals. Both geometries and all tested budgets are retained.

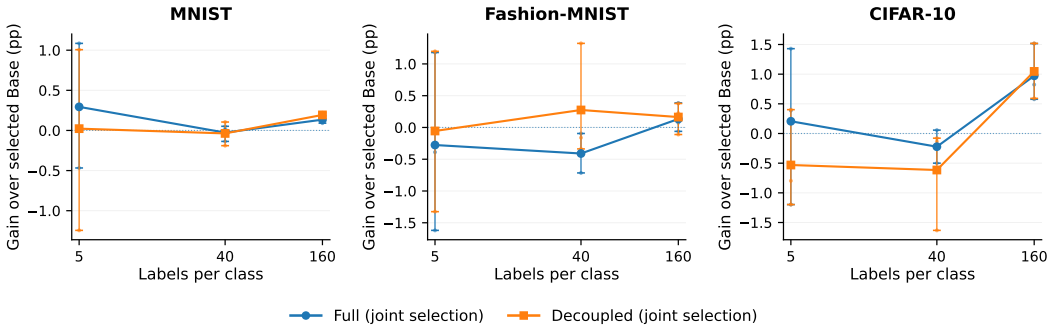


Figure 5: **A stronger linear procedure changes the practical comparison.** Independently selected Full and Decoupled minus selected Base, with three complete paired trajectories. All seed-wise effects and every tested budget are shown; whiskers are observed ranges. These are not the fixed-recipe effects in Figure 4. The absolute accuracies and Parallel/LapRLS results remain in Appendix C.

5.2 Independent selection and a shared-setting control

The joint-selection panel gives Base, Parallel, Full and Decoupled a common free intercept and access to G64/t_0 and G128/$(t_0/2)$. Parameters are selected only on folds within the observed labels. The linked grid is augmented with separate linear/quadratic penalties, giving 54 Base candidates and 90 candidates per enriched procedure. These independently selected procedures do not constitute a controlled coupling intervention; we evaluate that intervention separately at a shared setting below.

Mean Full–Base effects become $+0.13$, -0.18 and $+0.32$ points over the three budgets, with signs varying across trajectories. At $b = 160$, Full and Decoupled still exceed Base on CIFAR-10 by 0.97 and 1.05 points on average, with each of the three paired differences positive. At $b = 40$, Full is below Base in every graph mean. The ten-seed factorial geometry study independently shows why representation matters: at G128/$(t_0/2)$ the mean Full–Base gap is near zero, but the CIFAR-10 effect changes from -1.18 at $b = 5$ to $+1.25$ at $b = 160$. No single endpoint or integrated mean captures this interaction.

The intervention after shared selection. For each support, Base selects one geometry and linked penalty setting by the same observed-label validation. We then refit all four models at that setting, with a common free intercept. On CIFAR-10, adding the complement gains $+0.491$ point and recoupling loses 0.231 point; both signs hold on each of the three budget-averaged trajectories. This partial cancellation therefore survives this shared tuning rule. It is not universal: the mean complementary increment is $+0.020$ on MNIST and -0.170 on Fashion-MNIST. Table 11 retains all increments and

observed seed ranges. These three-seed results concern settings selected for Base, not an optimum shared by every model.

The result extends beyond residual regularization. LapRLS penalizes the whole predicted function. Expanding its ambient grid at fixed G64 improves CIFAR-10 averages by 4.73 points for linear features and 4.31 for quadratic features. Quadratic LapRLS then has the largest three-budget mean among the six complementary procedures on MNIST and Fashion-MNIST; Full has the largest on CIFAR-10. More directly, quadratic-minus-linear LapRLS is $+0.47$, $+1.18$ and $+0.76$ point, positive on every complete trajectory. Product benefits can thus occur in another regularization arrangement; they are not evidence for the universal superiority of joint residual fitting. Geometry choices and finite grids remain part of each procedure, and the smallest expanded LapRLS penalty is still selected in 27/54 decisions.

6 Which decisions actually depend on coupling?

Theorem 1 shows that recoupling can reverse useful decisions. On a real graph we need to locate, rather than assume, such sensitivity: *would this class assignment change if the same two blocks were coupled differently?* Keep geometry, penalties and annotations fixed and define

$$J_0 = PJP + QJQ, \quad D = (PJQ + QJP)/\alpha_q, \quad K_\gamma = H^{-1} + G/\alpha_\ell + J_0/\alpha_q + \gamma D. \tag{11}$$

Write U_γ for the fit under K_γ: $\gamma = 0$ gives Decoupled and $\gamma = 1$ Full. For $|\gamma| \le 1$, the feature covariance remains positive semidefinite, with unchanged trace and combined predictor span. Its cross term has rank at most $2(d+1)$ and a symmetric positive/negative spectrum. In the no-intercept formulation, the effective-precision difference has the same balanced inertia (Appendix A.5). This is a redistribution of regularization, not a uniform relaxation.

For class indices $c \neq h$, define $f_{ich}(\gamma) = U_{\gamma,ic} - U_{\gamma,ih}$. Normalizing the support-restricted response gives eigenvalues $\mu_j \in (-1, 1)$ and computable signed coefficients a_{ichj}, defined for both intercept conventions in Appendix A.6.

Theorem 3 (Whole-path class-margin diagnostic)**.** *The margin is exactly*

$$f_{ich}(\gamma) = f_{ich}(0) + \sum_j a_{ichj} \frac{\gamma}{1 + \mu_j \gamma}. \tag{12}$$

On $[u, v] \subseteq [0, 1]$ *it is bounded below by*

$$L_{ich}[u, v] = f_{ich}(0) + \sum_{a_{ichj} \ge 0} a_{ichj} \frac{u}{1 + \mu_j u} + \sum_{a_{ichj} < 0} a_{ichj} \frac{v}{1 + \mu_j v}. \tag{13}$$

If these bounds are positive against every rival on every interval of a partition, the initial winning class is unchanged on the entire path. Every path with strictly positive winning margins is certified by sufficiently fine finite refinement in exact arithmetic.

Unlike a norm bound, this test retains cancellations between signed contributions and compares the actual competing classes. The support and query factors are prepared once; varying γ requires no new graph solve. A witness of a negative margin proves sensitivity at that value; absence of a witness on a finite grid does not establish continuous-path stability.

Measured utility. Across 54 configurations and 221,184 query/configuration evaluations, the norm test covers 33.02%, the signed test 93.17%, and refinement raises coverage to 98.20%. Signs contribute 60.15 percentage points; subdivision adds another 5.03, or 11,122 diagnoses. At $b = 160$ with an intercept, coverage rises from 0.008% to 97.87% on Fashion-MNIST and from 0.42% to 97.56% on CIFAR-10. The remaining evaluations contain 3,983 recorded change witnesses and one unresolved case, not additional stable predictions.

Which error changes produce the recoupling loss? Write $\widehat{y}_i^D$ and $\widehat{y}_i^F$ for the two endpoint predictions on $T = S^c$. Let n_{repair} count wrong-to-correct changes and n_{damage} correct-to-wrong changes. Then

$$\mathcal{A}_T(\textsc{Full}) - \mathcal{A}_T(\textsc{Decoupled}) = \frac{100}{|T|}(n_{\text{repair}} - n_{\text{damage}}). \tag{14}$$

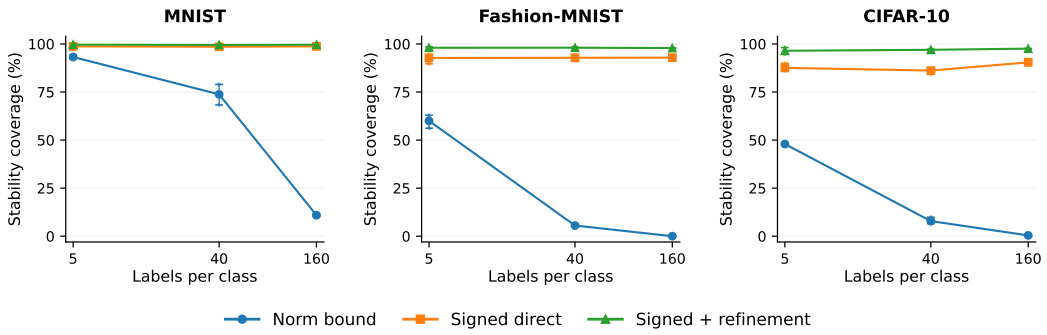


Figure 6: **Signs recover stability information; refinement adds a further gain.** Numerical whole-path coverage on the same 4,096 query vertices per support, with a free intercept. Three-seed means and observed ranges are shown. Coverage is a stability diagnostic, not classification accuracy. The norm, signed-direct and adaptive tests share the same path and query panel.

The recorded G64 full-complement diagnostic, with a common free intercept, two seeds and $b = 5, 40, 160$, supplies these counts. On CIFAR-10, recoupling repairs 0.616% of evaluated decisions but damages 1.041%, giving -0.425 point; Fashion-MNIST gives 0.621%, 0.945% and -0.324 point. These are within-support rates, averaged over budgets and seeds (Appendix C.4). Whole-path stable and change-and-return decisions contribute zero net endpoint error. The separate 4,096-query export lacks truth labels and final winners; its sensitivity flags cannot yet be joined to correction/harm labels. Coverage and these full-complement error counts are therefore reported separately.

What an analyst learns that endpoint accuracy cannot show. For CIFAR-10 query 38063, class 7 wins at both endpoints, with margins 0.008113 and 0.005007. At $\gamma = 0.390625$, class 4 overtakes it and the initial-class margin is -0.000164. Thus endpoint agreement would incorrectly suggest coupling-insensitivity. This is one of 31 recorded change-and-return cases (Figure 8 in the appendix). The diagnostic lets an analyst separate predictions invariant to this intervention from those requiring a coupling-sensitivity review. It neither identifies the correct class nor chooses the best coupling on evaluation labels.

Shared Green actions take 2,317.36 seconds, followed by a 105.83-second diagnostic phase with some writes included. These are not cold end-to-end costs. All direct signed diagnoses survive a stored cutoff of 10^{-6}; refinement-only minorants are unavailable for this sensitivity check. The tests are numerical, not outward-rounded finite-precision guarantees (Appendices B.6–B.7).

7 Discussion and Conclusion

What can an enrichment gain, or an ablation gap, establish? Exact compression shows that projected products can improve classification through anisotropic regularization without complementary parametric directions. The classification theorem establishes the converse warning: even identical Full and Parallel decisions do not imply redundancy, because coupling can hide a nonvanishing benefit that no positive quadratic reweighting of the same linear-feature space recovers in the constructed task. The common-intercept image controls exhibit opposing block-addition and recoupling effects at fixed settings. Recorded endpoint transitions show that recoupling loses more correct decisions than it repairs on two diagnostic graphs; independently selected procedures answer a different question.

The practical sequence is therefore to equalize the affine span through a common intercept, separate reweighting from independent block addition, and only then measure coupling at common settings. The exact scalar match answers a narrower attribution question than arbitrary anisotropy or independently selected procedures. Signed margin paths then identify the decisions sensitive to the specified intervention, including changes missed by endpoint agreement. Together, the proof and experiments turn a potentially misleading ablation into explicit, testable alternatives. The constructive theorem is an existence and non-identifiability result; the image conclusions remain conditional on the observed graphs, finite grids and annotation trajectories.

REPRODUCIBILITY STATEMENT

The supplementary material contains `code/` and `result/`: model implementations, configurations, per-seed observations, query diagnostics and witnesses. The README gives commands to install the dependencies, rerun the experiments from public graph inputs, and generate the tables and figures. Supplied observations can also be used for reconstruction without refitting. The command `python3 code/run.py check` runs the finite mathematical checks described below. Full large score fields and Green factors are not included. Appendix B specifies the paired studies and their different sample sizes.

AI USE STATEMENT

AI tools were used to assist with manuscript preparation by suggesting alternative wording, correcting grammar, and improving the clarity and readability of technical explanations. They also supported literature retrieval and discovery by helping identify potentially relevant publications and related work. The authors retain responsibility for the final text, the selection and interpretation of cited references, and all scientific claims.

ETHICS STATEMENT

The study uses existing image benchmarks, fixed representations and explicitly constructed graphs for mathematical witnesses, without collecting new personal data. The transductive evaluation and population information supplied to particular methods are explicit. Stability diagnostics do not establish prediction correctness, fairness or deployment safety.

REFERENCES

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003. doi: 10.1162/089976603321780317.

Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, 7: 2399–2434, 2006. URL https://jmlr.org/papers/v7/belkin06a.html.

Rajendra Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer, 1997. doi: 10.1007/978-1-4612-0653-8.

Jeff Calder, Brendan Cook, Matthew Thorpe, and Dejan Slepčev. Poisson learning: Graph based semi-supervised learning at very low label rates. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 1306–1316, 2020. URL https://proceedings.mlr.press/v119/calder20a.html.

Ronald R. Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006. doi: 10.1016/j.acha.2006.04.006.

Noureddine El Karoui. The spectrum of kernel random matrices. *The Annals of Statistics*, 38(1): 1–50, 2010. doi: 10.1214/08-AOS648. URL https://arxiv.org/abs/1001.0492.

Wolfgang Erb. Semi-supervised learning on graphs with feature-augmented graph basis functions. arXiv preprint arXiv:2003.07646, 2020. URL https://arxiv.org/abs/2003.07646.

Chester Holtz, Pengwen Chen, Zhengchao Wan, Chung-Kuan Cheng, and Gal Mishne. Continuous partitioning for graph-based semi-supervised learning. In *Advances in Neural Information Processing Systems*, volume 37, pp. 5615–5645, 2024. doi: 10.52202/079017-0182. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/0a4e065fb5ee13441c90bb4b4d6072d0-Abstract-Conference.html.

Qian Huang, Horace He, Abhay Singh, Ser-Nam Lim, and Austin R. Benson. Combining label propagation and simple models out-performs graph neural networks. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=8E1-f3VhX1o.

Vassilis N. Ioannidis, Meng Ma, Athanasios N. Nikolakopoulos, Georgios B. Giannakis, and Daniel Romero. Kernel-based inference of functions over graphs. In *Adaptive Learning Methods for Nonlinear System Modeling*, pp. 173–198. Elsevier, 2018. doi: 10.1016/B978-0-12-812976-0.00010-5. URL https://arxiv.org/abs/1711.10353.

Junteng Jia and Austin R. Benson. Residual correlation in graph neural network regression. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 588–598, 2020. doi: 10.1145/3394486.3403101. URL https://arxiv.org/abs/2002.08274.

Moyan Li and Raed Kontar. On negative transfer and structure of latent functions in multioutput Gaussian processes. *SIAM/ASA Journal on Uncertainty Quantification*, 10(4):1714–1732, 2022. doi: 10.1137/21M1436816. URL https://arxiv.org/abs/2004.02382.

Tianxi Li, Elizaveta Levina, and Ji Zhu. Prediction models for network-linked data. *The Annals of Applied Statistics*, 13(1):132–164, 2019. doi: 10.1214/18-AOAS1205. URL https://arxiv.org/abs/1602.01192.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006. URL https://gaussianprocess.org/gpml/.

Isaac J. Schoenberg. Positive definite functions on spheres. *Duke Mathematical Journal*, 9(1):96–108, 1942. doi: 10.1215/S0012-7094-42-00908-6.

William T. Stephenson, Soumya Ghosh, Tin D. Nguyen, Mikhail Yurochkin, Sameer Deshpande, and Tamara Broderick. Measuring the robustness of gaussian processes to kernel choice. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 3308–3331. PMLR, 2022. URL https://proceedings.mlr.press/v151/stephenson22a.html.

Rick Wilming, Céline Budding, Klaus-Robert Müller, and Stefan Haufe. Scrutinizing XAI using linear ground-truth data with suppressor variables. *Machine Learning*, 111(5):1903–1923, 2022. doi: 10.1007/s10994-022-06167-y. URL https://link.springer.com/article/10.1007/s10994-022-06167-y.

Liheng Zhang, Guo-Jun Qi, Liqiang Wang, and Jiebo Luo. AET vs. AED: Unsupervised representation learning by auto-encoding transformations rather than data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2547–2555, 2019. URL https://openaccess.thecvf.com/content_CVPR_2019/html/Zhang_AET_vs._AED_Unsupervised_Representation_Learning_by_Auto-Encoding_Transformations_Rather_CVPR_2019_paper.html.

Xiaojin Zhu, Zoubin Ghahramani, and John Lafferty. Semi-supervised learning using Gaussian fields and harmonic functions. In *Proceedings of the 20th International Conference on Machine Learning*, 2003. URL https://pages.cs.wisc.edu/~jerryzhu/pub/zgl.pdf.

A Proofs and computational implications

A.1 The additive estimator and its free-intercept form

For perturbations $(\Delta\Theta, \Delta V, \Delta c)$, the quadratic part of Eq. (3) is

$$\left\|E(Z\Delta\Theta + \Delta V + \mathbf{1}\Delta c^\top)\right\|_F^2 + \operatorname{tr}(\Delta\Theta^\top A\Delta\Theta) + \operatorname{tr}(\Delta V^\top H\Delta V).$$

The positive feature and correction penalties force $\Delta\Theta = \Delta V = 0$ if their terms vanish; $m > 0$ then forces $\Delta c = 0$. Thus the objective is strictly convex and coercive. Fixing $c = 0$ simply removes that variable.

Let $R_Y = Y - EU \in \mathbb{R}^{m\times C}$. The first-order equations give $A\Theta = Z^\top E^\top R_Y$ and $HV = E^\top R_Y$, hence $U = K_{:,S}R_Y + \mathbf{1}c^\top$. With $c = 0$, $(I + K_{SS})R_Y = Y$. With a free intercept, also $\mathbf{1}_m^\top R_Y = 0$, so $R_Y = T_S B_Y$, $B_Y \in \mathbb{R}^{(m-1)\times C}$, and

$$(I + T_S^\top K_{SS} T_S)B_Y = T_S^\top Y, \qquad c^\top = \bar{Y} - \mathbf{1}_m^\top K_{SS} T_S B_Y/m.$$

For $m = 1$, the free-intercept solution is $U = \mathbf{1}_N Y_{1,:}$, $\Theta = V = 0$; the contrast space is empty. For $m \geq 2$, substitution proves Eq. (9). This also shows that adding any constant covariance $\kappa \mathbf{1}_N \mathbf{1}_N^\top$ leaves a free-intercept fit unchanged whenever the estimator is well defined. The matrix on the contrast space is positive definite with minimum eigenvalue at least one when $K \succeq 0$.

At fixed total field U without intercept, minimizing over feature coefficients gives the effective penalty

$$H_{\text{eff}} = H - HZ(A + Z^\top HZ)^{-1} Z^\top H = (H^{-1} + ZA^{-1}Z^\top)^{-1}.$$

Completing the square and applying the matrix inversion lemma gives this precision form of the additive covariance (Rasmussen & Williams, 2006, Sec. 2.7 and App. A.3). The precision need not be a graph Laplacian; no corrected adjacency or random-walk interpretation is required.

A.2 Constant removal and exact projected compression

Write a symmetric quadratic coefficient $B_c \in \mathbb{R}^{d \times d}$ as $B_c = B_c^\circ + s_c I_d$, where $\operatorname{tr} B_c^\circ = 0$. Since $\|\phi_i\|_2 = 1$,

$$\phi_i^\top B_c \phi_i = \phi_i^\top B_c^\circ \phi_i + s_c, \qquad \|B_c\|_F^2 = \|B_c^\circ\|_F^2 + d s_c^2.$$

Moving s_c to a free intercept leaves scores unchanged and reduces the penalty unless $s_c = 0$. Using $\langle A, B \rangle_F = \operatorname{tr}(A^\top B)$, the traceless feature kernel is

$$\langle \phi_i \phi_i^\top - I_d/d, \phi_j \phi_j^\top - I_d/d \rangle_F = G_{ij}^2 - 1/d,$$

a positive-semidefinite Gram matrix with trace $N(1 - 1/d)$. This explicit Gram calculation is consistent with the classical theory of positive-definite functions on spheres (Schoenberg, 1942); row normalization is essential.

For Proposition 1, let $F = [B/\sqrt{\alpha_\ell}, B_q/\sqrt{\alpha_q}]$. Because $\mathbf{1} \in \operatorname{col}(\Psi)$, the columns of $[\Phi, P\Psi]$ span $\operatorname{col}([\mathbf{1}, \Phi])$. Consequently F has full row rank and $FF^\top = M \succ 0$, even if Φ itself is rank deficient. The minimum-norm solution of $Fv = \beta$ is $v_* = F^\top M^{-1} \beta$, since any other solution differs by a vector in $\ker F$ orthogonal to v_*. Thus

$$\min_{B\beta_\ell + B_q \beta_q = \beta} \{\alpha_\ell \|\beta_\ell\|_F^2 + \alpha_q \|\beta_q\|_F^2\} = \operatorname{tr}(\beta^\top M^{-1} \beta).$$

The equality is column-wise for all classes, and the correction and intercept remain unchanged. Factoring $M = C_M C_M^\top$ gives covariance $Q_0 C_M C_M^\top Q_0^\top$, so $Q_0 C_M$ with unit penalty is an exact implementation of the same model. Forming this basis may require reading the product features, but the fitted predictor need not retain all q columns. No approximation rank or compression-induced accuracy improvement is claimed.

Free intercept and the comparison space. The compression above retains the intercept convention. With a free intercept, set $\xi = Q_0^\top \mathbf{1}_N \in \mathbb{R}^{r_0}$ and write the total affine parametric component as $Q_0 \beta$. Minimizing its penalty over representations that differ by a constant gives

$$\min_c \frac{1}{2} \operatorname{tr}\big((\beta - \xi c^\top)^\top M^{-1} (\beta - \xi c^\top)\big) = \frac{1}{2} \operatorname{tr}(\beta^\top \Omega \beta),$$

where

$$\Omega = M^{-1} - \frac{M^{-1} \xi \xi^\top M^{-1}}{\xi^\top M^{-1} \xi}.$$

Differentiation in c gives the formula. Moreover $\Omega \succeq 0$ and $\Omega \xi = 0$, so constants are unpenalized rather than added as a cheaply penalized feature. Both Base with a free intercept and Parallel with a free intercept have affine span $\operatorname{col}([\mathbf{1}, \Phi])$. Without that intercept, their parametric spans need not agree. This elementary quotient-penalty calculation is used only to interpret the matched experiment, not as a new representer theorem.

A.3 Proof of the masking theorem and its nonvanishing extension

Here f_i is the scalar difference between the two class scores, and e_i is the coordinate unit vector at vertex i. The proof constructs diffusion coordinates, solves the classifiers, and rules out all admissible linear-feature anisotropies. Binary targets $y_S = (1, -1)^\top$ give the difference of the two scores fitted to centered one-hot targets. An error means that this scalar score has the wrong sign, with a fixed label-independent tie convention.

Lemma 1 (Contrast orientation under arbitrary linear-feature anisotropy)**.** *Let $B = [\mathbf{1}, \Phi]$, $S = \{s_+, s_-\}$, $v_s = e_{s_+} - e_{s_-}$ and $y_S = (1, -1)^\top$. Fix $H \succ 0$. Suppose two unannotated vertices satisfy*

$$B^\top(e_i - e_j) = \kappa B^\top v_s, \quad \kappa \geq 0, \qquad (e_i - e_j)^\top H^{-1} v_s > 0.$$

For every $F \succeq 0$ with $\mathrm{col}(F) \subseteq \mathrm{col}(B)$, the free-intercept fit with $K = H^{-1} + F$ satisfies $f_i > f_j$. In particular, it cannot classify both vertices correctly if their labels are $y_i = -1$ and $y_j = +1$.

Proof. Write $F = Q_0 M Q_0^\top$ in an orthonormal basis of $\mathrm{col}(B)$, with $M \succeq 0$. The contrast relation also holds with Q_0 in place of B, and therefore

$$(e_i - e_j)^\top F v_s = \kappa v_s^\top F v_s \geq 0.$$

The intercept equation makes the two dual coefficients opposite. Subtracting the support normal equations gives $\boldsymbol{r}_S = \rho(1, -1)^\top$, where

$$\rho = \frac{2}{2 + v_s^\top K v_s} > 0, \qquad f_i - f_j = \rho(e_i - e_j)^\top K v_s > 0.$$

The order required for the two specified labels is impossible. The result concerns fitted scores under the stated loss and fixed H, not the unconstrained set of all residual fields. □

A graph family with a simple second eigenvector. Fix an integer $r \geq 1$, put $N = 128r$, and form $63r$ ordinary pairs and r exceptional pairs. In this construction only, index vertices from 0 to $N - 1$; $v^{\oplus r}$ means r concatenated copies of v. Define

$$u = \frac{1}{\sqrt{24576r}} \left((1, -1)^{\oplus 63r},\ (\sqrt{12225}, -\sqrt{12225})^{\oplus r} \right)^\top, \qquad L = \frac{129}{128} \left(I - \frac{\mathbf{1}_N \mathbf{1}_N^\top}{N} \right) - \frac{uu^\top}{128}. \tag{15}$$

Here $\mathbf{1}^\top u = 0$ and $\|u\|_2 = 1$. For $i \neq j$,

$$W_{ij} = -L_{ij} = \frac{129}{16384r} + \frac{u_i u_j}{128} \geq \frac{65}{16384r} > 0,$$

because $|u_i| \leq 1/\sqrt{2r}$. Together with $L\mathbf{1} = 0$, this proves that L is a connected weighted-graph Laplacian. Its eigenvalues are 0 on $\mathbf{1}$, 1 on u, and $129/128$ on their orthogonal complement.

Choose $d = 2$ and $t = \log(4/3)$, with the natural logarithm. Equation (1) gives, after normalization,

$$(c_{\mathrm{o}}, \pm p_{\mathrm{o}}) = \frac{(16, \pm 1)}{\sqrt{257}}, \qquad (c_{\mathrm{e}}, \pm p_{\mathrm{e}}) = \frac{(16, \pm\sqrt{12225})}{\sqrt{12481}}. \tag{16}$$

Subscripts o, e name the ordinary and exceptional types; $p_{\mathrm{o}}, p_{\mathrm{e}} > 0$ are magnitudes, not signed vertex entries p_i. The same four rows arise spectrally for every r. Write $\Phi = [c, p]$. Both classes have $64r$ vertices: ordinary labels are $\mathrm{sign}(p)$ and exceptional labels are $-\mathrm{sign}(p)$. Observe any ordinary positive and any ordinary negative vertex. By permutation symmetry within types, relabel these as $0, 1$.

Fix $\alpha_\ell = 1$, $\alpha_q = 1/100$, and $H = I + L$ for all three models. Its inverse is

$$H^{-1} = hI + (1 - h)\frac{\mathbf{1}_N \mathbf{1}_N^\top}{N} + \epsilon uu^\top, \qquad h = \frac{128}{257}, \quad \epsilon = \frac{1}{514}. \tag{17}$$

The construction uses a weighted complete graph, two modes and unequal positive feature penalties. These are mathematical hypotheses, not the ten-neighbor image protocol or its reference penalties.

Projection and signed transmission. The even quadratic columns $c \odot c, p \odot p$ take two levels and lie in $\mathrm{span}(\mathbf{1}, c)$. The only odd column is $\sqrt{2}c \odot p$; the odd part of $\mathrm{col}[\mathbf{1}, \Phi]$ is spanned by p. Therefore

$$P(\sqrt{2}c \odot p) = \sqrt{2}\bar{c}\, p, \quad Q(\sqrt{2}c \odot p) = \sqrt{2}(c - \bar{c}\mathbf{1}) \odot p, \quad \bar{c} = \frac{63 c_{\mathrm{o}} p_{\mathrm{o}}^2 + c_{\mathrm{e}} p_{\mathrm{e}}^2}{63 p_{\mathrm{o}}^2 + p_{\mathrm{e}}^2}. \tag{18}$$

The common replication factor cancels. Swapping both signs of every pair leaves each kernel invariant and reverses the targets, so uniqueness gives a zero intercept. For each model,

$$f_i = \frac{K_{i0} - K_{i1}}{1 + K_{00} - K_{01}}, \qquad K_{i0} - K_{i1} = 2\epsilon u_i u_0 + 2p_i p_{\mathrm{o}} \Gamma(i) \quad (i \notin S), \tag{19}$$

with a strictly positive denominator and

$$\begin{aligned} \Gamma_{\textsc{Parallel}} &= 1 + 200\bar{c}^2, \\ \Gamma_{\textsc{Full}}(i) &= 1 + 200 c_i c_{\mathrm{o}}, \\ \Gamma_{\textsc{Decoupled}}(i) &= 1 + 200\big[\bar{c}^2 + (c_i - \bar{c})(c_{\mathrm{o}} - \bar{c})\big]. \end{aligned} \tag{20}$$

Every ordinary positive query has positive transmission for all three models; symmetry handles its negative counterpart. The two labeled scores also have their target signs. Parallel and Full have positive transmission to every exceptional positive query, so each makes $2r$ mistakes. Decoupled is different. Set

$$v = \bar{c}^2 + (c_{\mathrm{e}} - \bar{c})(c_{\mathrm{o}} - \bar{c}).$$

Exact comparisons of squared rationals show

$$.99805 < c_{\mathrm{o}} < .99806, \qquad .14321 < c_{\mathrm{e}} < .14322, \qquad .314 < \bar{c} < .315.$$

The last inequalities follow by substituting the first two in the positive weighted average in Eq. (18). These bounds give

$$v < 2(.315)^2 - .314(.998 + .143) + .999 \cdot .144 = -\frac{499}{31250}.$$

Since $p_{\mathrm{o}} > .06$, $p_{\mathrm{e}} > .98$ and $0 < u_i u_0 < 1/r$ for an exceptional positive query, its total transmission obeys

$$\epsilon \frac{u_i u_0}{p_{\mathrm{o}} p_{\mathrm{e}}} + 1 + 200v < \frac{1/514}{.06 \cdot .98} + 1 - 200\frac{499}{31250} = -\frac{51013759}{23611875} < 0. \tag{21}$$

All exceptional pairs are therefore correct for Decoupled. This proves the error counts and vertexwise Full–Parallel agreement without relying on floating-point signs.

No linear-feature anisotropy recovers the perfect classification. For every exceptional pair (i, j), Eq. (16) gives

$$B^\top (e_i - e_j) = \frac{p_{\mathrm{e}}}{p_{\mathrm{o}}} B^\top (e_0 - e_1), \qquad (e_i - e_j)^\top H^{-1} (e_0 - e_1) = 4\epsilon u_i u_0 > 0.$$

Its two true labels require the reverse order. Lemma 1 forces at least one mistake on each of the r disjoint exceptional pairs for every admissible F. Hence its error is at least $r/(128r - 2)$. This uniform statement includes every positive-definite coefficient penalty on the fixed linear-feature space; it does not range over a different geometry, H, loss or decoder.

Span, trace, perturbations and sampling. Unit row norms imply $\mathbf{1} \in \mathrm{col}(\Psi)$, so $\mathrm{col}[\Phi, \Psi] = \mathrm{col}[\Phi, P\Psi, Q\Psi]$. Orthogonality gives $\|P\Psi\|_F^2 + \|Q\Psi\|_F^2 = \|\Psi\|_F^2$ and hence equal weighted feature-covariance traces for Full and Decoupled. Their total kernels remain positive definite.

For each fixed r, the second eigenvalue is simple with gap $1/128$, every edge weight is strictly positive, and $[\mathbf{1}, \Phi]$ has rank three. These properties hold in a sufficiently small neighborhood. Choose the second eigenvector continuously there. The coordinates, projector, kernels and fitted scores then vary continuously. Every displayed classification has a strict margin, so all three classifications persist locally. This argument does not extend the universal anisotropy obstruction to arbitrary perturbations; that obstruction uses the exact contrast relations.

Each class contains $63r$ ordinary vertices out of $64r$. Uniform independent sampling of one label per class selects an ordinary positive and ordinary negative with probability $(63/64)^2 = 3969/4096$. Every such support is covered by the construction. In particular, the phenomenon is not limited to one specially indexed pair. Finally, $2/N = 1/(64r) \to 0$, but the Full–Decoupled error gap tends to $1/64$, and the all-anisotropy lower bound tends to $1/128$. This completes the proof of Theorem 1.

Corollary 1 (An exact coupling threshold in the constructed family)**.** *Use the path from Decoupled to Full in Eq.* (11)*. Define*

$$\zeta_r = \frac{\sqrt{12225}}{514 \cdot 24576 r\, p_{\mathrm{o}} p_{\mathrm{e}}}, \qquad \gamma_*(r) = -\frac{1 + 200v + \zeta_r}{200(c_{\mathrm{o}} c_{\mathrm{e}} - v)} \in (0, 1).$$

Every unannotated vertex is correct when $0 \le \gamma < \gamma_*(r)$; exactly $2r$ are incorrect when $\gamma_*(r) < \gamma \le 1$. The exceptional vertices have zero margin at the threshold. For $r = 1$, $\gamma_* \approx 0.08186$.

Proof. The denominator in Eq. (19) remains positive along the path. Dividing an exceptional query's numerator by $2p_{\mathrm{o}}p_{\mathrm{e}}$ leaves

$$1 + 200v + \zeta_r + 200\gamma(c_{\mathrm{o}} c_{\mathrm{e}} - v),$$

which is negative at zero, positive at one, and strictly increasing. Ordinary transmission is positive at both endpoints and affine in the numerator, so it stays positive throughout. This proves the threshold and the error counts. The resulting curve is an exact property of the construction, not a fitted explanation of an image-benchmark transition. □

Reproducible checks. The script `code/check_theory.py` verifies the rational sign bounds with exact fractions and checks the graph spectra and projected dictionaries for $r = 1, 2, 4$. Six independent primal/kernel comparisons and 129 direct path evaluations check the fitted scores and the closed-form threshold. Replication does not supply independent statistical samples: Figure 3 illustrates a proved family. These checks supplement the proof and do not alter image-benchmark observations.

A.4 Proof of the angular remainder and its prediction bound

The scalar identity $x^2 = 2x - 1 + (1 - x)^2$, applied entrywise to G, gives Eq. (7). Unit row norms imply $\|\phi_i - \phi_j\|_2^2 = 2(1 - G_{ij})$. In particular, $R_{ii} = 0$ and $R_{ij} \ge 0$. Except in degenerate cases, this remainder is not positive semidefinite: a positive-semidefinite matrix with zero diagonal is zero. None of the following argument requires $R \succeq 0$.

We have $K_F = K_L + \omega R - \omega \mathbf{1}_N \mathbf{1}_N^\top$. The last term has no action on the observed contrast space or on $\mathcal{C}$. Therefore

$$A_{K_F} = A_{K_L} + \omega R_S, \qquad \mathcal{C}_i(K_F) = \mathcal{C}_i(K_L) + \omega \mathcal{C}_i(R).$$

If $a_F = A_{K_F}^{-1} T_S^\top Y$ and $a_L = A_{K_L}^{-1} T_S^\top Y$, subtracting their equations yields $a_F - a_L = -\omega A_{K_L}^{-1} R_S a_F$. Substituting in the two prediction formulas proves Eq. (10) exactly.

Define the diameter over *all* normalized graph coordinates by

$$\delta = \max_{1 \le i,j \le N} \|\phi_i - \phi_j\|_2 \in [0, 2], \qquad r_* = \frac{\delta^4}{4}.$$

When the diffusion time varies, write $\delta(t)$ for this same maximum. Every entry of R is at most r_*. Since T_S has orthonormal columns, restriction to observed contrasts is norm nonexpansive, so

$$\|\mathcal{C}_i(R)\|_2 \le 2\sqrt{m}\, r_*, \qquad \|R_S\|_2 \le m r_*, \qquad \|a_F\|_2 \le \left\|T_S^\top Y\right\|_2.$$

The final inequality follows from $K_F \succeq 0$ and $A_{K_F} \succeq I$. The product inequality gives the explicit bound

$$\|(U_F - U_L)_i\|_2 \le \frac{\omega \delta^4}{4} (2\sqrt{m} + m \|D_i\|_2) \left\|T_S^\top Y\right\|_2. \tag{22}$$

A unique winning margin of U_L greater than $\sqrt{2}$ times this bound is sufficient for equal argmax, because the norm of a difference of two class coordinate vectors is $\sqrt{2}$.

Corollary 2 (When scalar reweighting is exactly sufficient)**.** *Fix $H \succ 0$, $a, \omega > 0$, a unit-row dictionary Φ, and a real ρ such that $a + \omega\rho > 0$. The free-intercept estimators with kernels $H^{-1} + aG + \omega J$ and $H^{-1} + (a + \omega\rho)G$ agree for every nonempty annotated set and every real target matrix if and only if*

$$G_{ij} \in \{1, \rho - 1\} \qquad \textit{for every } i, j.$$

For more than one distinct feature row, all distinct rows must therefore have the same mutual inner product, and there are at most $d + 1$ such rows.

Proof. Sufficiency follows from $J - \rho G = (1-\rho)\mathbf{1}_N\mathbf{1}_N^\top$ and invariance of the free-intercept fit to constant covariance. For necessity, take the annotated set to be all vertices. On the space orthogonal to $\mathbf{1}$, the fitted-score operator is $B(I+B)^{-1}$, where B is the covariance restricted to that space. The map is injective because $B = F(I-F)^{-1}$ for $F = B(I+B)^{-1}$. Thus equality for all targets implies $\Pi(J-\rho G)\Pi = 0$, with $\Pi = I - \mathbf{1}_N\mathbf{1}_N^\top/N$. A symmetric matrix with zero compression to $\mathbf{1}^\perp$ has the form $\mathbf{1}v^\top + v\mathbf{1}^\top$. Its diagonal here is constantly $1-\rho$, so $v = (1-\rho)\mathbf{1}/2$. Entrywise, $(G_{ij}-1)(G_{ij}-(\rho-1)) = 0$. Unit vectors have inner product one only if equal. The Gram matrix of p distinct rows with common inner product $c < 1$ is $(1-c)I_p + c\mathbf{1}_p\mathbf{1}_p^\top$, whose rank is at least $p-1$. Therefore $p \le d+1$. □

The constant-angle Gram geometry is classical. The corollary states exactly when it makes the two regularized estimators indistinguishable for all label problems. Actual classification labels on a particular support can agree even when this condition fails.

Diffusion contraction. In Eq. (1), divide z_i by its constant first coordinate and write

$$\phi_i(t) = \frac{(1, h_i(t)^\top)^\top}{\sqrt{1+\|h_i(t)\|_2^2}}, \qquad h_i(t) = \sqrt{N}\big(e^{-t\lambda_j/2}u_j(i)\big)_{j=2}^{d}.$$

The map $h \mapsto (1,h^\top)^\top/\sqrt{1+\|h\|_2^2}$ is 1-Lipschitz: its differential has squared norm in direction v equal to $\|v\|_2^2/(1+\|h\|_2^2) - (h^\top v)^2/(1+\|h\|_2^2)^2 \le \|v\|_2^2$. With

$$a_t^2 = N\max_i \sum_{j=2}^{d} e^{-t\lambda_j}u_j(i)^2 \le (N-1)e^{-t\lambda_2},$$

we obtain $\delta(t) \le 2a_t$. Also $\|D_i\|_2$ is uniformly bounded in t: the entries of G are bounded by one, H and the penalties are fixed, and $\big\|A_{K_L}^{-1}\big\|_2 \le 1$. Equation (22) then gives the stated $O(e^{-2t\lambda_2})$ rate. The constants depend on the fixed graph, support, penalties and targets, not on t.

Corollary 3 (Leading finite-graph term)**.** *Let $v_i \in \mathbb{R}^{d-1}$ have component $(v_i)_{j-1} = u_j(i)$ when $\lambda_j = \lambda_2$ and zero otherwise ($2 \le j \le d$). Thus the first-positive-eigenspace coordinates are embedded in the retained coordinate space. Set $R_{ij}^{(2)} = \|v_i - v_j\|_2^4$. Put $K_\infty = H^{-1}$, $A_\infty = I + T_S^\top(K_\infty)_{SS}T_S$, and $D_{\infty,i} = \mathcal{C}_i(K_\infty)A_\infty^{-1}$. For fixed $d \ge 2$,*

$$\lim_{t\to\infty} e^{2t\lambda_2}(U_F - U_L)_i = \frac{\omega N^2}{4}\big[\mathcal{C}_i(R^{(2)}) - D_{\infty,i}T_S^\top R_{SS}^{(2)}T_S\big]A_\infty^{-1}T_S^\top Y. \tag{23}$$

Proof. For the fixed finite spectrum, $h_i(t) = \sqrt{N}e^{-t\lambda_2/2}v_i + o(e^{-t\lambda_2/2})$ in the retained coordinate space. Expansion of the normalization gives $\|\phi_i(t) - \phi_j(t)\|_2^2 = Ne^{-t\lambda_2}\|v_i - v_j\|_2^2 + o(e^{-t\lambda_2})$. Thus $e^{2t\lambda_2}R(t) \to N^2R^{(2)}/4$. Both $G(t)$ and $J(t)$ tend to $\mathbf{1}_N\mathbf{1}_N^\top$, whose contribution vanishes in $A_K - I$ and $\mathcal{C}(K)$. Hence both support systems converge to A_∞ and $D_i \to D_{\infty,i}$. Taking the limit in Eq. (10) proves the formula. □

If the right-hand side is nonzero for some row and class, the rate is attained for that prediction difference; if it vanishes, label alignment can yield a faster rate. The corollary does not assume independent observations, a growing dimension, or a random graph. It gives a different setting and conclusion from random high-dimensional kernel linearization (El Karoui, 2010). It also does not convert an asymptotic score expansion into a guaranteed classification gain.

A.5 Signed coupling and preservation of predictor span

Let $A_p = P\Psi$, $B_p = Q\Psi$. Since $\mathbf{1} \in \mathrm{col}(\Psi)$, both $P\Psi$ and $Q\Psi$ belong to $\mathrm{col}([\Phi,\Psi])$; the reverse span inclusion follows from $\Psi = P\Psi + Q\Psi$. With the orthogonal reflection $R_P = 2P - I$,

$$J_\gamma = J_0 + \gamma(PJQ + QJP) = \frac{1+\gamma}{2}J + \frac{1-\gamma}{2}R_PJR_P.$$

This proves positive semidefiniteness for $|\gamma| \le 1$ and trace preservation. In orthonormal bases of $\mathrm{range}(P)$ and $\mathrm{range}(Q)$, write the cross term as $\left(\begin{smallmatrix} 0 & B_\times \\ B_\times^\top & 0 \end{smallmatrix}\right)$. Its nonzero eigenvalues are $\pm s_j(B_\times)$, where s_j are the positive singular values. Its rank is $2k$, $k = \operatorname{rank} B_\times \le d+1$. Also

$$\|PJQ + QJP\|_2 = \left\|A_p B_p^\top\right\|_2 \le \|A_p\|_F \|B_p\|_F = N\sqrt{\varepsilon(1-\varepsilon)}, \qquad \varepsilon = \|Q\Psi\|_F^2 / N.$$

Matrix square roots below are the symmetric positive-definite roots. For $0 < |\gamma| \le 1$, put $X = K_0^{-1/2}(K_\gamma - K_0)K_0^{-1/2}$. Congruence gives k positive and k negative eigenvalues of X. Since $K_\gamma \succ 0$, each eigenvalue x of X exceeds -1. The inverse difference is congruent to $(I+X)^{-1} - I$, with eigenvalues $-x/(1+x)$: signs reverse, but the counts remain k and k. At $\gamma = 0$, both X and $K_\gamma^{-1} - K_0^{-1}$ are zero. This redistributes penalties, without guaranteeing better accuracy. The argument combines block pinching (Bhatia, 1997) with the congruence calculation above.

A.6 Finite response, norm certificate and class-wise paths

For a free intercept take $\mathsf{T} = T_S$ and $\mathcal{C}$ from Eq. (9); without an intercept take $\mathsf{T} = I_m$ and $\mathcal{C}_i(K) = K_{iS}$. Put

$$A_0 = I + \mathsf{T}^\top (K_0)_{SS} \mathsf{T}, \quad B_S = \mathsf{T}^\top D_{SS} \mathsf{T}, \quad g_i = \mathcal{C}_i(D) - \mathcal{C}_i(K_0) A_0^{-1} B_S. \tag{24}$$

Let r_S be the number of columns of T. Write $\mathsf{M} = A_0^{-1/2} B_S A_0^{-1/2} = O\,\mathrm{diag}(\mu_j) O^\top$, with $O \in \mathbb{R}^{r_S \times r_S}$ orthogonal, $w_i = g_i A_0^{-1/2} O$ and $Z_c = O^\top A_0^{-1/2} \mathsf{T}^\top Y$. Here $w_i \in \mathbb{R}^{1 \times r_S}$ and $Z_c \in \mathbb{R}^{r_S \times C}$; all mode sums run over $j = 1, \dots, r_S$. The coefficient in Theorem 3 is $a_{ichj} = w_{ij}((Z_c)_{jc} - (Z_c)_{jh})$.

With the intercept convention encoded by $\mathsf{T}, \mathcal{C}$, put $A_\gamma = A_0 + \gamma B_S$. The prediction has the same constant part for every γ, and its varying part is $\mathcal{C}_i(K_\gamma) A_\gamma^{-1} \mathsf{T}^\top Y$. Subtracting the two support equations gives

$$(U_\gamma - U_0)_i = \gamma g_i A_\gamma^{-1} \mathsf{T}^\top Y. \tag{25}$$

Both $K_0 + D$ and $K_0 - D$ are positive definite, because their quadratic terms are J and $R_P J R_P$ divided by α_q. Consequently $I \pm \mathsf{M} \succ 0$ for $\mathsf{M} = A_0^{-1/2} B_S A_0^{-1/2}$, so $\eta = \|\mathsf{M}\|_2 < 1$. The normalized resolvent gives

$$b_i(\gamma) = \frac{|\gamma| \left\| g_i A_0^{-1/2} \right\|_2 \left\| A_0^{-1/2} \mathsf{T}^\top Y \right\|_2}{1 - |\gamma|\eta}, \qquad \|(U_\gamma - U_0)_i\|_2 \le b_i(\gamma). \tag{26}$$

For a unique initial winner $c_i = \arg\max_c U_{0,ic}$, its winning margin is $\min_{h \ne c_i} f_{ic_i h}(0)$. If this margin is greater than $\sqrt{2} b_i(\gamma)$, the winning class is unchanged. Since $x/(1 - x\eta)$ is increasing for $0 \le x \le 1$, $b_i(1)$ bounds the whole interval. The test with $b_i(1)$ is the uniform norm baseline used in the new path experiment; the older endpoint panel is retained separately in the numerical companion. For any labels, the absolute change in percentage accuracy is at most 100 times the fraction of changed decisions, itself bounded by 100 times the fraction not certified by a valid stability test. The latter is a deterministic bound, not a statistical interval.

Proof of Theorem 3. Diagonalizing M in Eq. (25) yields

$$U_{\gamma,i,:} - U_{0,i,:} = w_i\,\mathrm{diag}\left(\frac{\gamma}{1+\mu_j\gamma}\right) Z_c.$$

Taking a difference of two class columns proves Eq. (12), including $\mu_j = 0$. Every function $h_j(x) = x/(1+\mu_j x)$ is strictly increasing on $[0,1]$, since $h_j'(x) = (1+\mu_j x)^{-2} > 0$. Minimize each signed term separately over $[u, v]$ to obtain Eq. (13). The sum of these minima is a valid lower bound even though different terms attain their minima at different endpoints.

Refinement raises each local lower bound relative to its containing interval. Its conservatism is bounded by

$$(v-u) \sum_j \frac{|a_{ichj}|}{(1-|\mu_j|)^2}.$$

Indeed, this bounds the sum of ranges of all signed terms on the interval. If the winning class has a strictly positive margin against every rival throughout $[0, 1]$, continuity and compactness give a positive minimum margin. A finite mesh finer than this minimum divided by the largest displayed Lipschitz constant makes every lower bound positive. Therefore the procedure is complete for strictly positive-margin paths. A touching zero margin, an actual class change, or a prematurely terminated refinement need not be certified.

Operational distinction. The norm certificate uses a global label norm and η; the class-wise test retains signs and rival-specific label differences. Neither requires true labels on the evaluated set to certify stability. Once the support eigensystem and query factors are computed, subdivision uses scalar rational evaluations, not refitting. Numerically rigorous certification would additionally require outward-rounded bounds for all input and solver errors. The benchmark experiment and the supplied small-system implementation use floating point and are explicitly numerical evaluations, not such an interval-arithmetic guarantee.

A.7 Independent mathematical checks and numerical complexity

The same script checks 80 randomly generated small matrix problems, including projected compression, the angular remainder and its prediction bound. It compares both the derived path formula and the supplied implementation with 800 direct evaluations, with and without an intercept. It also checks 20 constant-angle scalar-recovery cases, a fixed connected path graph for the diffusion expansion, and interval refinement. These calculations support the proofs; they generate no image-classification results. The exact scalar-match experiment fixes its comparator from Theorem 2, without selecting it by test performance. The signed path experiment evaluates Theorem 3 with fixed geometry and penalties, not with a coupling chosen to maximize accuracy.

For the original estimator, $F = ZA^{-1}Z^\top$ and $M_S = I + F_{SS}$. Without an intercept, elimination gives

$$(H + E^\top M_S^{-1} E)V = E^\top M_S^{-1} Y, \quad R_Y = M_S^{-1}(Y - EV), \quad \Theta = A^{-1}Z^\top E^\top R_Y.$$

The implementation uses Cholesky on the support and batched class-wise conjugate-gradient updates for the sparse system. With $s = \mathrm{nnz}(L)$ the number of nonzero entries and I_{CG} conjugate-gradient iterations, the Gram-based Full path costs $O(m^2 d + m^3 + I_{\mathrm{CG}} C(s + m^2) + (N + m)d^2 C)$. Explicit product assembly replaces $m^2 d$ by $m^2 d^2$. Preparing projected blocks is additional. Positive σ guarantees definiteness, not a fixed iteration count. Computing certificate query blocks requires Green actions and support matrices, which are reused between endpoints; it is not free.

B Experimental specification

B.1 Inputs, supports and distinct empirical panels

The supplied image partitions are concatenated in the public neighbor-array order. Ten non-self neighbors are selected from eleven columns including self. Let d_{ij} be the distance in the supplied encoding and $d_{i,10}$ the distance to vertex i's tenth non-self neighbor (not to vertex 10). Directed weights are $\exp(-4d_{ij}^2/d_{i,10}^2)$; mean symmetrization and a zero diagonal give the fixed W. The resulting nonzero counts are 1,014,572, 978,786 and 1,039,490 for MNIST, Fashion-MNIST and CIFAR-10. The VAE/AET representations are not retrained (Calder et al., 2020; Zhang et al., 2019).

The dense campaign uses seeds 20270100 through 20270109. A class-wise permutation and its prefixes produce nested stratified supports. Its 52 budgets are $\{1, \ldots, 40\} \cup \{50, 60, \ldots, 160\}$. Seven-budget controls use $\{1, 5, 10, 20, 40, 80, 160\}$; rank/time and coupling controls use $\{5, 40, 160\}$. The reference model has $\alpha_\ell = \alpha_q = .03$, $\tau = 1$, $\sigma = .01$, $d = 64$ and $t = t_0$, without an intercept unless stated. The fixed factorial crossing has $d \in \{32, 64, 128\}$ and $t/t_0 \in \{.5, 1, 2\}$.

The complementary joint-selection and expanded-LapRLS panel uses seeds 20270100, 20270101 and 20270104, chosen by availability of prior computations rather than by final accuracies. Its 3 trajectories form a separate panel from the ten-seed study. The earlier full-query norm diagnostics use 20270100 and 20270101. The new angular and path panels use all three listed seeds; their query sample is not the earlier full-query population. All predictions are evaluated on $N - 10b$ unannotated

vertices. Vertices, validation folds, candidate evaluations and repeated numerical checks are not extra independent seed trajectories.

B.2 Feature controls and joint selection

BaseTrace doubles the linear covariance by halving its feature penalty. FullTrace and DecoupledTrace instead multiply the quadratic covariance by $\|Q\Psi\|_F^2 / \|\Psi\|_F^2$, leaving the linear block unchanged. Parallel retains projected products, and OQDiff retains the complementary products. These interventions are distinct. The mean-scaled control multiplies every penalty by $m/50$, preserving a fixed mean-loss interpretation rather than fixed summed-loss coefficients.

The original selected Base, Full and Decoupled share a free intercept and the 27 candidates obtained by multiplying the common feature penalty, τ and σ by $\{.1, 1, 10\}$. At $b = 1$, a predefined candidate is used without cross-validation. All other selection uses folds within the observed support, with $\min(3, b)$ folds, pooled validation accuracy and fixed configuration-order ties.

The joint-selection panel allows geometries G64 $(64, t_0)$ and G128 $(128, t_0/2)$. In each geometry, Base keeps the 27 linked candidates. For each enriched method, the 27 linked candidates are supplemented with the 18 combinations of unequal $(\alpha_\ell, \alpha_q) \in \{.003, .03, .3\}^2$, $(\tau, \sigma) \in \{(.1, .001), (1, .01), (10, .1)\}$. This gives 54 Base candidates and 90 for each enriched procedure across both geometries. The candidate budgets differ, so this is not an equal-compute comparison. A common unpenalized intercept and the same three folds are used for every method. Selection never sees evaluation labels; final fitting uses all observed labels. Chosen candidates remain attributes of individual trajectories and do not partition their reported method means.

For the shared-setting intervention, only Base's 54 candidate scores enter the choice. We use the same pooled integer validation counts and fixed-order tie rule, then hold the selected geometry, α_ℓ, τ and σ fixed for Base, Parallel, Decoupled and Full, with $\alpha_q = \alpha_\ell$. Every fit uses the same observed support and a free intercept. The reported extension reuses the verified Base validation counts and computes 108 new final fits over the three graphs, three budgets and three seeds. A complete rerun computes that validation itself and reuses its Base refit. The two decoders are recorded separately; neither oracle counts nor evaluation accuracy enter selection.

B.3 LapRLS and propagation comparators

With $X = \Phi$ or $X = [\Phi, \Psi]$, LapRLS solves

$$\min_{\Theta, c} \frac{1}{2m} \left\| X_S\Theta + \mathbf{1}_m c^\top - Y \right\|_F^2 + \frac{\lambda_A}{2} \|\Theta\|_F^2 + \frac{\lambda_I}{2N^2} \operatorname{tr}(\Theta^\top X^\top L X \Theta).$$

The explicit dictionary defines the kernel, not an approximation to an unspecified RBF. Multiplying by m gives summed-loss coefficients $m\lambda_A$ and $m\lambda_I/N^2$. Within a validation fold, m is its training count. Thus a numerically identical ambient penalty is not identical across loss conventions or label budgets.

The original grid is $\lambda_A \in \{.003, .03, .3\}$, $\lambda_I \in \{10^5, 10^6, 10^7\}$. The extended grid retains these values and adds $3 \times 10^{-7}, 3 \times 10^{-6}, 3 \times 10^{-5}, 3 \times 10^{-4}$. Old and extended candidates are selected on identical folds at G64 with a free intercept. Of 54 old/extended decisions, 40 change with strictly better validation accuracy; the other 14 reuse the same selected fit. A validation improvement need not improve evaluation accuracy. The extended minimum is still selected in 27 decisions, so no interior or global optimum is inferred.

C&S uses either a softmax predictor or a width-64 one-hidden-layer ReLU MLP on G64, followed by both correction and smoothing. The historical study trains for 200 epochs in float64 with Adam learning rate .03 and weight decay in $\{10^{-4}, 10^{-3}, 10^{-2}\}$. Here L_c, L_s are correction/smoothing iteration counts, not Laplacians, and α_c, α_s are their propagation parameters. The tuples $(L_c, \alpha_c, L_s, \alpha_s)$ are $(50, .5, 50, .8)$, $(50, .8, 50, .8)$ and $(100, .8, 100, .9)$, combined with autoscaling or fixed scale 1 or .5, giving 27 candidates per family. Both stages use symmetrically normalized adjacency in this declared adaptation of Huang et al. (2021). Predictor outputs reused by postprocessing candidates are not separate neural trainings. Poisson Learning uses GraphLearning on the same W; PoissonMBO uses the fixed diffusion/volume recipe with population proportions. No new C&S or Poisson result is in-

troduced into the three-seed joint-selection table. CutSSL remains related work without a quantitative comparison in this experiment set.

B.4 Statistics and numerical decision diagnostics

Here pp means percentage points. For seed s, let $\Delta_s(b)$ be the paired difference in percentage accuracy and $b_1 < \cdots < b_{52}$ the measured budgets. Its dense summary is

$$\overline{\Delta}_s = \frac{1}{\log 160} \sum_{j=1}^{51} \frac{\Delta_s(b_j) + \Delta_s(b_{j+1})}{2} \log \frac{b_{j+1}}{b_j}.$$

The compact and new panels average their seven and three budgets, respectively, inside each seed before averaging seeds. For multi-budget summaries, sample standard deviations measure variation between seed-wise averages, not between budgets. For the historical intervals, 10,000 bootstrap draws with seed 271828 resample complete paired trajectories. Intervals are exploratory and marginal, conditional on the fixed graphs.

Three-trajectory displays use observed paired effects and ranges. The companion retains the individual observations needed to recalculate these summaries. Neither a range nor an interval containing zero establishes equivalence.

B.5 Exact scalar-match experiment

The two geometries are G64/t_0 and G128/$(t_0/2)$. For each graph, budget $5, 40, 160$ and seed $20270100, 20270101, 20270104$, Full uses $\alpha_\ell = \alpha_q = .03$, $\tau = 1$, $\sigma = .01$ and a free intercept. The matched linear-feature model changes only its coefficient penalty to $\alpha_{\mathrm{match}} = .01$. The graph, support, labels and intercept convention agree within each pair. Two model fits and an auxiliary linear-system response check provide the scores and the finite identity. Per-vertex score differences, margin quantiles and disagreement counts are recorded; all accuracies use the same unannotated complement.

The 18 graph/geometry/budget means and ranges are generated directly from the 54 individual pairs. Disagreement counts are summed over the three supports and divided by the sum of their evaluated counts when expressed as fractions. The maximum absolute response discrepancy is 2.58×10^{-8} (relative maximum 1.10×10^{-8}). At the tested finite times, $t\lambda_2 \in [0.296, 1.191]$. The median ratio of the uniform curvature bound to the largest rowwise score difference is 3.52×10^{9}: the limiting theorem is not used as an empirically tight bound.

B.6 Whole-path diagnostic protocol and costs

The new panel crosses three graphs, budgets $5, 40, 160$, the three seeds above and both intercept conventions: 27 supports and 54 configurations. G64 and penalties $(.03, .03, 1, .01)$ are fixed. A SHA-256-derived PCG64 rule selects 4,096 vertices uniformly without replacement from each unannotated complement; their sorted indices are reused across the three diagnostics and both intercepts. Selection uses neither held-out labels nor prediction scores. The supplementary sample records allow reconstruction of these indices.

The norm baseline uses the uniform bound $b_i(1)$ in Eq. (26). The signed direct test applies Eq. (13) on $[0, 1]$ for the initial winner against every rival. The adaptive variant subdivides uncertified intervals with maximum depth 16 and at most 8,192 examined intervals. The sign threshold is 10^{-10}. Modes with $|\mu_j| \le 10^{-10}$ are grouped with an explicit error envelope subtracted from the bounds. A fixed grid of 65 equally spaced coupling values supplies negative-margin witnesses and filters these cases before further refinement; it never establishes continuous stability by absence of a witness. Ties or uncertain margins and refinement limits are separately recorded. A path with strictly positive margins is finitely certifiable in exact arithmetic by Theorem 3; the finite computational limits need not certify every such path.

On this panel, the observed sets satisfy norm $\subseteq$ signed $\subseteq$ adaptive. This nesting is an empirical property of these records, not a universal dominance theorem. Counts pool query/configuration evaluations; the same vertex can appear under both intercepts or at multiple budgets. The 31 change-and-return records correspond to 24 distinct support/vertex identities after merging the two intercepts.

One free-intercept CIFAR-10/$b = 5$ record reaches the refinement limit and is not called stable. No positive diagnosis has a contrary recorded grid witness.

All implementations use floating point and approximate sparse solves. The maximum recorded endpoint-response discrepancy is 2.83×10^{-8}. Solver residuals and mode-grouping bounds do not constitute directed rounding of the complete chain. We therefore report numerical coverage of the theoretical tests, not formal finite-precision certificates. Stability does not imply a correct class, and an unchanged endpoint pair need not be unchanged throughout the path.

B.7 Recorded margin safety and cutoff sensitivity

For the initial winning class $c_i = \arg\max_c U_{0,ic}$, let $\ell_i = \min_{h \neq c_i} L_{ic_ih}[0, 1]$ be the stored direct signed lower bound. This section applies stricter cutoffs to those same records; it does not recompute a factorization, path or sparse solve. Among the 206,078 direct diagnoses, the minimum ℓ_i is 2.2432×10^{-6}; its 1st percentile is 8.906×10^{-3} and its median is 0.6622. Table 1 gives the retained counts. All direct diagnoses survive 10^{-6}; only 31 are lost at 10^{-4}.

Table 1: Postprocessing sensitivity of the *direct signed* test to its stored positive-bound cutoff. Coverage uses the full 221,184-query panel. No solver tolerance is changed and no numerical error bound is inferred.

Stored-bound cutoff	Retained signed	Removed	Panel coverage (%)
10^{-10}	206,078	0	93.17
10^{-6}	206,078	0	93.17
10^{-5}	206,072	6	93.17
10^{-4}	206,047	31	93.16
10^{-3}	205,850	228	93.07

For an exact lower bound approximated with a proven overestimation error at most ε, requiring the reported bound to exceed ε would preserve the certificate. Here the recorded maximum endpoint-response discrepancy, 2.83×10^{-8}, is a diagnostic, not such a uniform bound along the path. The cutoff sweep therefore supports the separation of the direct test from its decision threshold, not formal finite-precision certification.

The 11,122 cases certified only after subdivision do not have their final minimum over accepted interval bounds in the compact export. The negative original ℓ_i for those cases is not that final minimum and is not reused as if it were. Their per-case safety under a tighter solve or a larger final cutoff is not established by this sweep. Full path factors and an export of accepted leaf bounds would be needed for that separate analysis. The reported adaptive coverage remains a numerical diagnostic at the declared tolerances.

Table 2: Measured path-diagnostic costs, in seconds. The internal rows are included in the post-Green phase and must not be added to it. The extra refinement is run after grid-witness filtering; its time is not the cost of a standalone no-grid algorithm.

Measured component	Seconds
New Green actions: 26 supports, 16,850 columns	2317.36
Post-Green diagnostic phase	105.83
Support factorization, included	24.77
Query factors, included	6.69
Norm test, included	0.081
Signed direct evaluation, included	4.55
Additional refinement, included	1.62
Grid witnesses, included	7.72

One Green action is reused from an existing support. Shared phases are counted once per support, not once per intercept. The post-Green time includes some large score writes and orchestration, but excludes preceding loads. Both measured phases total about 40 minutes; this is not a cold end-to-end

runtime from images. The recorded platform is the Apple M4 Pro system with 24 GB RAM and one numerical thread per worker, with preparation shared between the two intercept conventions.

Prepared fitting time. The separate CPU session alternates 72 warmed measurements of Full and PoissonMBO: three repetitions on one support per graph at budgets $1, 5, 40, 160$. Within-cell median ratios range from 3.80 to 5.23. The recorded arm64/macOS environment uses Python 3.11.16, NumPy 2.3.5, SciPy 1.17.0 and single-threaded Apple Accelerate; the inspected machine has an Apple M4 Pro and 24 GB RAM. These prepared-fit times exclude preprocessing, validation, decoding and storage. Spectral preparation alone takes approximately 68, 138 and 79 seconds; total-cost amortization is not inferred from it.

C Additional evidence without pooling protocols

C.1 Fixed-representation controls and geometry

The common-free-intercept attribution uses 210 matched model quadruples: three graphs, seven budgets and ten seeds. All four refits have $d = 64$, $t = t_0$, $(\alpha_\ell, \alpha_q, \tau, \sigma) = (.03, .03, 1, .01)$, summed loss and the same unannotated evaluation set. We average each contrast over the seven budgets within a seed, then over ten seeds. Bootstrap resampling acts on complete trajectories. The first contrast is within a common affine parametric span; the subsequent two add the independent complement and then restore the cross covariance.

Table 3: Ordered attribution with a common free intercept: seven-budget mean accuracy differences (pp), ten paired trajectories, marginal 95% bootstrap intervals. Each interval is recomputed for its own paired contrast. These are the increments of Figure 2.

Paired contrast	MNIST	Fashion-MNIST	CIFAR-10
Parallel − Base	$+0.86\ [+0.55, +1.24]$	$+0.50\ [+0.40, +0.62]$	$+1.34\ [+1.05, +1.62]$
Decoupled − Parallel	$+0.00\ [-0.07, +0.06]$	$+0.33\ [+0.29, +0.38]$	$+0.57\ [+0.48, +0.66]$
Full − Decoupled	$-0.04\ [-0.06, -0.02]$	$-0.18\ [-0.22, -0.13]$	$-0.40\ [-0.45, -0.34]$
Full − Base	$+0.82\ [+0.57, +1.15]$	$+0.66\ [+0.57, +0.76]$	$+1.52\ [+1.15, +1.85]$

Table 4: The same seven budgets and ten supports with the intercept fixed at zero. The first contrast may add the constant direction and therefore is not a pure within-span penalty contrast. This table and Table 3 keep the aggregation weights fixed when comparing intercept conventions.

Paired contrast	MNIST	Fashion-MNIST	CIFAR-10
Parallel − Base	$+0.88\ [+0.59, +1.26]$	$+0.58\ [+0.47, +0.69]$	$+1.45\ [+1.06, +1.78]$
Decoupled − Parallel	$+0.03\ [+0.01, +0.06]$	$+0.32\ [+0.29, +0.35]$	$+0.60\ [+0.50, +0.71]$
Full − Decoupled	$-0.03\ [-0.05, -0.01]$	$-0.19\ [-0.24, -0.16]$	$-0.46\ [-0.53, -0.39]$

The nine rank/time configurations form a fixed factorial control, not selection by evaluation accuracy. After fixed-geometry penalty selection, Decoupled–Full averages are $.01, .17, .23$ points; the native CIFAR-10 interval spans zero. All nine geometries and both selection protocols remain separately identified in the numerical companion.

A separate fixed-penalty, common-intercept control has Full below its predictor-only and sequential alternatives on the compact CIFAR-10 mean, by 1.16 and 1.31 points. This supports the paper's separation between usefulness of features and superiority of the joint residual organization. The same data do not establish an optimally selected sequential procedure.

C.2 Exact scalar match at every evaluated budget

The score-response records and per-seed effects in `result/` preserve every matched geometry and budget, without independent optimization of the two procedures.

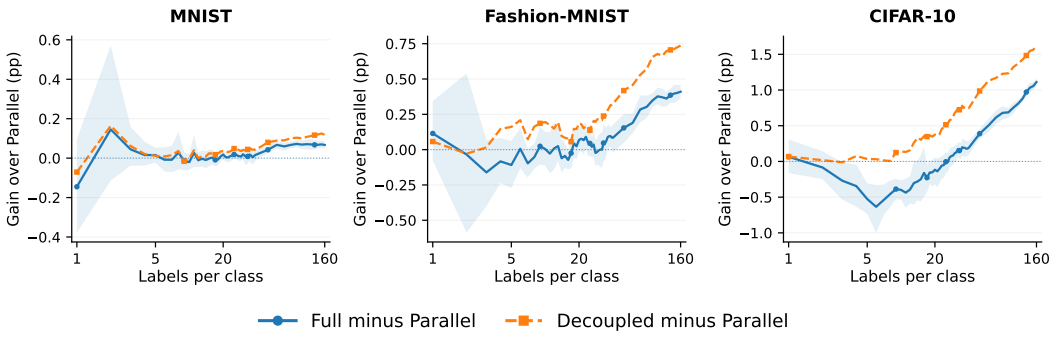


Figure 7: The zero-intercept complementary-block comparison over every annotation budget. Decoupled minus Parallel adds the separately fitted complementary block; Full minus Parallel also restores coupling. Each line is a ten-trajectory paired mean, with the Full–Parallel pointwise interval. The near-zero integrated gap can conceal a budget-dependent change of sign.

Table 5: Zero-intercept dense reference contrasts on ten annotation trajectories. Entries are normalized log-budget means in percentage points, with marginal paired 95% intervals. Parallel–Base can also introduce a constant direction. Figure 2 instead uses a common free intercept on the seven-budget panel; the two estimands must not be pooled.

Paired contrast	MNIST	Fashion-MNIST	CIFAR-10
Full − Base	$+0.68$ $[+0.54, +0.83]$	$+0.53$ $[+0.39, +0.65]$	$+1.61$ $[+1.27, +1.93]$
Parallel − Base	$+0.65$ $[+0.49, +0.83]$	$+0.47$ $[+0.35, +0.58]$	$+1.58$ $[+1.29, +1.86]$
Full − Parallel	$+0.03$ $[-0.03, +0.10]$	$+0.06$ $[-0.00, +0.13]$	$+0.03$ $[-0.05, +0.11]$
Decoupled − Full	$+0.03$ $[-0.00, +0.05]$	$+0.17$ $[+0.14, +0.21]$	$+0.44$ $[+0.38, +0.50]$

Table 6: Historical, ten-trajectory argmax controls. The dense row uses normalized log-budget area; the other rows average the seven compact budgets. Entries are paired mean differences and marginal 95% bootstrap intervals in percentage points. Different rows have different estimands.

Full minus control	MNIST	Fashion-MNIST	CIFAR-10
Base (dense)	$+0.68$ $[+0.54, +0.83]$	$+0.53$ $[+0.39, +0.65]$	$+1.61$ $[+1.27, +1.93]$
Base, common intercept	$+0.82$ $[+0.57, +1.15]$	$+0.66$ $[+0.57, +0.76]$	$+1.52$ $[+1.15, +1.85]$
BaseTrace	$+0.46$ $[+0.40, +0.53]$	$+0.48$ $[+0.42, +0.54]$	$+0.66$ $[+0.38, +0.93]$
Base, selected penalties	$+0.58$ $[+0.34, +0.86]$	$+0.62$ $[+0.40, +0.83]$	$+0.75$ $[+0.28, +1.17]$

Table 7: Full minus its exact scalar-matched linear comparator (pp). Each cell is the mean and observed range of three paired seeds. These are not confidence intervals. All graphs, budgets and geometries of the matched panel are retained.

Graph	b	$64, t_0$: mean [range]	$128, t_0/2$: mean [range]
MNIST	5	$+0.164$ $[-0.029, +0.413]$	$+0.027$ $[-0.040, +0.114]$
MNIST	40	$+0.278$ $[+0.244, +0.343]$	$+0.017$ $[-0.049, +0.105]$
MNIST	160	$+0.253$ $[+0.221, +0.304]$	$+0.139$ $[+0.056, +0.230]$
Fashion-MNIST	5	$+0.073$ $[-0.146, +0.269]$	$+0.196$ $[-0.237, +0.733]$
Fashion-MNIST	40	$+0.466$ $[+0.204, +0.631]$	-0.139 $[-0.220, -0.060]$
Fashion-MNIST	160	$+0.796$ $[+0.684, +0.930]$	$+0.217$ $[+0.102, +0.345]$
CIFAR-10	5	-0.012 $[-0.225, +0.130]$	-0.059 $[-0.494, +0.337]$
CIFAR-10	40	$+0.589$ $[+0.544, +0.622]$	$+0.125$ $[-0.262, +0.386]$
CIFAR-10	160	$+1.324$ $[+1.182, +1.464]$	$+1.047$ $[+0.868, +1.166]$

C.3 PATH COVERAGE BY GRAPH AND ANNOTATION BUDGET

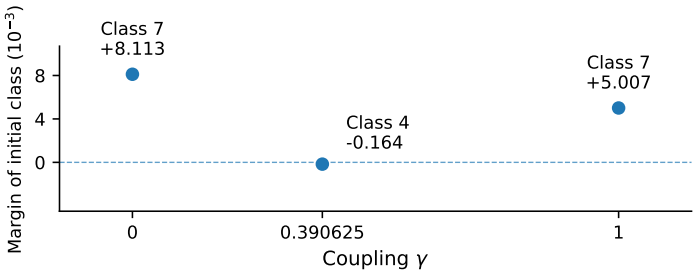


Figure 8: A real change-and-return example: CIFAR-10 query 38063, $b = 40$, seed 20270100, free intercept and fixed G64 penalties. Only three recorded initial-class margins are plotted; the negative value is a witness, not an interpolated minimum. Endpoint agreement does not establish stability on the path.

Table 8: Free-intercept whole-path diagnostics, on the same 12,288 query/configuration evaluations per row (4,096 queries on three supports). Coverage is numerical, not accuracy; witness counts come from negative grid margins. The other intercept convention is retained in the companion.

Graph	b	Norm (%)	Signed (%)	Adaptive (%)	Witnesses
MNIST	5	93.21	98.75	99.62	47
MNIST	40	73.79	98.55	99.48	64
MNIST	160	10.91	98.75	99.58	51
Fashion-MNIST	5	60.01	92.72	98.02	243
Fashion-MNIST	40	5.55	92.80	98.06	239
Fashion-MNIST	160	0.01	92.90	97.87	262
CIFAR-10	5	47.95	87.57	96.45	435
CIFAR-10	40	7.86	86.17	96.96	373
CIFAR-10	160	0.42	90.42	97.56	300

Adaptive coverage aggregated separately on each seed is 97.94%, 98.47% and 98.19%. These are three support trajectories, not independent query-level replications. Refinement adds 11,122 diagnoses beyond the direct signed test.

C.4 ENDPOINT ERRORS AND THE SCOPE OF THE QUERY DIAGNOSTIC

For each fixed support, changes partition into wrong-to-correct, correct-to-wrong and wrong-to-different-wrong. The last type changes a decision but not its correctness; unchanged decisions also contribute zero. Summing the correctness indicators gives Eq. (14). No independence assumption is used in that accounting identity.

Table 9: Endpoint correctness balance in the full-complement, free-intercept diagnostic: average percentage-point contributions over $b = 5, 40, 160$ within two seeds. Geometry, penalties and support are common to the two endpoints. This diagnostic panel is not pooled with the ten-seed attribution or the 4,096-query panel.

Graph	Errors corrected	Correct labels lost	Net Full − Decoupled
MNIST	+0.147	−0.173	−0.026
Fashion-MNIST	+0.621	−0.945	−0.324
CIFAR-10	+0.616	−1.041	−0.425

The transition counts come from the two-seed endpoint diagnostic, not from extrapolation of the query sample. Its 36 configurations retain both intercept conventions; their count-derived differences

reproduce the recorded endpoint accuracies. The direct comparison to the older fit register agrees for 71 of 72 endpoint values; one independently computed Decoupled endpoint differs by 15 correct decisions (0.0252 point, CIFAR-10/$b = 40$, seed 20270101, free intercept). We retain the diagnostic's paired counts together rather than mixing endpoints across producers. This does not change any historical observation.

The shared-query panel records 3,952 endpoint changes and 31 additional change-and-return witnesses, with 217,200 numerically stable evaluations and one unresolved case. All endpoint changes lie outside the stable set. However, its compact rows omit the true query class and the final winning class. Thus no cross-tabulation of repair/damage against adaptive status is claimed. The 31 returns diagnose a form of sensitivity that has exactly zero endpoint-accuracy contribution, regardless of the unknown true label. An exact per-query error attribution would require both endpoint winners and true classes for every graph, support, intercept and vertex; identities and margins alone do not supply these labels. No real-data join is reported.

C.5 Joint-selection results at every evaluated budget

Table 10 gives the absolute classification accuracies behind Figure 5. The companion retains individual trajectories from which three-budget means and dispersions can be recalculated; no final budget replaces that trajectory summary.

Table 10: All three budgets of the new selection panel, 3 paired seeds, mean argmax accuracy (%). LapRLS uses its expanded grid. No endpoint is used as a substitute for the trajectory summary. The maximum mean in each row is bold.

Graph	b	Base	Parallel	Full	Decoupled	LapRLS-L	LapRLS-Q
MNIST	5	95.13	94.76	95.43	95.15	94.40	**95.70**
MNIST	40	**97.06**	97.04	97.03	97.02	96.99	96.98
MNIST	160	97.23	97.24	97.36	**97.42**	97.21	97.35
Fashion-MNIST	5	72.15	72.79	71.87	72.09	73.14	**73.66**
Fashion-MNIST	40	79.77	79.58	79.36	80.05	79.55	**80.21**
Fashion-MNIST	160	83.37	83.29	83.50	**83.53**	80.72	83.06
CIFAR-10	5	56.06	55.22	**56.27**	55.53	54.93	53.89
CIFAR-10	40	**70.16**	69.55	69.94	69.54	68.10	68.27
CIFAR-10	160	73.01	73.13	73.98	**74.06**	69.22	72.39

C.6 Sensitivity of an auxiliary historical control

Three stored observations of the mean-scaled Decoupled control at $b = 5$ differ from its equivalent fixed-loss recipe: Fashion-MNIST seeds 20270102 and 20270105, and CIFAR-10 seed 20270104. Their argmax differences are $+0.2173$, -0.0486 and -0.3236 points, respectively. Controlled replays agree, but do not identify the cause of the historical differences. The original records remain unchanged. In the explicitly tabular equivalence scenario, substituting the corresponding fixed-recipe metrics moves affected ten-seed means by at most 0.0324 point; all 42 archived endpoint contrasts at $b = 160$ remain unchanged. Removing the affected common supports is a separate sensitivity calculation, not a rule for selecting a preferred ranking. These three mean-scaled observations supply neither the exact scalar-match panel nor the adaptive-path panel. The retained individual observations identify the three cases and allow both sensitivity calculations.

C.7 Extended LapRLS and replication scope

The old and extended observations, tied validation candidates and effective penalties are retained for each paired support.

The replication extension contains a complete ten-seed comparison for MNIST/$b = 5$ only; the other eight graph/budget cells and complete three-budget trajectories remain at three seeds. On that ten-seed cell, Full minus Base is -0.123 point with exploratory bootstrap interval $[-1.078, 0.536]$, rather than $+0.294$ on the three common seeds. No incomplete trajectory is filled or treated as a

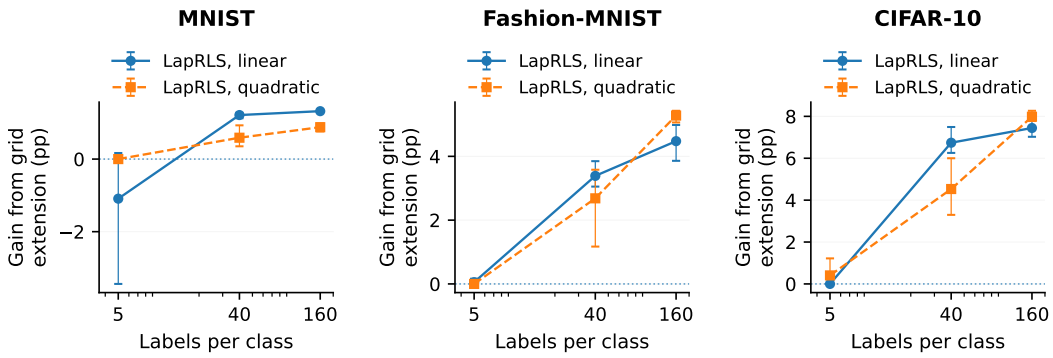


Figure 9: **Changing the allowed regularization range changes the comparison.** Extended minus original LapRLS accuracy on identical supports and folds. Points average three seeds; whiskers show their observed range. No selection uses evaluation labels.

ten-seed mean. The sensitivity uses the completed observations of the same selection protocol and does not change the separate angular or path panels.

C.8 Propagation, C&S and decoding

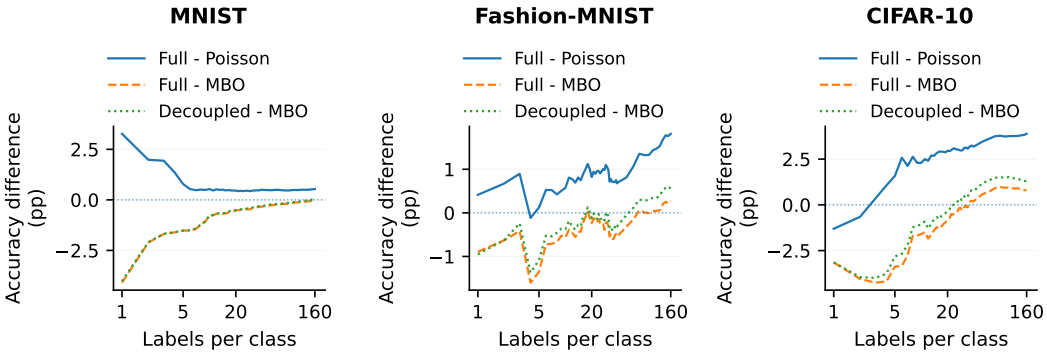


Figure 10: **Learning a feature rule and volume-informed propagation have different annotation regimes.** Dense, ten-seed paired contrasts against Poisson Learning and PoissonMBO, keeping all 52 budgets. PoissonMBO uses population proportions during fitting. These curves are not comparisons with the newly selected three-seed procedures.

The dense Full–Poisson integrated differences are $.99, .80, 2.05$ points. Against PoissonMBO they are $-1.12, -.48, -1.58$; initial losses and later crossings are preserved. For the separate seven-budget fixed-geometry selection study at ten seeds, Full minus C&S-MLP on CIFAR-10 is $+.02$ points with interval $[-.31, +.32]$. At its $b = 160$ endpoint, C&S-MLP reaches 73.35%, exceeding Full at 72.60%. The new selected values in Table 10 use different procedures and only three of these seeds; a comparison of those two published means would not be paired evidence.

For scores U on T, the oracle chooses assignments $\widehat{y}_i$ maximizing $\sum_{i \in T} U_{i,\widehat{y}_i}$ subject to $|\{i \in T : \widehat{y}_i = c\}| = q_c$, where q_c is the known remaining count of class c. Replacing U_{ic} by $aU_{ic} + r_i + h_c$, $a > 0$, multiplies the original objective by a and adds the same $\sum_{i \in T} r_i + \sum_c q_c h_c$ to every feasible assignment, so its optimizer set is unchanged. In contrast, class offsets may alter argmax. This invariance explains why readouts require separate reporting; it does not infer a cause of accuracy changes. All oracle selections use the same parameters chosen by observed-label argmax validation. No hidden evaluation label is used for choosing a parameter or breaking a tie.

C.9 Interventions at a common validation-selected setting

The 27 choices and all 108 final fits are retained separately from independently selected models. On CIFAR-10, the complementary increment is positive and the recoupling increment negative on each of the three complete trajectories. Fashion-MNIST instead has complementary increments $+0.064$, $+0.032$ and -0.607 point, so the fixed-setting pattern does not persist uniformly under this selection rule. Full–Base varies in sign across seeds on both graphs. The experiment tests one Base-selected linked-penalty rule on three development graphs; it does not establish robustness to other selectors or new graphs.

Table 11: Interventions at one Base-selected setting per support. Entries are mean accuracy differences (pp) over three paired trajectories, with their observed ranges, not confidence intervals. Each trajectory equally averages $b = 5, 40, 160$. Geometry, penalties and intercept are common to the four models. This three-budget estimand is distinct from the seven-budget fixed-setting attribution.

Paired contrast	MNIST	Fashion-MNIST	CIFAR-10
Parallel − Base	$+0.190$ $[+0.112, +0.334]$	$+0.043$ $[-0.503, +0.546]$	$+0.098$ $[-0.192, +0.370]$
Decoupled − Parallel	$+0.020$ $[+0.004, +0.034]$	-0.170 $[-0.607, +0.064]$	$+0.491$ $[+0.374, +0.590]$
Full − Decoupled	-0.008 $[-0.017, +0.008]$	-0.052 $[-0.248, +0.217]$	-0.231 $[-0.422, -0.085]$
Full − Base	$+0.201$ $[+0.110, +0.377]$	-0.180 $[-0.893, +0.452]$	$+0.359$ $[-0.104, +0.659]$