Nonlinear Label Budgets: How Geometry Determines Prediction Risk

Anonymous authors
Paper under double-blind review

Abstract

Nonlinear models can exploit low-dimensional structure, but this benefit depends on how observations are processed. We ask how many labels need nonlinear processing when an initial sample configures an affine predictor for the remaining labels. For a known compact smooth model manifold, we characterize the minimax risk over this complete class. The leading risk follows a staircase of derivative-space dimensions as the nonlinear budget changes; at the square-root boundary, curvature determines an exact continuous transition. We construct attaining readouts by correcting the mean displacement and shrinking normal uncertainty. A growing-dimensional construction shows that a vanishing pilot fraction can remove a diverging cost of fixed affine prediction. Allowing a quadratic readout removes the leading curvature penalty with a smaller sufficient pilot budget, while minimax laws quantify source uncertainty and prescribe shrinkage under model error. Conditional low-rank readouts turn these principles into a configuration method with an explicit gain criterion. Experiments test the risk laws and learned readouts; on 100 captured objects, configuration reduces MSE by a per-object average of 16.3% over matched fixed readouts.

1 Introduction

Low-dimensional structure can reduce the observations needed for prediction (Bickel & Li, 2007; Kpotufe, 2011). Nonlinear models exploit it by locating the target and adapting to nearby geometry. Must every response enter a nonlinear computation, or can a small initial sample configure a simple readout that retains this statistical benefit?

Acquisition pipelines can fit a model on a host and export coefficients for subsequent responses. Single-pixel imaging separates measurement from reconstruction (Duarte et al., 2008), with implementations using precomputed linear reconstruction and field-programmable gate array (FPGA) readouts (Czajkowski et al., 2018; Hoshi et al., 2022). We study the statistical value of this configuration: a predictor processes k labels arbitrarily, then fixes an affine map of the remaining $n-k$ labels before they arrive (Figure 1, left). Following two-stage estimation, we call the initial sample a *pilot* (Wang et al., 2020) and the resulting class *pilot-affine*. All n labels are scalar responses, including the pilot. We ask how optimal prediction risk changes with k.

One-step corrections yield first-order efficiency (Le Cam, 1956; van der Vaart, 1998); curvature and higher-order expansions describe remaining errors (Efron, 1975; Robins et al., 2017). A smaller pilot presents a different difficulty: uncertainty about the target's location leaves higher-order errors that may remain visible at the final prediction scale. A readout must decide which directions to retain and which to suppress, balancing geometric error against noise. To identify the optimal tradeoff, we derive matching upper and lower bounds over all pilot-affine predictors, allowing the pilot to choose both the matrix and the intercept.

The key is to compare geometric uncertainty with the noise in the main observations. A pilot leaves tangent-location error of order $k^{-1/2}$, whereas curvature produces normal displacement of order k^{-1}. Higher derivatives produce further uncertainty scales, yielding a staircase of risks indexed by derivative-space dimensions as k changes. At $k \asymp \sqrt{n}$, curvature displacement and main noise have the same scale. The optimal affine readout then corrects the mean displacement and shrinks the remaining normal uncertainty according to curvature. A quadratic readout can reconstruct that

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.

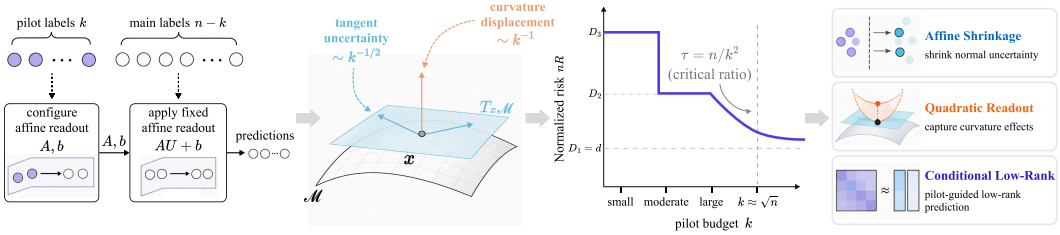

Figure 1: **From nonlinear label budgets to geometric prediction risk.** From left to right: pilot labels fix the affine coefficients before main-block access; tangent uncertainty and curvature displacement scale as $k^{-1/2}$ and k^{-1}; the schematic risk staircase descends through osculating dimensions D_j, with a curvature-dependent transition near $k \asymp \sqrt{n}$; affine shrinkage, quadratic readouts and conditional low-rank prediction exploit this geometry.

displacement, explaining how a richer operation recovers intrinsic-dimensional risk with a smaller sufficient pilot budget.

These laws also identify what a readout must learn and preserve. Source estimation introduces uncertainty in the geometry, while model error adds normal directions that the readout must retain. We characterize both costs, then express configuration through conditional moments for general observation designs. Controlled simulations test the gain criterion, and captured views evaluate configuration at a fixed readout capacity.

Our main contributions are:

1. **Sharp geometric budget laws.** We establish matching minimax bounds for pilot-affine prediction, including a derivative-dimension staircase and an exact curvature transition. A growing-dimensional construction establishes a diverging advantage over fixed affine prediction.
2. **Quadratic recovery and uncertainty correction.** We construct quadratic readouts that recover intrinsic-dimensional risk with a smaller sufficient pilot budget. We also derive minimax laws and attaining readouts for learned geometry and normal model error.
3. **Provable gains from conditional configuration.** We derive a gain criterion for low-rank readouts with estimated moments and certify a finite-budget readout for ambiguous pilots. Simulations and captured-view experiments test the risk laws and configuration gains.

2 Related work

Two-stage estimation and geometric risk. One-step estimation and statistical curvature connect preliminary fitting to higher-order risk (Le Cam, 1956; Efron, 1975; Robins et al., 2017). Su & Wang (2025, Theorem 1) derive the distribution of a specified subsampled one-step estimator, including its square-root boundary. Our bounds optimize all ambient readouts affine in the main labels after arbitrary pilot processing, a different class from affine parameter updates followed by nonlinear embedding. Linear minimax theory and optimal recovery relate affine risk to parameter-set geometry (Donoho et al., 1990; Donoho, 1994). Butler & Levit (2007, Theorem 4.1) derive geometric corrections to small-noise Bayes risk on manifolds. In our two-block experiment, pilot uncertainty brings curvature into the leading minimax constant, yielding a budget-dependent shrinkage law.

Learned geometry. Manifold estimation studies recovery of a model and its local geometry from data (Genovese et al., 2012; Fefferman et al., 2018; Aamari & Levrard, 2019; Aizenbud & Sober, 2025). Representation learning connects source tasks to target sample complexity (Du et al., 2021; Tripuraneni et al., 2021; Zhang et al., 2022; Tinati & Tu, 2026). Our learned-feature law connects source prediction leverage, central to optimal design (Kiefer & Wolfowitz, 1960), to target minimax risk.

Conditional prediction. Neural processes use context observations to construct predictors (Garnelo et al., 2018; Kim et al., 2019; Gordon et al., 2020); ALPaCA learns features and priors for online

Bayesian linear regression (Harrison et al., 2018). We study the risk gained by configuring the main-data matrix, with rank controlling readout capacity. Reduced-rank regression provides its algebraic foundation (Izenman, 1975; Bunea et al., 2011); our conditional gain criterion accounts for the error in estimating the moments that configure this matrix.

Measurement and readout constraints. Random projections preserve manifold distances (Baraniuk & Wakin, 2009); adaptive sensing selects measurements using earlier responses (Arias-Castro et al., 2013). Our observation design is fixed, while the pilot selects the subsequent affine predictor. We characterize prediction risk as a function of this nonlinear-access budget.

3 Geometric laws for nonlinear label budgets

Let $M \subset \mathbb{R}^N$ be a fixed known compact connected smooth embedded manifold without boundary, with dimension $d \geq 1$ (Lee, 2013). We estimate an unknown $v \in M$ from independent blocks

$$Z_0 = v + k^{-1/2}\varepsilon_0, \qquad Z_1 = v + (n-k)^{-1/2}\varepsilon_1, \qquad \varepsilon_i \sim N(0, I_N).$$

We call an estimator *pilot-affine* if it has the form $A(Z_0, \omega)Z_1 + b(Z_0, \omega)$, where A, b are arbitrary Borel functions and ω is an independent random seed. The estimator may process the pilot nonlinearly; the main labels enter affinely. Write $R_M(k, n-k)$ for the infimum of worst-case squared Euclidean risk over this entire class. These blocks are sufficient statistics for actual scalar labels $y_i = X_i v + \xi_i$ with independent standard Gaussian noise and fixed $X_0^T X_0 = kI_N$, $X_1^T X_1 = (n-k)I_N$. Under an isotropic test covariate, excess prediction loss is exactly coefficient loss.

From pilot uncertainty to risk. Both risk laws follow from conditional optimization. For a prior π on M, let $C_k = \mathrm{Cov}(V \mid Z_0)$ and $r = n - k$. The optimal pilot-affine Bayes risk is

$$\mathbb{E}_\pi \operatorname{tr}\{C_k(I + rC_k)^{-1}\} = \mathbb{E}_\pi \sum_{j=1}^{N} \frac{\lambda_j(C_k)}{1 + r\lambda_j(C_k)}. \tag{1}$$

A direction whose pilot uncertainty is asymptotically larger than $1/r$ costs approximately $1/r$ in risk; uncertainty of smaller order contributes $o(1/r)$. Local priors turn this identity into minimax lower bounds. Geometry determines which covariance directions cross the threshold, and the attaining readouts retain those directions.

The relevant dimension counts the directions exposed by local derivatives. For a chart h_x centered at x, define

$$V_j(x) = \operatorname{span}\{\partial^\alpha h_x(0) : 1 \leq |\alpha| \leq j\}, \qquad D_j = \max_{x \in M} \dim V_j(x).$$

The chain rule makes V_j independent of the chart. These are osculating spaces; $D_1 = d$, and higher ranks may vary across M.

Theorem 3.1 (Geometric cost of a nonlinear pilot)**.** *Fix M, N and an integer $q \geq 1$. If $k \to \infty$, $k^q/n \to 0$ and $n/k^{q+1} \to 0$, then*

$$nR_M(k, n-k) \longrightarrow D_q. \tag{2}$$

Thus $k = \lfloor n^\beta \rfloor$, $0 < \beta < 1$ and $1/\beta \notin \mathbb{N}$ give $D_{\lfloor 1/\beta \rfloor}$. The theorem allows nonconstant osculating ranks and requires no group action or analyticity.

The upper bound is attained by a geometric projection. Fit a nearest point $\widetilde{v} \in M$ using Z_0, let $P_q(x)$ project onto $V_q(x)$, and use

$$\widehat{v}_q = \widetilde{v} + P_q(\widetilde{v})(Z_1 - \widetilde{v}). \tag{3}$$

The risk consists of variance at most $D_q/(n-k)$ and residual squared bias of order $k^{-(q+1)}$, with an exponentially small localization term. A local posterior lower bound forces every pilot-affine rule to pay for the same D_q directions.

The one-step case $q = 1$ recovers tangent correction (Le Cam, 1956; Simpson et al., 1992). For a concrete higher-order example, take the closed curve $h(t) = (\cos t, \sin t, \cos 2t, \sin 2t)$. Its first three derivatives are linearly independent at every phase, so $(D_1, D_2, D_3) = (1, 2, 3)$. With $k = \lfloor n^\beta \rfloor$, the optimal leading risks are therefore $1, 2, 3$ on $\beta \in (1/2, 1), (1/3, 1/2), (1/4, 1/3)$, respectively. The same one-dimensional model thus exposes three risk levels as nonlinear access decreases.

A growing-dimensional consequence. The budget advantage can grow without bound. For dense rotation templates of growing ambient dimension, both predictors receive the same noisy source calibration. With $k = \lfloor n^\beta \rfloor$ and $1/2 < \beta < 1$, the pilot-affine normalized risk tends to 3, while the source-configured fixed-affine normalized risk grows as Q^2, where $Q \log Q = \log n$. Thus a vanishing pilot fraction removes a diverging cost on the same model sequence. Appendix D.5 specifies the sequence and proves the uniform bounds.

3.1 Curvature shrinkage at the square-root boundary

Between the first two steps, curvature uncertainty is comparable to main-block noise, so the threshold in (1) becomes a continuous tradeoff. At $n/k^2 \to \tau \in (0, \infty)$, tangent projection leaves normal bias at this same scale. Projecting onto the entire second-order space removes that bias but pays full variance in each normal direction. The optimal update instead uses the magnitude of curvature to balance the two terms.

Let P_x project onto T_xM, and define the second fundamental form by the normal graph

$$h_x(u) = x + u + \tfrac{1}{2}\mathrm{II}_x(u,u) + O(\|u\|^3).$$

For any orthonormal tangent basis $e_1, \dots, e_d$, set

$$H_x = \sum_a \mathrm{II}_x(e_a, e_a), \qquad K_x = \tfrac{1}{2}\sum_{a,b} \mathrm{II}_x(e_a, e_b)\mathrm{II}_x(e_a, e_b)^T. \tag{4}$$

Here H_x is the unnormalized mean-curvature vector and K_x is a positive semidefinite ambient operator supported on the normal space. Both are basis independent. If G is a standard tangent Gaussian, Wick's formula (Isserlis, 1918) gives $\mathbb{E}[\mathrm{II}_x(G,G)] = H_x$ and $\mathrm{Cov}(\mathrm{II}_x(G,G)/2) = K_x$.

Theorem 3.2 (Critical risk and curvature shrinkage)**.** *Fix the manifold and ambient dimension of Section 3. For any fixed $\tau \in (0, \infty)$, as $k \to \infty$ and $n/k^2 \to \tau$,*

$$nR_M(k, n-k) \longrightarrow F_M(\tau) := d + \max_{x \in M} \mathrm{tr}\{\tau K_x (I + \tau K_x)^{-1}\}. \tag{5}$$

Let p be a Borel nearest point to Z_0, and define

$$\begin{aligned} &\mu_p = p + H_p/(2k), \qquad t = (n-k)/k^2, \\ &B_p = tK_p(I + tK_p)^{-1}, \qquad \widehat{v} = \mu_p + (P_p + B_p)(Z_1 - \mu_p). \end{aligned} \tag{6}$$

Uniformly in $v \in M$, this rule has normalized risk $d + \mathrm{tr}\{\tau K_v (I + \tau K_v)^{-1}\} + o(1)$ and hence attains the minimax limit. The theorem allows the rank of K_x to vary.

Mean correction centers the pilot's quadratic normal error. The tangent covariance has scale k^{-1} and the normal covariance has scale $k^{-2}K_p$. Substitution into (1) yields the critical trace. The resolvent stays smooth at zero curvature. The critical-limit function joins the first two steps: $F_M(\tau) \to d$ as $\tau \downarrow 0$, and $F_M(\tau) \to D_2$ as $\tau \uparrow \infty$.

3.2 Quadratic recovery of intrinsic risk

The critical law identifies the cost of keeping the main readout affine. A quadratic readout uses the main observations to reconstruct the normal displacement responsible for that cost. Set $r = n - k$. At a pilot projection p, write $u = P_p(Z_1 - p)$ and

$$q = p + u + \tfrac{1}{2}\mathrm{II}_p(u,u) - H_p/(2r). \tag{7}$$

For a known fixed manifold, its risk is at most $d/r + Ck^{-3} + C/(rk) + C/r^2 + Ce^{-ck}$. Thus $k/n \to 0$ and $n/k^3 \to 0$ suffice to recover intrinsic-dimensional leading risk uniformly. Appendix A proves the staircase, critical and quadratic risk laws.

4 Prediction with learned and uncertain geometry

Quadratic recovery uses the local geometry to remove normal displacement. When that geometry is estimated or the target departs from it, normal components also carry uncertainty that prediction must preserve. We characterize the resulting risk through source precision and model error.

4.1 Geometry estimated from source data

To quantify how source estimation affects quadratic recovery, let $x = Bf$, where f belongs to a known compact smooth embedded boundaryless d-manifold $\mathcal{F} \subset \mathbb{R}^D$. We take $\mathbb{R}^D = \mathrm{span}(\mathcal{F})$, and express the source design identity below in these effective feature coordinates. The entire $N \times D$ matrix B is unknown. Independent source responses $Y_j = Bf_j + \sigma_s \epsilon_j$, $\epsilon_j \sim N(0, I_N)$ independently, with known $m^{-1} \sum_j f_j f_j^T = I_D$, give $W = B + \sigma_s E/\sqrt{m}$ with independent standard Gaussian entries in E.

For prediction, let $\widehat{f}$ minimize $\|Z_0 - Wf\|$ over $\mathcal{F}$. Compute (7) on $W\mathcal{F}$ at $p = W\widehat{f}$, and return $(1-w)q + wZ_1$, with $w = r\widehat{\nu}/(1+r\widehat{\nu})$ and $\widehat{\nu} = \sigma_s^2 \|\widehat{f}\|^2/m$. Apply this rule when $\|W\|_F \le 2L$ and $\sigma_{\min}(W) \ge s/2$, and output Z_1 otherwise; L, s are the class bounds below.

Theorem 4.1 (Learned-feature minimax risk). *Fix $N \ge D$, $s, \sigma_s > 0$, $L > s\sqrt{D}$, and $\mathcal{K} = \{B : \|B\|_F \le L, \sigma_{\min}(B) \ge s\}$. If $k/n \to 0$, $n/k^3 \to 0$, and $n\sigma_s^2/m \to \gamma < \infty$, the guarded rule has uniform normalized risk*

$$d + (N-d)\frac{\gamma \|f\|^2}{1+\gamma\|f\|^2} + o(1).$$

Replacing $\|f\|^2$ by $\max_{f\in\mathcal{F}} \|f\|^2$ gives the asymptotic minimax constant over all estimators using the same source sample and this task's target blocks.

A finite feature net with covering radius $O(k^{-1/2})$ preserves the limit, removing continuous pilot optimization. For a fixed source Gram matrix $\Sigma \succ 0$, replace $\|f\|^2$ by $f^T\Sigma^{-1}f$. Thus source prediction leverage determines the cost of learning geometry: uncertainty in poorly measured source directions propagates to target risk.

4.2 Normal model error

For targets $x + u/\sqrt{n}$ with $x \in M$, $u \perp T_x M$ and $\|u\| \le \varrho$, normal model error remains visible at the prediction scale. Let $\lambda_j(x)$ be the normal eigenvalues of K_x, including zeros. At $n/k^2 \to \tau$, replace each critical shrinkage weight by

$$b_j^*(x) = \frac{\max\{\tau\lambda_j(x), c_x\}}{1 + \max\{\tau\lambda_j(x), c_x\}}, \qquad \sum_j (c_x - \tau\lambda_j(x))_+ = \varrho^2. \tag{8}$$

For $\varrho = 0$, take $c_x = 0$. The pilot-affine minimax constant is $d + \max_x \sum_j b_j^*(x)$, attained by substituting these weights into (6). The weight floor preserves main-label information even in directions with zero curvature. Appendix B proves the learned-geometry and normal-error minimax laws.

5 Configuring readouts from conditional moments

The geometric laws arise by optimizing a readout conditional on the pilot. The same principle configures prediction under general observation designs: a source model supplies moments of pilot observations Z, main inputs U and targets X, while a rank constraint controls readout capacity. Write their conditional means as $m_X(Z), m_U(Z)$, and set $C(Z) = \mathrm{Cov}(X, U \mid Z)$ and $S(Z) = \mathrm{Cov}(U \mid Z) \succ 0$. Conditional reduced-rank regression (Izenman, 1975) gives

$$\widehat{X} = m_X(Z) + B_R(Z)\{U - m_U(Z)\}, \qquad B_R(Z) = [C(Z)S(Z)^{-1/2}]_R S(Z)^{-1/2}, \tag{9}$$

Here R is the rank cap, and $[\cdot]_R$ retains the leading R singular components (Eckart & Young, 1936). After offline source fitting, the pilot computes the moments and exports the map before main acquisition.

Configuration improves risk by allowing the matrix to depend on the pilot. Write $T = CS^{-1/2}$ and $\bar{T} = (\mathbb{E}C)(\mathbb{E}S)^{-1/2}$. Relative to the best fixed rank-R matrix with an arbitrary nonlinear pilot intercept, the oracle average-risk improvement under the joint distribution of (Z, U, X) is

$$G_R = \mathbb{E}\|[T(Z)]_R\|_F^2 - \|[\bar{T}]_R\|_F^2 \ge 0. \tag{10}$$

Moment-estimation error determines whether this gain is realized. With exact conditional means and S, let $\widehat{T}$ be estimated from independent source data, and define $\delta = \mathbb{E}\|\widehat{T} - T\|_F^2$ and $\mathcal{E}_R =$

$\mathbb{E}\|T - [T]_R\|_F^2$. Using $\widehat{B} = [\widehat{T}]_R S^{-1/2}$ adds at most $4\sqrt{\mathcal{E}_R\delta} + 4\delta$ to the optimal configured risk. Hence $4\sqrt{\mathcal{E}_R\delta}+4\delta < G_R$ guarantees improvement over every fixed rank-R matrix. When T already has rank at most R, the sufficient condition reduces to $4\delta < G_R$. Appendix C proves the identity and extends the guarantee to jointly estimated means and covariances.

In a controlled model with nonzero rank-truncation residual, 128 source samples per pilot state yield a guaranteed risk reduction of .1493 and a measured reduction of .5104 $\pm$ .0010 (SE) over the population-optimal fixed rank-one matrix (Appendix F.5). Gains already occur below the sufficient threshold, illustrating its conservative source-accuracy requirement.

When a pilot leaves several plausible locations, the conditional moments average over them. For the nonconvex-curve experiment, we discretize pilot observations into cells and store one affine rule per cell; interval bounds certify its continuous worst-case risk (Appendix E).

6 Experiments

CPU simulations test the Gaussian risk laws, while held-out captured views test configuration gains at fixed capacity using sparse pixels and source-fitted conditional moments. On the harmonic curve, Figure 2a shows the predicted risk staircase. At $k = \lfloor n^{.4} \rfloor$, affine and quadratic risks approach 2 and 1. Uncertainty uses independent pilots or source cohorts in simulations, and objects or sequences in reconstruction. Appendices F–G collect the protocols, controls and implementation details.

6.1 Curvature and the critical risk

To test curvature-dependent shrinkage, we vary both curvature rank and strength using the unit circle, the nonconvex plane curve $h(t) = ((1+.3\cos 2t)\cos t, (1+.3\cos 2t)\sin t)$, and $S^1(1) \times S^1(1/2) \subset \mathbb{R}^4$. The curve has four curvature-zero phases; the torus has two normal curvature eigenvalues, $1/2$ and 2. We use $n = 1024, 4096, 16384$ and $k/\sqrt{n} = 1/2, 1, 2$, with 32 independent pilots per target. Figure 2b–c tests budget and phase dependence; Table 1 compares seven rules on shared labels. Entries are normalized conditional risks averaged equally over fixed phases, with pilot-sampling standard errors. Posterior calculations use a uniform prior on each angle.

At $n = 16384, k = 128$, torus curvature risk $2.986 \pm .047$ agrees with the limit 3.000; removing either correction or shrinkage increases risk (Table 1). The curve phase average is $1.312 \pm .041$, versus its critical prediction 1.273. With 2,048 pilots per target, independent confirmation places the curve's grid peak at 1.586 and 1.546 for $n = 16384$ and 262144, approaching the predicted 1.522. The curvature rule improves on raw tangent projection in all 27 conditional phase averages and on full second-order projection in 26. At $n = 1024, k = 16$ on the nonconvex curve, localization errors raise its mean risk to 4.995, versus 2.032 for full second-order projection. Global conditional moments retain these competing locations instead of committing to one local fit.

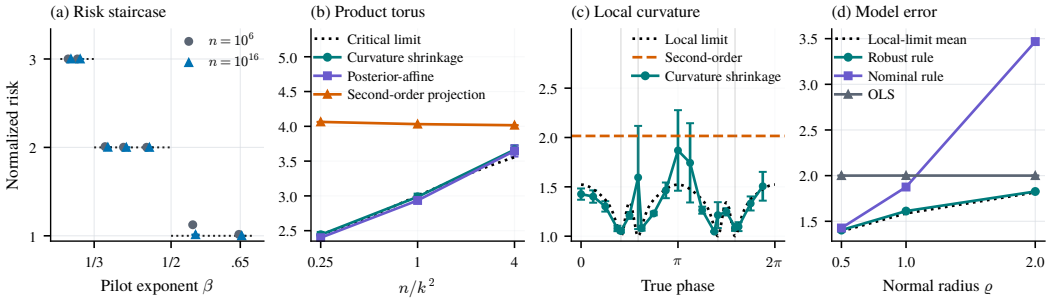

Figure 2: **Geometric prediction costs.** (a) Harmonic-curve risk at noncritical pilot exponents; bars show 95% intervals, with horizontal offsets $\pm.003$ for visibility. (b–d) Torus budget dependence, local curvature and normal model error at $n = 16384$ ($k = 128$ in c–d); bars show SEs. Dotted lines give theoretical limits, averaged over offsets in (d); gray lines mark curvature zeros.

Table 1: Component ablations at $n = 16384, k = 128$ (mean $\pm$ SE).

Rule	Circle	Curve	Torus
Raw tangent	$1.881 \pm .151$	$1.751 \pm .146$	$5.653 \pm .418$
Corrected tangent	$1.614 \pm .123$	$1.539 \pm .116$	$4.383 \pm .311$
Shrinkage without correction	$1.509 \pm .068$	$1.391 \pm .050$	$3.211 \pm .063$
Curvature shrinkage	$1.389 \pm .055$	$1.312 \pm .041$	$2.986 \pm .047$
Full second-order	$2.016 \pm .000$	$2.016 \pm .000$	$4.031 \pm .000$
Global posterior-affine	$1.382 \pm .055$	$1.285 \pm .031$	$2.934 \pm .038$
All-label OLS	$2.001 \pm .001$	$1.999 \pm .001$	$4.000 \pm .001$

Global readouts for ambiguous pilots. For this nonconvex curve at $n = 1024, k = 16$, retaining alternative locations yields a 10,404-cell rule with certified continuous worst normalized risk at most 1.81, below all-label OLS (2). Independent confirmation over 64 phases with 8,192 Gaussian pairs gives conditional risk $1.711 \pm .010$ and actual loss $1.736 \pm .022$, versus local curvature risk $2.392 \pm .296$.

6.2 Quadratic recovery with learned geometry

To measure how estimating geometry affects quadratic recovery, we estimate all 40 coefficients of an eight-dimensional curve $h(t) = B\phi(t)$, with known Fourier features through order two, from independent source data of noise standard deviation .02. Two shapes have relative second-harmonic amplitudes $b = 1, .3$. Each source cohort costs $8m$ scalar responses; target observations use Gaussian sufficient statistics with $k = \sqrt{n}$. The affine baseline corrects mean curvature and shrinks normal uncertainty from both pilot localization and source estimation. The quadratic rule uses the source-aware mixture from Section 4.1; joint nonlinear prediction fits all target observations while profiling the same source uncertainty. All three use the same data and numerical phase fitting.

Table 2: Prediction risk with learned geometry at $b = 1$.

n	m	Affine	Quadratic	Joint	Limit
4096	128	1.677	1.431	1.407	1.421
4096	512	1.368	1.123	1.089	1.110
4096	2048	1.277	1.012	0.991	1.028
16384	128	2.573	2.397	2.383	2.427
16384	512	1.652	1.415	1.399	1.421
16384	2048	1.343	1.105	1.094	1.110

Theorem 4.1 predicts $1 + 7a/(1 + a)$ with $a = 5n(.02)^2/m$ for both shapes. Table 2 reports the corresponding n-normalized risks at $b = 1$, averaged over 128 independent source cohorts with 16 targets each. Increasing source precision reduces the predicted geometry-estimation cost, and quadratic risk tracks this change toward the intrinsic-dimensional limit. Across all 12 cells, source-cohort paired 95% intervals favor quadratic over affine and joint over quadratic. Complementing these learned-geometry results, independent matched-input confirmation on four known manifolds favors quadratic over affine in all 16 contrasts at $k = \sqrt{n}$, with simultaneous intervals below zero. At $k = \sqrt{n}/4$, three of eight settings reverse in mean.

6.3 Preserving information under model error

We test normal-uncertainty correction on the circle, curve and torus with $n = 4096, 16384$, $k = \sqrt{n}$, and radii $\varrho = .5, 1, 2$. Five rules share 16,896 pilot/main pairs across 18 configurations. The robust rule receives the radius, but not the displacement direction, and uses (8) without target-loss tuning. At $\varrho = 2$, the curve's robust conditional risk is $1.826 \pm .003$, versus nominal $3.468 \pm .033$ and all-label OLS 2.000; its actual loss is $1.889 \pm .051$. Figure 2d shows how the robust rule retains information as model error increases. Robust conditional averages are lower than nominal averages across all 18 cells.

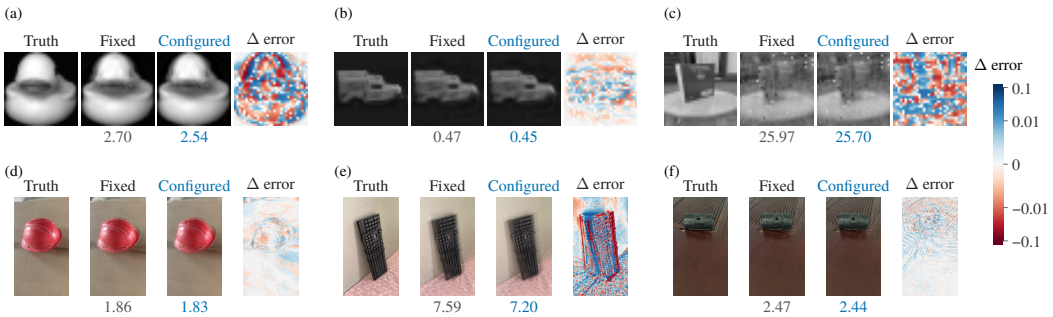

Figure 3: **Matched captured-view readouts.** (a–c) COIL-20, COIL-100 and CO3D; (d–f) MVImgNet RGB at long edge/observed locations 256/80, 512/80 and 512/320. Pilot fraction is $1/5$ and rank is two; examples use median gains at object/sequence and frame levels. Numbers give MSE ($\times 10^3$). Error differences use a shared asinh color scale: blue denotes improvement, red degradation. RGB channels are averaged; observed pixels are masked. Predictions are unclipped for all metrics.

Table 3: Captured-view reconstruction with matched rank-two readouts.

Collection	Mean MSE ($\times 10^3$) Configured	Fixed	Over fixed matrix Gain (%) [95%]	Wins
COIL-20	3.280	3.777	14.83 [10.54,19.10]	19/20
COIL-100	1.568	1.876	16.30 [13.81,18.66]	94/100
CO3D	9.125	9.526	4.38 [3.20,5.63]	16/16
MVImgNet RGB, 256/80	8.723	9.191	5.97 [4.30,7.73]	24/24
MVImgNet RGB, 512/80	10.086	10.602	5.45 [4.22,6.74]	24/24
MVImgNet RGB, 512/320	9.923	10.237	2.89 [1.95,3.95]	22/24

6.4 Configuration gains on captured views

Task and source models. Sparse-pixel image completion (Garnelo et al., 2018) tests configuration when source views provide a model for the unobserved pixels. We begin with 32×32 grayscale views: 16 pilot and 64 main pixels predict 944 targets using three shared masks and rank-two readouts. Object or sequence identity is known; target pose and frame position are hidden. COIL-20/100 provide rotating-platform views (Nene et al., 1996a;b), with 36 source, 18 validation and 18 test views per object. Fourier models and variance multipliers selected on COIL-20 transfer to COIL-100. Figure 3 shows representative reconstructions.

CO3D and MVImgNet extend the task to natural captures (Reizenstein et al., 2021; Yu et al., 2023). Settings chosen on development categories transfer to eight other categories per dataset: 16 CO3D and 24 MVImgNet sequences, with 216 and 169 test frames. Source-fitted cosine models with correlated residuals supply conditional moments. CO3D uses the grayscale protocol; MVImgNet RGB frames preserve their aspect ratios at long edges 256 and 512. Both RGB resolutions use 80 locations, each providing three scalars; 320 locations at long edge 512 approximately preserve the 256-setting's density. Matched comparisons allocate one fifth to the pilot.

Matched comparisons. The fixed control shares the source model, rank and nonlinear pilot means, so the comparison isolates how configuration improves the use of main observations. Table 3 averages frames and masks within objects/sequences, then weights these units equally. Gains average within-unit percentage reductions, and wins count improved units. The 95% bootstrap intervals resample whole units (Efron, 1979). Increasing capacity further reduces error: at 512/80, unrestricted linear prediction reaches 9.464×10^{-3} MSE versus 10.086×10^{-3} for rank-two configuration.

Configuration and budget allocation. Figure 4 varies the pilot/main split over 80 fixed grayscale locations at rank two. Source-selected 64/16 improves on 16/64 by 4.90%, 7.35% and 2.23% across the three datasets, with bootstrap intervals above zero. As the pilot explains more uncertainty, total risk can fall while the additional gain from matrix adaptation shrinks. RGB budget selection

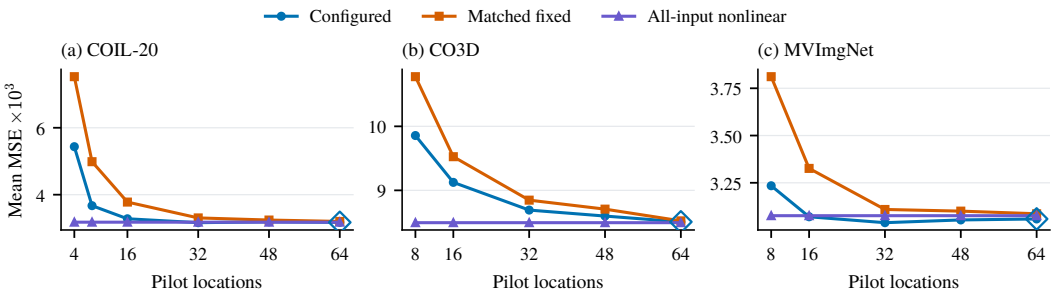

Figure 4: **Allocating a fixed observation budget.** Splits share 80 grayscale locations and 944 targets. Fixed and configured matrices share rank two and pilot means; the nonlinear reference uses all observations. Curves are object/sequence means; diamonds mark source-risk selections.

calibrates source-risk differences on development sequences; error at most ε guarantees excess risk at most 2ε within the candidate family (Proposition C.2). RGB calibration mainly corrects shared ranking bias: a post-hoc development-fixed split matches five of six setting–rank pairs and slightly improves the sixth.

Readout implementation. For COIL, configuration uses two main-stage accumulators and multiply–accumulates per pixel, versus 18 for nonlinear prediction. Both readouts upload 18 scalars; configuration additionally requires 128 downloaded coefficients and host feedback after 16 pilot pixels. CPU timings favor nonlinear prediction.

7 Discussion and conclusion

Intrinsic dimension alone does not determine prediction cost under limited nonlinear access. Uncertainty about the target's location extends into higher-order geometric directions, whose estimation can raise risk even on a low-dimensional manifold. The budget laws explain when a configured readout can suppress this uncertainty and when it must retain additional directions. A larger pilot reduces localization uncertainty; a quadratic readout reconstructs the leading curvature displacement from main observations; more precise source data reduce uncertainty in the geometry itself. Model error, in turn, requires preserving information beyond the fitted manifold. These mechanisms explain why pilot size, readout capacity and source precision should be considered together.

For constrained readouts, the gain criterion specifies when source accuracy suffices for useful adaptation. Controlled experiments test this guarantee against the optimal fixed matrix; captured views evaluate gains at matched observations and rank. Budget experiments show why these gains must be assessed alongside total risk: a larger pilot can improve accuracy while leaving less uncertainty for matrix adaptation to resolve. Thus total risk guides budget selection, while configuration gain measures the value of adaptation.

A small pilot can therefore configure a predictor that retains intrinsic-dimensional leading risk while processing the remaining labels affinely. Our sharp risk laws characterize the geometric conditions and budgets that make this possible, and richer readouts extend the attainable tradeoffs. Nonlinear access is thus a statistical resource whose value can be quantified through geometry.

AI USE STATEMENT

In this work, we used generative AI tools to assist with research ideation or execution and proving mathematical claims. Additionally, we used these tools to aid or polish writing and to retrieve and discover related literature. We have reviewed all AI-assisted work, including the manuscript text, references, mathematical proofs and experimental results. The authors are responsible for the final manuscript and all AI-assisted content, claims and artifacts.

REPRODUCIBILITY STATEMENT

Appendices A–E provide the assumptions, proofs and risk-certification procedure. Appendix F specifies the simulation models, random seeds, replication counts and uncertainty estimates. Appendix G documents the captured-view data splits, preprocessing, source-model fitting, readout selection and evaluation protocols, together with implementation details and resource accounting in Appendix G.5.

REFERENCES

Eddie Aamari and Clément Levrard. Nonasymptotic rates for manifold, tangent space and curvature estimation. *The Annals of Statistics*, 47(1):177–204, 2019. doi: 10.1214/18-AOS1685. URL https://arxiv.org/abs/1705.00989.

Yariv Aizenbud and Barak Sober. Estimation of local geometric structure on manifolds from noisy data. *Journal of Machine Learning Research*, 26(64):1–89, 2025. URL https://www.jmlr.org/papers/v26/25-0183.html.

Ery Arias-Castro, Emmanuel J. Candès, and Mark A. Davenport. On the fundamental limits of adaptive sensing. *IEEE Transactions on Information Theory*, 59(1):472–481, 2013. doi: 10.1109/TIT.2012.2215837. URL https://arxiv.org/abs/1111.4646.

Richard G. Baraniuk and Michael B. Wakin. Random projections of smooth manifolds. *Foundations of Computational Mathematics*, 9(1):51–77, 2009. doi: 10.1007/s10208-007-9011-z.

Peter J. Bickel and Bo Li. Local polynomial regression on unknown manifolds. In *Complex Datasets and Inverse Problems: Tomography, Networks and Beyond*, volume 54 of *IMS Lecture Notes–Monograph Series*, pp. 177–186. Institute of Mathematical Statistics, 2007. URL https://arxiv.org/abs/0708.0983.

Florentina Bunea, Yiyuan She, and Marten H. Wegkamp. Optimal selection of reduced rank estimators of high-dimensional matrices. *The Annals of Statistics*, 39(2):1282–1309, 2011. doi: 10.1214/11-AOS876. URL https://arxiv.org/abs/1004.2995.

Leo T. Butler and Boris Levit. A bayesian approach to the estimation of maps between riemannian manifolds. *Mathematical Methods of Statistics*, 16(4):281–297, 2007. doi: 10.3103/S1066530707040011. URL https://arxiv.org/abs/0705.2540.

Krzysztof M. Czajkowski, Anna Pastuszczak, and Rafał Kotyński. Real-time single-pixel video imaging with fourier domain regularization. *Optics Express*, 26(16):20009–20022, 2018. doi: 10.1364/OE.26.020009.

David L. Donoho. Statistical estimation and optimal recovery. *The Annals of Statistics*, 22(1):238–270, 1994. doi: 10.1214/aos/1176325367. URL https://web.stanford.edu/dept/statistics/cgi-bin/donoho/wp-content/uploads/2018/08/SEOR.pdf.

David L. Donoho, Richard C. Liu, and Brenda MacGibbon. Minimax risk over hyperrectangles, and implications. *The Annals of Statistics*, 18(3):1416–1437, 1990. doi: 10.1214/aos/1176347758. URL https://doi.org/10.1214/aos/1176347758.

Simon S. Du, Wei Hu, Sham M. Kakade, Jason D. Lee, and Qi Lei. Few-shot learning via learning the representation, provably. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=pW2Q2xLwIMD.

Marco F. Duarte, Mark A. Davenport, Dharmpal Takhar, Jason N. Laska, Ting Sun, Kevin F. Kelly, and Richard G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, 2008. URL https://mdav.ece.gatech.edu/publications/ddtlskb-spm-2008.pdf.

Carl Eckart and Gale Young. The Approximation of One Matrix by Another of Lower Rank. *Psychometrika*, 1(3):211–218, 1936. doi: 10.1007/bf02288367.

Bradley Efron. Defining the curvature of a statistical problem (with applications to second order efficiency). *The Annals of Statistics*, 3(6):1189–1242, 1975. doi: 10.1214/aos/1176343282.

Bradley Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1): 1–26, 1979. doi: 10.1214/aos/1176344552.

Herbert Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93(3): 418–491, 1959. doi: 10.1090/s0002-9947-1959-0110078-1.

Charles Fefferman, Sergei Ivanov, Yaroslav Kurylev, Matti Lassas, and Hariharan Narayanan. Fitting a putative manifold to noisy data. In *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pp. 688–720, 2018. URL https://proceedings.mlr.press/v75/fefferman18a.html.

Marta Garnelo, Dan Rosenbaum, Christopher Maddison, Tiago Ramalho, David Saxton, Murray Shanahan, Yee Whye Teh, Danilo Rezende, and S. M. Ali Eslami. Conditional neural processes. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1704–1713. PMLR, 2018. URL https://proceedings.mlr.press/v80/garnelo18a.html.

Christopher Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Minimax manifold estimation. *Journal of Machine Learning Research*, 13:1263–1291, 2012. URL https://jmlr.org/papers/v13/genovese12a.html.

Richard D. Gill and Boris Y. Levit. Applications of the van Trees Inequality: A Bayesian Cramér-Rao Bound. *Bernoulli*, 1(1/2):59–79, 1995. doi: 10.2307/3318681.

Jonathan Gordon, Wessel P. Bruinsma, Andrew Y. K. Foong, James Requeima, Yann Dubois, and Richard E. Turner. Convolutional conditional neural processes. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=Skey4eBYPS.

Alexander Grigor'yan. Gaussian upper bounds for the heat kernel on arbitrary manifolds. *Journal of Differential Geometry*, 45:33–52, 1997. URL https://www.math.uni-bielefeld.de/~grigor/superg.pdf.

Brian C. Hall. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*. Springer International Publishing, 2 edition, 2015. doi: 10.1007/978-3-319-13467-3.

Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, et al. Array programming with NumPy. *Nature*, 585(7825):357–362, 2020. doi: 10.1038/s41586-020-2649-2.

James Harrison, Apoorva Sharma, and Marco Pavone. Meta-learning priors for efficient online bayesian regression. *arXiv preprint arXiv:1807.08912*, 2018. URL https://arxiv.org/abs/1807.08912.

Nicholas J. Higham. *Accuracy and Stability of Numerical Algorithms*. Society for Industrial and Applied Mathematics, 2 edition, 2002. doi: 10.1137/1.9780898718027.

I. Hoshi, T. Shimobaba, T. Kakue, and T. Ito. Real-time single-pixel imaging using a system on a chip field-programmable gate array. *Scientific Reports*, 12:14097, 2022. doi: 10.1038/s41598-022-18187-8.

Leon Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1–2):134–139, 1918. doi: 10.1093/biomet/12.1-2.134.

Alan Julian Izenman. Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*, 5(2):248–264, 1975. doi: 10.1016/0047-259X(75)90042-1.

Fredrik Johansson. Arb: Efficient arbitrary-precision midpoint-radius interval arithmetic. *IEEE Transactions on Computers*, 66(8):1281–1292, 2017. URL https://fredrikj.net/math/arbpaper.pdf.

J. Kiefer and J. Wolfowitz. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366, 1960. doi: 10.4153/CJM-1960-030-4. URL https://doi.org/10.4153/CJM-1960-030-4.

Hyunjik Kim, Andriy Mnih, Jonathan Schwarz, Marta Garnelo, Ali Eslami, Dan Rosenbaum, Oriol Vinyals, and Yee Whye Teh. Attentive neural processes. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=SkE6PjC9KX.

Samory Kpotufe. k-NN regression adapts to local intrinsic dimension. In *Advances in Neural Information Processing Systems*, volume 24, 2011. URL https://arxiv.org/abs/1110.4300.

S. N. Lahiri. *Resampling Methods for Dependent Data*. Springer New York, 2003. doi: 10.1007/978-1-4757-3803-2.

Béatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338, 2000. doi: 10.1214/aos/1015957395.

Lucien Le Cam. On the asymptotic theory of estimation and testing hypotheses. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pp. 129–156. University of California Press, 1956. URL https://digicoll.lib.berkeley.edu/record/112827/files/math_s3_v1_article-11.pdf.

John M. Lee. *Introduction to Smooth Manifolds*. Springer, 2 edition, 2013. doi: 10.1007/978-1-4419-9982-5.

Sameer A. Nene, Shree K. Nayar, and Hiroshi Murase. Columbia object image library (COIL-20). Technical Report CUCS-005-96, Columbia University, 1996a. URL https://www1.cs.columbia.edu/CAVE/software/softlib/coil-20.php.

Sameer A. Nene, Shree K. Nayar, and Hiroshi Murase. Columbia object image library (COIL-100). Technical Report CUCS-006-96, Columbia University, 1996b. URL https://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php.

Jeremy Reizenstein, Roman Shapovalov, Philipp Henzler, Luca Sbordone, Patrick Labatut, and David Novotny. Common objects in 3D: Large-scale learning and evaluation of real-life 3D category reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.

James Robins, Lingling Li, Rajarshi Mukherjee, Eric Tchetgen Tchetgen, and Aad van der Vaart. Minimax estimation of a functional on a structured high-dimensional model. *The Annals of Statistics*, 45(5):1951–1987, 2017. doi: 10.1214/16-AOS1515.

Yuan Shen, Duygu Ceylan, Paul Guerrero, Zexiang Xu, Niloy J. Mitra, Shenlong Wang, and Anna Frühstück. SuperGaussian: Repurposing video models for 3D super resolution. In *European Conference on Computer Vision*, 2024.

D. G. Simpson, D. Ruppert, and R. J. Carroll. On one-step GM estimates and stability of inferences in linear regression. *Journal of the American Statistical Association*, 87(418):439–450, 1992. doi: 10.1080/01621459.1992.10475224.

Miaomiao Su and Ruoyu Wang. Subsampled one-step estimation for fast statistical inference, 2025. URL https://arxiv.org/abs/2407.13446v2. Version 2, September 24, 2025.

Mohammad Tinati and Stephen Tu. On the asymptotics of self-supervised pre-training: Two-stage M-estimation and representation symmetry. In *Proceedings of Thirty Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pp. 6197–6309, 2026. URL https://proceedings.mlr.press/v336/tinati26a.html.

Nilesh Tripuraneni, Chi Jin, and Michael I. Jordan. Provable meta-learning of linear representations. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 10434–10443, 2021. URL https://proceedings.mlr.press/v139/tripuraneni21a.html.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17 (3):261–272, 2020. doi: 10.1038/s41592-019-0686-2.

Feifei Wang, Danyang Huang, Yingqiu Zhu, and Hansheng Wang. Efficient estimation for generalized linear models on a distributed system with nonrandomly distributed data, 2020. URL https://arxiv.org/abs/2004.02414.

Xianggang Yu, Mutian Xu, Yidan Zhang, Haolin Liu, Chongjie Ye, Yushuang Wu, Zizheng Yan, Chenming Zhu, Zhangyang Xiong, Tianyou Liang, Guanying Chen, Shuguang Cui, and Xiaoguang Han. MVImgNet: A large-scale dataset of multi-view images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9150–9161, 2023.

Xuhui Zhang, Jose Blanchet, Soumyadip Ghosh, and Mark S. Squillante. A class of geometric structures in transfer learning: Minimax bounds and optimality. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 3794–3820, 2022. URL https://proceedings.mlr.press/v151/zhang22a.html.

Appendix guide

Appendices A–C prove the geometric risk laws, their extensions to learned and uncertain geometry, and conditional configuration gains. Appendix D establishes the growing-dimensional separation; Appendix E certifies a finite-budget global readout. Appendix F pairs each theoretical prediction with its simulation protocol and results. Appendix G organizes captured-view evidence by collection, source and target budgets, and readout implementation.

A Geometric budget laws

A.1 Pilot-affine minimax risk on a smooth model manifold

The proof has two parts. Derivative projection bounds the residual bias and variance uniformly, giving the upper bound. A local prior then reduces every pilot-affine rule to a conditional affine Bayes risk; the covariance has D_q directions above the main-noise scale, which forces the matching lower bound.

Use the model, derivative spaces and estimator class of Theorem 3.1, and set $V_0 = \{0\}$. All geometric constants in this subsection may depend on the fixed M, N, q.

Uniform upper bound. Let $\widehat{v}$ be a measurable nearest point of M to Z_0. It exists by compactness and is unique in a fixed tubular neighborhood (Federer, 1959; Lee, 2013). Let $P_q(x)$ be the orthogonal projector onto $V_q(x)$. It is Borel across rank changes: use a finite measurable chart selection and the Moore–Penrose inverse of the derivative-column matrix. The admissible rule

$$\widehat{v}_q = \widehat{v} + P_q(\widehat{v})(Z_1 - \widehat{v})$$

satisfies, for constants depending only on M, N, q and sufficiently large k,

$$\sup_{v \in M} \mathbb{E}_v \|\widehat{v}_q - v\|^2 \le D_q/(n-k) + Ck^{-(q+1)} + Ce^{-ck}. \tag{11}$$

Indeed, compact smooth embedded manifolds have a positive tubular radius and a finite collection of centered graph charts with uniform derivative bounds and local inverse-Lipschitz bounds. On

$\|Z_0 - v\| < \rho$ for a sufficiently small fixed ρ, $v, \widehat{v}$ lie in a common chart centered at $\widehat{v}$, and $\|\widehat{v} - v\| \le 2\|Z_0 - v\|$. Taylor expansion through order q gives

$$\|(I - P_q(\widehat{v}))(\widehat{v} - v)\| \le C\|\widehat{v} - v\|^{q+1}.$$

All nonconstant Taylor terms lie in $V_q(\widehat{v})$; no lower singular-value bound is needed. Outside this event the residual is bounded by the diameter of M, and its probability is $O(e^{-ck})$. Gaussian moments yield the bias term in (11). The independent main-noise variance is $\operatorname{rank} P_q/(n-k) \le D_q/(n-k)$. The budget conditions imply $k/n \to 0$ and $ne^{-ck} \to 0$, proving the upper limit.

Choose a neighborhood U_{reg} where V_q has maximal rank D_q: a maximal nonzero minor stays nonzero nearby. Successively restrict this neighborhood to maximal-rank patches for $V_1, \dots, V_{q-1}$. Every derivative space then has constant rank on U_{reg}.

Lemma A.1 (Uniform posterior moments in a buffered patch)**.** *Choose sufficiently small coordinate balls $U_0 \Subset U_1$ on M. Let p be a smooth probability density relative to induced volume with compact support $K \Subset U_0$, positive on the interior of K. Write $\pi = p\,\mathrm{dvol}_M$ and use the joint law $V \sim \pi$, $Z_0 = V + k^{-1/2}\varepsilon_0$. Let x be the nearest projection of Z_0. On $\overline{U}_1$ choose a smooth orthonormal tangent frame E_x. There are centered graph charts*

$$h_x(u) = x + E_x u + g_x(u), \quad u \in B_{2r}(0), \quad g_x(u) \perp T_x M, \quad g_x(0) = Dg_x(0) = 0,$$

for a common $r > 0$, jointly smooth in (x, u), with all fixed-order derivatives uniformly bounded. Define $T = \sqrt{k}h_x^{-1}(V)$ when $x \in U_1$ and $V \in h_x(B_r)$, and $T = 0$ otherwise. Every auxiliary that is only defined on U_1 is extended measurably by zero off U_1 whenever a full-law random quantity is formed. For every fixed polynomial P, under the joint Bayes law,

$$\mathbb{E}[P(T) \mid Z_0] - \mathbb{E}[P(G)] \longrightarrow 0 \quad \textit{in probability}, \qquad G \sim N(0, I_d).$$

For every fixed finite degree, the same conclusion holds for a polynomial family P_x whose coefficients are continuous and uniformly bounded in $x \in \overline{U}_1$, comparing $\mathbb{E}[P_x(T) \mid Z_0]$ with $\mathbb{E}[P_x(G)]$; extend this family by zero off U_1. Posterior contributions outside $h_x(B_r)$, multiplied by any fixed power of k, vanish in probability.

Proof. The parameter-dependent inverse-function theorem and compactness of $\overline{U}_1$ give the common graph radius, uniform derivatives and volume factors J_x with $J_x(0) = 1$. In particular $\|g_x(u)\| \le C\|u\|^2$, and points of K outside $h_x(B_r)$ are separated from x by a uniform $\delta > 0$.

Choose compact prior-interior buffers $K_a \Subset \{p > 0\}$ with $\pi(K_a) \uparrow 1$. For fixed a, $p \ge c_a > 0$ near K_a. On $V \in K_a$, $\|\varepsilon_0\| \le b_k = k^{1/8}$, the pilot projection lies in this neighborhood for large k, and

$$\|x - V\| \le 2b_k/\sqrt{k}, \qquad e = Z_0 - x \perp T_x M, \qquad \|e\| \le b_k/\sqrt{k}.$$

In scaled coordinates, the exact likelihood contrast is

$$Q_k(t) = \frac{1}{2}\|t\|^2 + \frac{k}{2}\|g_x(t/\sqrt{k})\|^2 - k\langle e, g_x(t/\sqrt{k})\rangle.$$

Uniformly on this event, $Q_k(t) \ge \|t\|^2/4$ for $\|t\| < r\sqrt{k}$, and $\sup_{\|t\| \le L} |Q_k(t) - \|t\|^2/2| \le C_L(k^{-1} + b_k/\sqrt{k}) \to 0$. Outside the chart, $d_w = \|w - x\| \ge \delta$ and $\|e\| \le \delta/4$ give contrast at least $k(d_w^2 - 2\|e\|d_w)/2 \ge k\delta^2/4$.

Removing the common likelihood factor and scaling by $k^{d/2}$, the local normalizer is

$$L_k = \int_{B_{r\sqrt{k}}} e^{-Q_k(t)} p(h_x(t/\sqrt{k}))J_x(t/\sqrt{k})\,dt.$$

The integrand is bounded by $Ce^{-\|t\|^2/4}$; integration on B_1 gives $L_k \ge c_a' > 0$. The outside-chart part is at most $k^{d/2}e^{-k\delta^2/4}$. On bounded t-sets, $p(h_x(t/\sqrt{k}))/p(x) \to 1$ and $J_x(t/\sqrt{k}) \to 1$ uniformly. Dominated convergence, using $(1 + \|t\|^s)Ce^{-\|t\|^2/4}$ for each fixed degree s, yields

$$L_k/p(x) \to (2\pi)^{d/2}, \qquad \frac{\int P(t)e^{-Q_k(t)}p(h_x(t/\sqrt{k}))J_x(t/\sqrt{k})\,dt}{L_k} \to \mathbb{E}P(G).$$

The discarded mass is at most $(c'_a)^{-1}k^{d/2}e^{-k\delta^2/4}$, so it remains negligible after any fixed polynomial scaling. The same envelope bounds all needed absolute moments; expansion in finitely many monomials covers the coefficient-continuous families.

All limits are uniform on the buffered event. Its complement has probability at most $1 - \pi(K_a) + \Pr(\|\varepsilon_0\| > b_k)$. Taking $k \to \infty$ and then $a \to \infty$ proves convergence under the full Bayes law, including priors that vanish at their support boundary. □

Lemma A.2 (Posterior covariance on a regular patch)**.** *Choose the buffered balls of Lemma A.1 inside the regular patch U_{reg} above, fix one such prior, and let $C_k = \mathrm{Cov}(V \mid Z_0)$. There is $c > 0$, depending on M, q and the chosen buffered regular patch, such that under the joint law $V \sim \pi$, $Z_0 \mid V$,*

$$\Pr\{\lambda_{D_q}(C_k) \geq ck^{-q}\} \longrightarrow 1,$$

with eigenvalues in decreasing order.

Proof. Put $W_j(x) = V_j(x) \ominus V_{j-1}(x)$. On a smaller regular patch containing $\overline{U}_1$ choose smooth orthonormal frames $B_j(x)$ for each nonzero W_j and concatenate them as $B(x)$. The chart family of Lemma A.1 has uniformly bounded derivatives through every fixed order. Projection to W_j annihilates Taylor orders below j. On the local chart,

$$k^{j/2}B_j(x)^T(V - x) = H_{j,x}(T) + R_{j,k}, \qquad \|R_{j,k}\| \leq Ck^{-1/2}\|T\|^{j+1},$$

where $H_{j,x}$ consists of homogeneous degree-j polynomials. The bound on posterior moments of degree $2j+2$ gives $\mathbb{E}[\|R_{j,k}\|^2 \mid Z_0] = o_P(1)$ on the local chart. Outside it, V, x lie in fixed compact sets, so the squared scaled feature is bounded by Ck^j; the final conclusion of Lemma A.1 makes its conditional contribution $o_P(1)$. Thus the concatenated scaled coordinates equal the concatenated polynomials plus $o_{L^2(\pi(\cdot|Z_0))}(1)$ in probability. Their conditional second moments remain $O_P(1)$. Conditional Cauchy–Schwarz and the joint polynomial-moment limits consequently give, in operator norm,

$$S_kB(x)^TC_kB(x)S_k - \Gamma(x) \longrightarrow 0 \quad \text{in probability,}$$
$$\Gamma(x) = \mathrm{Cov}((H_{j,x}(G))_{j=1}^q), \quad G \sim N(0, I_d),$$

where S_k scales block W_j by $k^{j/2}$ and $x = \widehat{v}$.

This matrix $\Gamma(x)$ is positive definite. A zero-variance linear combination is a polynomial constant on all of $\mathbb{R}^d$, since the Gaussian has full support. It has no constant term, so it vanishes identically. Distinct homogeneous degrees cannot cancel. Within degree j, a vector in W_j annihilating every order-j derivative already annihilates V_{j-1} and is thus both in and orthogonal to V_j; it is zero.

Continuity on the compact regular set $\overline{U}_1$ gives $\inf_{x \in \overline{U}_1} \lambda_{\min}\Gamma(x) = 2c > 0$. The pilot projection lies there with probability tending to one because $K \Subset U_0$. Therefore with probability tending to one,

$$S_kB(x)^TC_kB(x)S_k \succeq cI, \quad B(x)^TC_kB(x) \succeq cS_k^{-2} \succeq ck^{-q}I.$$

Compression interlacing proves the ambient eigenvalue claim. □

Minimax lower bound. For the fixed prior from Lemma A.2, conditional affine square completion, including the offset and independent randomization, gives

$$\inf_{A,b} \mathbb{E}_\pi\|A(Z_0, \omega)Z_1 + b(Z_0, \omega) - V\|^2 = \mathbb{E}\,\mathrm{tr}\{C_k(I + (n-k)C_k)^{-1}\}.$$

This Bayes risk lower-bounds minimax risk. On the covariance event its first D_q eigenvalues contribute at least

$$\frac{D_q}{n-k}\frac{(n-k)ck^{-q}}{1 + (n-k)ck^{-q}}.$$

The event probability tends to one, $(n-k)/k^q \to \infty$ and $n/(n-k) \to 1$. Hence the normalized lower limit is D_q, matching (11) and proving the theorem.

A.2 Proof of the critical curvature law

The argument refines the first two derivative layers. We first expand the attaining rule into tangent noise, centered quadratic normal error and a uniform remainder. We then derive the posterior covariance at the same scales and optimize its normal block. Both calculations give the resolvent in Theorem 3.2, including singular curvature.

Use the notation and attaining rule of Theorem 3.2, with $r = n - k$ and $A_p = P_p + B_p$. For a standard tangent Gaussian G_x, Wick's formula (Isserlis, 1918) gives

$$\mathbb{E}\mathrm{II}_x(G_x, G_x) = H_x, \qquad \mathrm{Cov}\big(\tfrac{1}{2}\mathrm{II}_x(G_x, G_x)\big) = K_x. \tag{12}$$

The tensors and their derivatives are bounded on the fixed compact manifold; K_x may change rank. We prove the uniform pointwise risk claimed in the theorem and then its minimax lower bound.

A.2.1 Uniform frequentist upper bound

Compactness gives a Borel nearest selection; in a fixed tubular neighborhood it is the unique smooth nearest projection. The resulting A_p and offset $(I - A_p)\mu_p$ are Borel, even at rank changes of K. Moreover $0 \preceq A_p \preceq I$, since $P_p B_p = 0$ and the eigenvalues of B_p lie in $[0, 1)$.

Fix a sufficiently small tubular radius $\rho > 0$. On $\|\varepsilon_0\| \le \rho\sqrt{k}$, after reducing ρ if needed, p, v lie in a graph chart centered at p with uniform inverse bounds. The nearest inequality gives $\|p - v\| \le 2\|\varepsilon_0\|/\sqrt{k}$. Smoothness of nearest projection, with derivative P_v on the manifold, gives

$$p - v = k^{-1/2} P_v \varepsilon_0 + O(\|\varepsilon_0\|^2/k), \qquad \|P_p - P_v\| \le C\|\varepsilon_0\|/\sqrt{k}. \tag{13}$$

The constants are uniform in v by a finite compact atlas. Define $\alpha = \sqrt{k} P_p(v - p) \in T_p M$. Then $\|\alpha\| \le 2\|\varepsilon_0\|$ and

$$\|\alpha + P_v \varepsilon_0\| \le C\|\varepsilon_0\|^2/\sqrt{k}.$$

The graph representation with tangent coordinate $\alpha/\sqrt{k}$ yields

$$v - p = \frac{\alpha}{\sqrt{k}} + \frac{1}{2k}\mathrm{II}_p(\alpha, \alpha) + R, \qquad \|R\| \le C\|\varepsilon_0\|^3/k^{3/2}.$$

Because $(I - A_p)\alpha = 0$,

$$k(I - A_p)(\mu_p - v) = \tfrac{1}{2}(I - B_p)\{H_p - \mathrm{II}_p(\alpha, \alpha)\} + O(\|\varepsilon_0\|^3/\sqrt{k}). \tag{14}$$

Here $I - B_p$ can be used since both H_p and $\mathrm{II}_p(\alpha, \alpha)$ are normal.

Write $B_v^\tau = \tau K_v (I + \tau K_v)^{-1}$. The map $(t, K) \mapsto tK(I + tK)^{-1}$ is Lipschitz on a bounded set of positive semidefinite K and a compact interval about τ; this follows from the resolvent identity and $\|(I + tK)^{-1}\| \le 1$. Consequently the right side of (14) converges in L^2, uniformly in v, to

$$W_v = \tfrac{1}{2}(I - B_v^\tau)\{H_v - \mathrm{II}_v(P_v\varepsilon_0, P_v\varepsilon_0)\}.$$

The differences of H and B are bounded by $C\|\varepsilon_0\|/\sqrt{k} + o(1)$; the difference of the quadratic tensors is bounded by $C(\|\varepsilon_0\|^3/\sqrt{k} + \|\varepsilon_0\|^4/k)$. All required Gaussian moments are finite for fixed N. On the complementary event, $\|(I - A_p)(\mu_p - v)\|$ is at most $\mathrm{diam}(M) + C/k$, hence its squared scaled contribution is $O(k^2 e^{-ck})$. Polynomial Gaussian tails also remove the corresponding contribution of W_v. This justifies the stated uniform L^2 convergence without conditioning away rare errors.

Equation (12) implies

$$\mathbb{E}\|W_v\|^2 = \mathrm{tr}\{(I - B_v^\tau)^2 K_v\}.$$

Conditional on the pilot, the exact main-noise risk is

$$\|(I - A_p)(\mu_p - v)\|^2 + \frac{1}{r}\,\mathrm{tr}\, A_p^2, \qquad \mathrm{tr}\, A_p^2 = d + \mathrm{tr}\, B_p^2.$$

Uniform continuity, Gaussian localization and boundedness give $\mathbb{E}\,\mathrm{tr}\, B_p^2 \to \mathrm{tr}((B_v^\tau)^2)$ uniformly. Since $n/r \to 1$ and $n/k^2 \to \tau$, the normalized risk tends uniformly to

$$d + \mathrm{tr}((B_v^\tau)^2) + \tau\,\mathrm{tr}\{(I - B_v^\tau)^2 K_v\} = d + \mathrm{tr}\{\tau K_v (I + \tau K_v)^{-1}\}.$$

The final identity follows on each eigenvalue $s = \tau\lambda \ge 0$ from $(s/(1+s))^2 + s/(1+s)^2 = s/(1+s)$. This proves the pointwise assertion and the minimax upper bound in Theorem 3.2. The maximum exists by continuity on compact M.

A.2.2 Posterior covariance at the critical scale

Choose the buffered prior and local coordinates of Lemma A.1, writing p for the pilot projection. Smooth tangent and normal frames exist on a smaller chart. The lemma supplies all fixed polynomial moments and power-weighted tail bounds used below.

Lemma A.3 (Scaled tangent and normal covariance)**.** *For this buffered prior, let $C_k = \mathrm{Cov}(V \mid Z_0)$, and express it in orthonormal tangent and normal frames at p. Writing its blocks as C_{TT}, C_{TN}, C_{NN}, we have*

$$kC_{TT} \to I_d, \qquad k^{3/2}C_{TN} \to 0, \qquad k^2 C_{NN} - K_p^{\mathrm{nor}} \to 0$$

in probability. Here K_p^{nor} is the normal-coordinate representation of K_p; it may be singular.

Proof. Inside the chart the tangent coordinate of $V - p$ is exactly $T/\sqrt{k}$, while its normal coordinate is

$$\frac{1}{2k}\mathrm{II}_p(E_pT, E_pT) + O(\|T\|^3/k^{3/2}).$$

Lemma A.1 through degree six makes the scaled remainder $o_{L^2(\pi(\cdot|Z_0))}(1)$ in probability. Outside-chart contributions are negligible after multiplying squared coordinate bounds by k^2, since M is compact and the posterior tail beats every fixed power. Conditional Cauchy–Schwarz then permits replacing the coordinates by their linear and quadratic polynomials. Gaussian second and fourth moments yield I_d and K_p^{nor}, and the centered linear/quadratic covariance vanishes by odd Gaussian moments. Smooth coefficients on the buffered compact patch give the required uniformity. No inverse of K_p is used. □

Lemma A.4 (Resolvent limit with a singular normal block)**.** *For the preceding covariance and $r/k^2 \to \tau$, under the joint law,*

$$\mathrm{tr}\{rC_k(I + rC_k)^{-1}\} - d - \mathrm{tr}\{\tau K_V(I + \tau K_V)^{-1}\} \longrightarrow 0$$

in probability and in L^1.

Proof. In the moving orthonormal frames put $L = I + rC_{TT}$, $D = rC_{TN}$, $F = I + rC_{NN}$. Then

$$L/k \to \tau I_d, \quad D = \sqrt{k}\, o_P(1), \quad F = I + \tau K_p^{\mathrm{nor}} + o_P(1).$$

Thus $L^{-1} = O_P(k^{-1})$ and $D^T L^{-1} D = o_P(1)$. The Schur complement $S = F - D^T L^{-1} D$ converges to $I + \tau K_p^{\mathrm{nor}}$ in the sense that their difference vanishes. Its inverse differs by $o_P(1)$ from $(I + \tau K_p^{\mathrm{nor}})^{-1}$, whose norm is at most one. For example restrict to events where the error is below $1/2$, apply the resolvent identity there, and exhaust their probability. Block inversion gives lower-right block S^{-1} and upper-left block $L^{-1} + L^{-1}DS^{-1}D^TL^{-1}$, whose trace tends to zero. Since $\mathrm{tr}\{rC(I + rC)^{-1}\} = N - \mathrm{tr}(I + rC)^{-1}$, the claimed limit follows. Smoothness and $p - V \to 0$ replace K_p by K_V. Both trace functionals lie in $[0, N]$, so convergence in probability implies L^1 convergence by boundedness. □

A.2.3 Full-class minimax lower bound and consequences

Given Z_0, affine square completion under the prior gives minimum conditional risk

$$\mathrm{tr}\{C_k(I + rC_k)^{-1}\}.$$

This covers arbitrary coefficients and offsets. Independent randomization cannot reduce the conditional quadratic minimum. The corresponding Bayes risk lower-bounds the supremum risk of every admissible estimator. Lemma A.4 and $n/r \to 1$ therefore imply

$$\liminf nR_M(k, r) \geq d + \int \mathrm{tr}\{\tau K_v(I + \tau K_v)^{-1}\}\, \pi(dv).$$

Let v_* maximize the continuous trace functional. For each $\epsilon > 0$ choose a coordinate neighborhood where its value is at least its maximum minus ϵ, and place a smooth compactly supported probability density inside that neighborhood. The preceding covariance and resolvent arguments apply there, without requiring constant second-order rank. First take the statistical limit for this fixed prior, then let $\epsilon \downarrow 0$. This proves the lower bound in Theorem 3.2.

Endpoint limits and sphere. Since $\operatorname{range} K_x$ is the span of the second fundamental form, $\dim V_2(x) = d + \operatorname{rank} K_x$. Uniform boundedness of K gives $F_M(\tau) \to d$ as $\tau \downarrow 0$; the rank upper bound and any maximal-rank point give $F_M(\tau) \to D_2$ as $\tau \uparrow \infty$. These limits follow the critical limit at each fixed τ. On a radius-R_{sph} sphere, $K = d\nu\nu^T/(2R_{\mathrm{sph}}^2)$, hence $F_M(\tau) = d + \tau d/(2R_{\mathrm{sph}}^2 + \tau d)$.

A.3 Quadratic-readout risk

In the same fixed-manifold model, use the quadratic readout (7), denoted $\widehat{x}_2$, with pilot projection p and $r = n - k$. Write the true target as x, $Z_0 = x + G_0/\sqrt{k}$, $Z_1 = x + G_1/\sqrt{r}$ for independent standard Gaussian G_i, and $v = P_p(Z_1 - p)$, $Q_p = I - P_p$.

Theorem A.5 (Uniform quadratic-readout upper bound)**.** *There exist $C, c > 0$, depending only on fixed M, N, such that*

$$\sup_{x \in M} \mathbb{E}_x \|\widehat{x}_2 - x\|^2 \le \frac{d}{r} + \frac{C}{k^3} + \frac{C}{rk} + \frac{C}{r^2} + Ce^{-ck}.$$

Consequently, if $k \to \infty$, $k/n \to 0$ and $n/k^3 \to 0$, then

$$n \sup_{x \in M} \mathbb{E}_x \|\widehat{x}_2 - x\|^2 \longrightarrow d.$$

In particular this holds at $n/k^2 \to \tau \in (0, \infty)$, where Theorem 3.2 gives the affine-after-pilot constant $d + \max_x \operatorname{tr}[\tau K_x (I + \tau K_x)^{-1}]$. This is strictly larger than d when $K_x \neq 0$ at some point. The affine lower bound is imported from that theorem.

A.3.1 Exact conditional risk and uniform remainder

Compact embeddedness supplies a positive tubular radius, uniformly bounded derivatives of local normal graphs through order three, and uniformly bounded II. A measurable nearest-point selection exists by the compact argmin correspondence. The geometric tensors are smooth on M, so the displayed rule is Borel.

For every realization of the pilot put $u = P_p(x - p)$ and

$$e_p(x) = Q_p(x - p) - \tfrac{1}{2}\mathrm{II}_p(u, u).$$

Conditionally on the pilot, $v = u + \eta$, where $\eta \sim N(0, P_p/r)$ is independent of it. The error decomposes into the orthogonal tangent part η and normal part

$$-e_p(x) + \mathrm{II}_p(u, \eta) + \tfrac{1}{2}\{\mathrm{II}_p(\eta, \eta) - H_p/r\}.$$

The linear and centered quadratic Gaussian terms have zero means and zero cross covariance by odd Gaussian moments. Wick's identity therefore gives the exact conditional risk

$$\mathbb{E}[\|\widehat{x}_2 - x\|^2 \mid Z_0] = \frac{d}{r} + \|e_p(x)\|^2 + \frac{1}{r}\sum_i \|\mathrm{II}_p(u, e_i)\|^2 + \frac{1}{2r^2}\sum_{i,j} \|\mathrm{II}_p(e_i, e_j)\|^2.$$

The sums are basis independent.

By the nearest-point property, $\|x - p\| \le 2\|Z_0 - x\| = 2\|G_0\|/\sqrt{k}$ even when the projection is not unique. Hence $\mathbb{E}\|u\|^2 \le 4N/k$ and $\mathbb{E}\|x - p\|^6 \le C/k^3$, uniformly in x. On $\|x - p\|$ less than a sufficiently small fixed radius, the normal graph at p gives $\|e_p(x)\| \le C\|u\|^3 \le C\|x - p\|^3$. On the complement $e_p(x)$ is uniformly bounded, since $p, x \in M$ and II is bounded. The nearest-point inequality and a fixed-dimensional Gaussian tail bound make that complement's probability at most Ce^{-ck}. Therefore

$$\mathbb{E}\|e_p(x)\|^2 \le C/k^3 + Ce^{-ck}.$$

Substituting into the conditional identity proves the finite upper bound. Its leading lower inequality is also immediate from the exact identity: risk is at least d/r. Under the stated limits $n/r \to 1$, and each remaining normalized upper term vanishes. If the noise-mean correction is omitted, replace $-e_p(x)$ by $-e_p(x) + H_p/(2r)$; Young's inequality retains the same stated upper order.

Using the full aggregate. Replacing Z_1 by $\bar{Z} = (kZ_0 + rZ_1)/n$ preserves the limit. To see this, let F_p denote the quadratic map and $\Delta = \bar{Z} - Z_1$. Gaussian moments give $\mathbb{E}\|\Delta\|^{2j} \le C_j\{k/(nr)\}^j$ for $j = 1, 2$, while $\mathbb{E}\|P_p(Z_1 - p)\|^4 \le C(k^{-2} + r^{-2})$. The quadratic difference identity yields $\mathbb{E}\|F_p(\bar{Z}) - F_p(Z_1)\|^2 \le Ck/(nr) = o(n^{-1})$. Cauchy–Schwarz then preserves normalized risk. Omitting the uniformly $O(r^{-1})$ correction $H_p/(2r)$ also preserves the limit.

A.3.2 Unrestricted lower bound

The unrestricted benchmark follows from the multivariate van Trees inequality (Gill & Levit, 1995, Section 4). In a normal graph $x(u) = x_0 + Eu + g(u)$, take $E^TE = I_d$, $E^Tg = 0$ and $\|Dg(u)\| \le C\epsilon$ on a ball of radius ϵ. Choose a smooth interior prior with finite Fisher matrix J_π and zero boundary terms. Projection onto E contracts ambient squared error, and the two independent observation precisions sum to n. Applying the inequality to the chart coordinate gives

$$\mathbb{E}\|\widehat{x} - x(u)\|^2 \ge \operatorname{tr}[(n\mathbb{E}[I_d + Dg(u)^T Dg(u)] + J_\pi)^{-1}] \ge \frac{d}{n(1 + C\epsilon^2) + \|J_\pi\|}.$$

Letting $n \to \infty$ and then $\epsilon \downarrow 0$ gives the normalized lower bound d, attained by the quadratic rule under its pilot conditions.

B Learning and uncertainty in the geometry

This section first quantifies source-estimation error in quadratic prediction, then gives finite-search and source-design extensions. The final part treats normal model error, which requires retaining additional main-block information.

B.1 Learning the coefficient manifold

The upper bound separates target noise from source-estimation error, controlling their dependence through the learned geometry. The lower bound perturbs the embedding in normal directions while allowing target motion in tangent directions. The resulting information bound identifies source prediction leverage as the additional leading risk.

Use the source and target experiment, coefficient class $\mathcal{K}$, guarded readout and assumptions of Theorem 4.1. Here $p = D$ denotes feature dimension, $r = n - k$, and $Z_i = Bf + G_i/\sqrt{n_i}$ with $n_0 = k, n_1 = r$ and independent standard Gaussian G_i. Let $M_W = W\mathcal{F}$ and $p_0 = W\widehat{f}$; write P_0, II_0, H_0 for its learned tensors at p_0. Write $\mathcal{G} = \{\|W\|_F \le 2L, \sigma_{\min}(W) \ge s/2\}$, q for the quadratic output, and $\widehat{\beta} = r\widehat{\nu}/(1 + r\widehat{\nu})$. Thus $\widehat{x} = (1 - \widehat{\beta})q + \widehat{\beta}Z_1$ on $\mathcal{G}$, with fallback Z_1. Set $R^2 = \max_{f\in\mathcal{F}} \|f\|^2$. We prove uniform pointwise convergence and the unrestricted minimax constant over $(B, f) \in \mathcal{K} \times \mathcal{F}$.

B.1.1 Uniform expansion with estimated geometry

For W in the compact guarded class, the fixed feature atlas and the lower singular-value bound give uniform normal graphs through order three, locally Lipschitz tangent projectors and a common tube radius. Indeed tangent Gram inverses are uniformly bounded, while $\|W(f - g)\| \ge (s/2)\|f - g\|$ separates nonlocal branches. The parameter-dependent inverse-function theorem supplies the local derivative bounds. Compact minimization on $\mathcal{F}$ gives joint Borel selection of p_0.

Write $D = W - B$, $e_0 = Df$, $y = Wf$, and let P, Q be the true tangent and normal projectors at $x = Bf$. Then

$$e_0 \sim N(0, \nu(f)I_N), \qquad \nu(f) = \sigma_s^2\|f\|^2/m,$$

independently of the target noises. Nearest projection and inverse stability give

$$\|p_0 - y\| \le 2(\|G_0\|/\sqrt{k} + \|e_0\|), \qquad \|\widehat{f} - f\| \le (2/s)\|p_0 - y\|.$$

Thus, for every fixed $a \ge 1$,

$$\mathbb{E}[1_{\mathcal{G}}\{\|p_0 - y\|^a + \|\widehat{f} - f\|^a + \|P_0 - P\|^a\}] \le C_a(k^{-a/2} + m^{-a/2}). \tag{LG.1}$$

Let $\mathcal{L} = \{\|G_0\|/\sqrt{k} + \|e_0\| \le c_\rho\}$, where the fixed c_ρ is small enough for these charts. On $\mathcal{G} \cap \mathcal{L}$, $\|P_0 - P\| \le C(\|D\|_F + \|\widehat{f} - f\|)$. Gaussian tails bound chart failure by $C(e^{-ck} + e^{-cm})$ and guard rejection by Ce^{-cm}, uniformly. The latter follows because $\|D\|_F \le \min(L, s/2)$ implies the guard.

Set $u = P_0(x - p_0)$ and $e = Q_0(x - p_0) - \mathrm{II}_0(u, u)/2$, where $Q_0 = I - P_0$. The exact decomposition

$$\begin{aligned} e + Qe_0 = Q_0(y - p_0) - \tfrac{1}{2}\mathrm{II}_0(P_0(y - p_0), P_0(y - p_0)) \\ + (Q - Q_0)e_0 + \mathrm{II}_0(P_0(y - p_0), P_0 e_0) - \tfrac{1}{2}\mathrm{II}_0(P_0 e_0, P_0 e_0) \end{aligned}$$

retains the dependence of source error and learned geometry. The first term is $O(\|y - p_0\|^3)$ in a common graph. Equation (LG.1), Gaussian moments and Cauchy–Schwarz bound each mixed term without independence assumptions. For example, the squared projector/source product is bounded by $C(k^{-1} + m^{-1})m^{-1}$. Consequently

$$\mathbb{E}[1_{\mathcal{G}}\|e + Qe_0\|^2] \le C\{k^{-3} + (km)^{-1} + m^{-2} + e^{-ck} + e^{-cm}\}. \tag{LG.2}$$

The four terms above have squared orders $k^{-3} + m^{-3}$, $(k^{-1} + m^{-1})m^{-1}$, $(k^{-1} + m^{-1})m^{-1}$, and m^{-2}. On $\mathcal{G} \cap \mathcal{L}^c$, all geometric and source factors are bounded. Main noise is independent of $\mathcal{L}$ and appears with degree at most two, so its restricted moments contribute exponential tails. This proves (LG.2) and the exceptional parts of the moment estimates below.

Conditionally on (W, Z_0), put $\eta = P_0 G_1/\sqrt{r}$. The exact identity is

$$q - x = \eta - e + \mathrm{II}_0(u, \eta) + \tfrac{1}{2}\{\mathrm{II}_0(\eta, \eta) - H_0/r\}.$$

It follows that $q - x = PG_1/\sqrt{r} + Qe_0 + R_0$ on the guard, with

$$\begin{aligned} \mathbb{E}[1_{\mathcal{G}}\|R_0\|^2] \le C\{k^{-3} + (km)^{-1} + m^{-2} + (rk)^{-1} \\ + (rm)^{-1} + r^{-2} + e^{-ck} + e^{-cm}\}. \end{aligned} \tag{LG.3}$$

The same calculation in L^4, using twelfth moments for the cubic graph remainder and eighth moments for products, gives

$$\|1_{\mathcal{G}} R_0\|_{L^4} \le C\{k^{-3/2} + (km)^{-1/2} + m^{-1} + (rk)^{-1/2} + (rm)^{-1/2} + r^{-1} + e^{-ck} + e^{-cm}\}. \tag{LG.4}$$

Constants in tail exponents may decrease between lines. The assumed rates imply $m \ge c_0 n$, $r/n \to 1$, and that (LG.3) is $o(n^{-1})$ and (LG.4) is $o(n^{-1/2})$, uniformly.

B.1.2 Estimated source-variance weight

Let $\beta(f) = r\nu(f)/(1 + r\nu(f))$. Compactness of $\mathcal{F}$, (LG.1), and bounded $r\sigma_s^2/m$ imply

$$\|1_{\mathcal{G}}(\widehat{\beta} - \beta(f))\|_{L^4} \le C(k^{-1/2} + m^{-1/2}).$$

By (LG.4), $\|1_{\mathcal{G}}(Z_1 - q)\|_{L^4} = O(n^{-1/2})$. Hölder's inequality therefore makes replacing $\widehat{\beta}$ by $\beta(f)$ an $o(n^{-1/2})$ change in L^2.

The leading error is now

$$A = PG_1/\sqrt{r} + \beta(f)QG_1/\sqrt{r} + (1 - \beta(f))Qe_0.$$

Guard rejection contributes $o(n^{-1})$ to the squared difference from A, using its Gaussian tail moments and the fallback Z_1. Globally $\sup_{B,f} n\mathbb{E}\|\widehat{x} - x - A\|^2 \to 0$. Independence of the two leading noises and orthogonal supports give

$$\mathbb{E}\|A\|^2 = \frac{d}{r} + (N - d)\frac{\nu(f)}{1 + r\nu(f)}. \tag{LG.5}$$

Its normalized value is uniformly bounded. Cauchy–Schwarz for the residual cross term proves the asserted uniform risk law and its minimax upper bound.

B.1.3 Unrestricted lower bound

Select f_0 maximizing $\|f\|^2 = R^2 > 0$ and an interior $B_0 \in \mathcal{K}$. The source identity design implies $R^2 \geq p$, so the denominator below is nonzero. Choose a local chart $f(u)$, $u \in \mathbb{R}^d$, such that $f(0) = f_0$ and $B_0 Df(0) = T_0$ is orthonormal. Let $U_0 \in \mathbb{R}^{N\times(N-d)}$ be an orthonormal true normal frame. Define

$$B(a) = B_0 + U_0 a f_0^T / R^2, \qquad F(u,a) = B(a) f(u).$$

For a sufficiently small fixed parameter ball all coefficients stay inside $\mathcal{K}$. At zero, $O_0 := DF(0) = [T_0, U_0]$ is orthogonal. Smoothness on a fixed compact local neighborhood bounds D^2F. Integrating this derivative along line segments gives

$$\sup_{\|(u,a)\|\leq\epsilon} \|DF(u,a) - O_0\|_{\mathrm{op}} \leq C\epsilon.$$

In particular $DF^T DF = I_N + O(\epsilon)$ uniformly. Source observations depend only on a, with exact information $\mathrm{diag}(0_d, \nu_0^{-1} I_{N-d})$, where $\nu_0 = \sigma_s^2 R^2/m$. The target blocks contribute $nDF^T DF$.

Apply the multivariate van Trees inequality for the target function F (Gill & Levit, 1995, Section 4), using a smooth interior prior with finite Fisher matrix J_π. The Gaussian likelihood is smooth on the compact parameter ball and the prior vanishes at its boundary. With

$$J = \mathbb{E}_\pi DF, \qquad I = n\mathbb{E}_\pi DF^T DF + \mathrm{diag}(0_d, \nu_0^{-1} I_{N-d}) + J_\pi,$$

the Bayes risk is at least $\mathrm{tr}(JI^{-1}J^T)$, also for randomized estimators after conditioning on their independent seed. The direct derivative bound gives $\|J - O_0\|_{\mathrm{op}} \leq C\epsilon$, so $\sigma_{\min}(J) \geq 1 - C\epsilon$ and $J^T J \succeq (1 - C'\epsilon) I_N$ for small ϵ. Matrix order, with constants renamed, gives

$$\mathrm{tr}(JI^{-1}J^T) \geq (1 - C\epsilon) \left\{ \frac{d}{n(1+C\epsilon) + \|J_\pi\|} + \frac{N-d}{n(1+C\epsilon) + \nu_0^{-1} + \|J_\pi\|} \right\}. \tag{LG.6}$$

First let $n \to \infty$ with this prior fixed, then $\epsilon \downarrow 0$. For $\gamma > 0$ the limit lower bound is the claimed constant; for $\gamma = 0$ the first term gives d. The admissible local subfamily lower-bounds unrestricted minimax risk over the full model, completing the proof.

Finite pilot search. Use the model, guard $\mathcal{G}$, and assumptions of Appendix B.1. Let $\mathcal{F}_k \subset \mathcal{F}$ be a deterministic finite Euclidean h_k-net, with $h_k \leq C_0 k^{-1/2}$. Select the first minimizer of $\|Wg - Z_0\|$ over $g \in \mathcal{F}_k$ in a fixed order, denoted g_k. Set $\widehat{f} = g_k$, or use any Borel $\widehat{f} \in \mathcal{F}$ satisfying

$$\|W\widehat{f} - Z_0\| \leq \|Wg_k - Z_0\|.$$

Evaluate the exact learned tangent and second fundamental form at $p_0 = W\widehat{f}$ and otherwise use the same guarded quadratic rule and estimated source-variance weight. The finite-search rule has the same limit. Choose $g \in \mathcal{F}_k$ with $\|g - f\| \leq h_k$. On $\mathcal{G}$, write $y = Wf$, $e_0 = (W - B)f$ and $a = \|G_0\|/\sqrt{k} + \|e_0\|$. Finite minimization and $\|W\|_{\mathrm{op}} \leq 2L$ give

$$\|p_0 - y\| \leq \|p_0 - Z_0\| + \|Z_0 - y\| \leq \|Wg - Z_0\| + \|Z_0 - y\| \leq 2a + 2Lh_k,$$
$$\|\widehat{f} - f\| \leq (2/s)\|p_0 - y\|.$$

Thus moments of order a' are $O(k^{-a'/2} + m^{-a'/2})$. For large k, the deterministic net error lies within half the common chart radius, leaving exceptional probability $C(e^{-ck} + e^{-cm})$.

Appendix B.1 uses nearest-point optimality only for this localization step. Its graph expansion holds for any two nearby points on $W\mathcal{F}$, and its conditional Gaussian calculation for any source/pilot-measurable p_0. The source-error products, estimated-weight remainder and guard fallback therefore retain their bounds and leading risk. The unchanged experiment gives the same lower bound. A finite smooth atlas gives $O(h^{-d}) = O(k^{d/2})$ net points; a fixed tie order makes the minimizer Borel. Appendix F.3.3 specifies the grid used in the experiments.

Positive-definite source designs. Keep the coefficient class and target experiment of Theorem 4.1, but let the known source features satisfy $m^{-1}\sum_j f_j f_j^T = \Sigma \succ 0$ for a fixed matrix Σ. Source least squares gives

$$W = B + \frac{\sigma_s}{\sqrt{m}} E\Sigma^{-1/2}, \qquad \ell_\Sigma(f) = f^T\Sigma^{-1}f, \qquad \widehat{\nu} = \frac{\sigma_s^2}{m}\ell_\Sigma(\widehat{f}).$$

Use the original observable regularity guard, quadratic rule, and weight with this $\widehat{\nu}$. Under the rates of Theorem 4.1, the resulting uniform risk law is

$$n\mathbb{E}\|\widehat{x} - Bf\|^2 = d + (N-d)\frac{\gamma\ell_\Sigma(f)}{1+\gamma\ell_\Sigma(f)} + o(1).$$

The unrestricted minimax constant replaces $\ell_\Sigma(f)$ by its maximum over $\mathcal{F}$. This holds for a fixed positive-definite design, uniformly over finitely many such designs, and for the finite-search implementation. To obtain the upper bound, substitute $e_0 \sim N(0, \sigma_s^2\ell_\Sigma(f)I_N/m)$ into (LG.1)–(LG.5). Fixed positive definiteness preserves the moment and guard bounds, and ℓ_Σ is Lipschitz on compact $\mathcal{F}$, preserving the estimated-weight argument.

For the lower bound choose a maximizer f_0, put $\ell_0 = \ell_\Sigma(f_0)$ and $b = \Sigma^{-1}f_0/\ell_0$, and use the local family $B(a) = B_0 + U_0 a b^T$ at an interior B_0. Since $b^T f_0 = 1$, the target derivative is still $[T_0, U_0]$, while the normal source information is

$$(m/\sigma_s^2) b^T\Sigma b\, I_{N-d} = m/(\sigma_s^2\ell_0)\, I_{N-d}.$$

The family stays in $\mathcal{K}$ on a small fixed neighborhood, so the same van Trees argument (LG.6) applies with $\nu_0 = \sigma_s^2\ell_0/m$.

Minimizing this worst limiting risk at fixed source cost is therefore the classical maximum-leverage design problem (Kiefer & Wolfowitz, 1960).

B.2 Proof for local normal model error

Use the fixed manifold and curvature notation of Section 3.1. For $\varrho \geq 0$, define

$$\mathcal{M}_{n,\varrho} = \{x + u/\sqrt{n} : x \in M,\ u \perp T_x M,\ \|u\| \leq \varrho\}.$$

Let $R_{M,\varrho}(k, r)$ be the pilot-affine minimax risk over this set under the same independent Gaussian blocks, with $r = n - k$. The nominal manifold and uncertainty radius are known; the normal displacement is unknown. Write $h = N - d \geq 1$.

For $t > 0$ let $\lambda_1(x), \ldots, \lambda_h(x) \geq 0$ be the normal eigenvalues of K_x. If $\varrho > 0$, let $c_x(t, \varrho) > 0$ be the unique solution of

$$\sum_{j=1}^{h}(c_x - t\lambda_j(x))_+ = \varrho^2. \tag{15}$$

For $\varrho = 0$ set $c_x(t, 0) = 0$. Define normal operators, in any orthonormal eigenbasis of K_x,

$$S_x^* = \mathrm{diag}((c_x - t\lambda_j)_+), \quad T_x^* = tK_x + S_x^*, \quad B_x^* = T_x^*(I + T_x^*)^{-1}.$$

All are zero on the tangent space. The definitions are basis independent at repeated eigenvalues. Set

$$F_{M,\varrho}(\tau) = d + \max_{x\in M}\sum_{j=1}^{h}\frac{\max\{\tau\lambda_j(x), c_x(\tau,\varrho)\}}{1 + \max\{\tau\lambda_j(x), c_x(\tau,\varrho)\}}.$$

Theorem B.1. *For fixed M, N, ϱ and $\tau \in (0, \infty)$, as $k \to \infty$ with $n/k^2 \to \tau$,*

$$nR_{M,\varrho}(k, n-k) \longrightarrow F_{M,\varrho}(\tau).$$

An attaining rule uses a Borel nearest point p of Z_0 on the nominal M, $t = r/k^2$, $\mu_p = p + H_p/(2k)$, and

$$\widehat{v} = \mu_p + (P_p + B_p^*(t, \varrho))(Z_1 - \mu_p).$$

No constant-curvature-rank assumption is imposed. For $\varrho = 0$ this reduces to the base critical theorem. For fixed τ, $F_{M,\varrho}(\tau) \to N$ as $\varrho \to \infty$, an iterated limit only.

B.2.1 The normal covariance game

For every finite-dimensional PSD matrix T, define $f(T) = \operatorname{tr}\{T(I+T)^{-1}\}$, where I has the same dimension as T. Extending T by zero to a larger space leaves $f(T)$ unchanged. For an $h \times h$ PSD matrix L, we have

$$\min_B \max_{\|u\| \le \varrho} \{\|B\|_F^2 + \operatorname{tr}((I-B)L(I-B)^T) + \|(I-B)u\|^2\} = \max_{S \succeq 0, \operatorname{tr} S \le \varrho^2} f(L+S). \tag{16}$$

For any feasible covariance $S = \sum_j s_j e_j e_j^T$, a centered distribution in the radius-ϱ ball places mass $s_j/(2\varrho^2)$ at $\pm\varrho e_j$ and the rest at zero. Its average risk lower-bounds the worst-offset risk. Affine square completion then gives $f(L+S)$, attained at $B = (L+S)(I+L+S)^{-1}$; this proves the lower inequality. For $\varrho = 0$, take $S = 0$.

Let l_j be the eigenvalues of L and choose $S^* = \operatorname{diag}((c-l_j)_+)$ with $\sum(c-l_j)_+ = \varrho^2$. For $\varrho > 0$, continuity and strict monotonicity above the smallest eigenvalue give a unique root. The function f has gradient $(I+T)^{-2}$ and is concave: with $R = (I+T)^{-1}$, the second directional derivative of $\operatorname{tr}(I+T)^{-1}$ is $2\operatorname{tr}(RDRDR) \ge 0$.

At $T^* = L + S^*$, every gradient eigenvalue is at most $(1+c)^{-2}$, with equality on the support of S^*. Thus

$$\operatorname{tr}\{(I+T^*)^{-2}(S-S^*)\} \le (1+c)^{-2}(\operatorname{tr} S - \varrho^2) \le 0.$$

Concavity proves that S^* maximizes f. For $B^* = T^*(I+T^*)^{-1}$, $\|I-B^*\|_{\mathrm{op}}^2 = (1+c)^{-2}$ and

$$\max_{\|u\| \le \varrho} \|(I-B^*)u\|^2 = \varrho^2/(1+c)^2 = \operatorname{tr}((I-B^*)S^*(I-B^*)^T).$$

Consequently its worst-case objective equals $f(T^*)$, proving the other inequality and attainment. When $\varrho = 0$, take $S^* = 0$ and ordinary square completion.

For fixed ϱ, the water level and its spectral function are continuous, including at eigenvalue collisions. Thus the ambient contraction B_x^* is uniformly continuous on compact M, even when curvature rank changes.

B.2.2 Uniform upper bound

For large n, the tubular-neighborhood theorem gives the unique representation $v = x + u/\sqrt{n}$. The projection expansion from Appendix A.2 remains uniform over $\|u\| \le \varrho$:

$$\sqrt{k}(p-x) = P_x\varepsilon_0 + o_{L^s}(1), \qquad s < \infty.$$

The extra displacement has scaled size $O(\sqrt{k/n}) = o(1)$; the tube complement still has exponentially small probability.

Apply the graph at p to the nominal point x, exactly as in the base upper proof. The uniform graph remainder and uniform continuity of the contraction B^* give

$$k(I - P_p - B_p^*)(\mu_p - v) \longrightarrow (I - B_x^*(\tau,\varrho))\left\{\tfrac{1}{2}[H_x - \mathrm{II}_x(G_x,G_x)] - u/\sqrt{\tau}\right\} \tag{17}$$

uniformly in L^2 over the compact normal-radius bundle, where $G_x = P_x\varepsilon_0$. The added term is $-(k/\sqrt{n})(I - P_p - B_p^*)u$. Its tangent leakage vanishes because $P_x u = 0$ and $P_p \to P_x$. Uniform continuity controls tensor replacement on $\|\varepsilon_0\| \le L$; Gaussian moments control the remainder, bounded after scaling by $C\|\varepsilon_0\|^3/\sqrt{k}$ plus vanishing bounded-offset terms. Take $k \to \infty$ before $L \to \infty$. The tube complement contributes $O(k^2 e^{-ck})$.

The centered curvature term in (17) has zero mean and covariance K_x, so its cross term with deterministic u vanishes. Conditional main noise has exact risk $\{d + \operatorname{tr}((B_p^*)^2)\}/r$. Since $n/r \to 1$, the normalized frequentist risk converges uniformly to

$$d + \operatorname{tr}((B_x^*)^2) + \tau\operatorname{tr}((I-B_x^*)^2 K_x) + \|(I-B_x^*)u\|^2.$$

The game with $L = \tau K_x$ gives worst case $d + f(\tau K_x + S_x^*)$. Taking the maximum over x proves the upper bound. This also proves the claimed rule is valid: all coefficients depend only on Z_0 and known geometry.

B.2.3 Lower bound with normal uncertainty

Fix a chart neighborhood and a smooth compactly supported probability density for a nominal $X \in M$ in that neighborhood. Choose a smooth orthonormal normal frame $\mathcal{E}_x : \mathbb{R}^h \to (T_xM)^\perp$ there. Let $U \in \mathbb{R}^h$ have a fixed finite centered distribution supported in $\|U\| \le \varrho$, covariance S. Independently of X, draw U and put

$$V_n = X + \mathcal{E}_X U/\sqrt{n}.$$

This is an admissible n-dependent prior on $\mathcal{M}_{n,\varrho}$.

Apply Lemma A.1 to this finite mixture. In its buffered graph coordinates $X = h_p(T/\sqrt{k})$, the likelihood exponent for a normal atom z changes by

$$\frac{k}{\sqrt{n}}\langle Z_0 - X, \mathcal{E}_X z\rangle - \frac{k}{2n}\|z\|^2.$$

On bounded T and $\|\varepsilon_0\| \le k^{1/8}$ this is uniformly $o(1)$, since $k/\sqrt{n} = O(1)$ and $\|Z_0 - X\| = O(k^{-3/8}) + O(\|T\|/\sqrt{k})$. Over the chart, its bound $o(1) + C\sqrt{k/n}\|T\|$ is absorbed into the lemma's Gaussian envelope. Outside the chart it is $O(1)$ on the compact prior support, preserving exponential localization. The finite positive mixture weights preserve the buffered normalizer bound.

Consequently, conditionally on Z_0, (T, U) converges in all fixed moments, in prior-predictive probability, to independent $G \sim N(0, I_d)$ and the original U. As in the lemma, exhaust prior-interior buffers after taking the statistical limit. Frame variation is $O(\|T\|/\sqrt{k})$ and vanishes after multiplication by $k/\sqrt{n} = O(1)$.

Let $C_n = \mathrm{Cov}(V_n \mid Z_0)$, written in tangent and normal coordinates at p. The graph expansion and the preceding joint moment convergence give

$$k(C_n)_{TT} \to I_d, \qquad k^{3/2}(C_n)_{TN} \to 0, \qquad k^2(C_n)_{NN} - K_p^{\mathrm{nor}} - S/\tau \to 0 \tag{18}$$

in prior-predictive probability. The leading normal variable is $\mathrm{II}_p(G, G)/2 + \mathcal{E}_p U/\sqrt{\tau}$; centering removes its mean $H_p/2$, and independence removes its mixed covariance. The leading tangent/normal cross covariance vanishes by odd Gaussian moments and $\mathbb{E}U = 0$. The covariance S in (18) is in the chosen normal frame.

The base Schur-complement proof therefore applies with the limiting normal matrix $\tau K_p^{\mathrm{nor}} + S$. It is PSD and may be singular. Hence

$$\mathrm{tr}\{rC_n(I + rC_n)^{-1}\} - d - f(\tau K_X + \mathcal{E}_X S \mathcal{E}_X^T) \to 0$$

in probability and expectation, using the universal bound $0 \le \mathrm{tr}\{rC_n(I + rC_n)^{-1}\} \le N$. Conditional affine square completion lower-bounds $nR_{M,\varrho}$ by (n/r) times this expectation, even for random coefficients, because the independent randomization cannot lower conditional Bayes risk.

Choose x_* attaining the maximum of the continuous function $d + f(\tau K_x + S_x^*)$. For any $\epsilon > 0$, place the nominal prior in a sufficiently small neighborhood of x_* and choose S equal to the maximizing $S_{x_*}^*$ in its normal frame. Realize S by the finite radius-ϱ distribution used in the game. Continuity of curvature and the frame ensures $d + f(\tau K_X + \mathcal{E}_X S \mathcal{E}_X^T) \ge F_{M,\varrho}(\tau) - \epsilon$ on that neighborhood. First take the statistical limit for this fixed prior, then $\epsilon \downarrow 0$. This proves the lower bound and Theorem B.1.

For the stated iterated large-ϱ limit, all normal eigenvalues of K are uniformly bounded. Equation (15) forces $c_x(\tau, \varrho) \to \infty$ uniformly in x. Each of the h fractions tends to one, so $F_{M,\varrho}(\tau) \to d + h = N$.

C Configuration gain and estimation error

We derive the gain from pilot-dependent matrices and bound the cost of estimating their conditional moments. The resulting risk contrasts then support readout comparison and budget selection.

C.1 Configuration gain and the accuracy of a learned readout

Let Z take values in a standard Borel space and let $X \in \mathbb{R}^q$, $U \in \mathbb{R}^d$ have finite second moments, with $q, d \ge 1$. Write $m_X(Z) = \mathbb{E}[X \mid Z]$, $m_U(Z) = \mathbb{E}[U \mid Z]$, and assume $S(Z) = \mathrm{Cov}(U \mid Z)$

is positive definite almost surely. Define

$$C(Z) = \mathbb{E}[(X - m_X(Z))(U - m_U(Z))^\top \mid Z], \quad T(Z) = C(Z)S(Z)^{-1/2}.$$

Write $V = \mathbb{E}\|X - m_X(Z)\|^2$, $\bar{S} = \mathbb{E}S(Z)$ and $\bar{T} = (\mathbb{E}C(Z))\bar{S}^{-1/2}$. Fix $1 \le R \le \min(q, d)$ and let A_R denote best rank-at-most-R Frobenius approximation with deterministic Borel SVD tie breaking (Eckart & Young, 1936). Use measurable moment versions, setting $S = I_d$ on exceptional null sets. The adaptive class contains all Borel predictors $b(Z) + B(Z)U$ with finite coefficients and $\operatorname{rank} B(Z) \le R$ almost surely. The fixed-main class requires constant B but retains arbitrary $b(Z)$. Both allow an independent uniform random seed, with ranks holding jointly almost surely and risks averaging the seed.

Conditional reduced-rank regression (Izenman, 1975) and Eckart–Young approximation give

$$R_{\mathrm{ad}} = V - \mathbb{E}\|T_R\|_F^2, \qquad R_{\mathrm{fix}} = V - \|\bar{T}_R\|_F^2, \qquad G_R = \mathbb{E}\|T_R\|_F^2 - \|\bar{T}_R\|_F^2 \ge 0.$$

Indeed, center X, U conditionally on Z and whiten the main residual by $S^{-1/2}$. The optimal slope is $T_R S^{-1/2}$; for a constant slope, integrate the centered moments first, giving $\bar{T}_R \bar{S}^{-1/2}$. Both classes retain the intercept $m_X - Bm_U$. Measurable SVD attains the optima, $\|T\|_F^2 \le \operatorname{tr}\operatorname{Cov}(X \mid Z)$ ensures integrability, and conditioning on an independent seed leaves them unchanged. Nonnegativity follows by class inclusion.

For learned moments, one perturbation bound below covers errors in both the intercept and slope. With exact means and S, take $\widehat{C} = \widehat{T}S^{1/2}$ to recover the main-text bound for any operator estimate $\widehat{T}$ with finite mean-square error.

Joint moment estimation. Let D be source data independent of (Z, X, U). All expectations below include D; a fixed-main matrix may depend on D but not on Z.

Theorem C.1 (Joint conditional-moment estimation)**.** *Let $\widehat{m}_X, \widehat{m}_U, \widehat{C}, \widehat{S} \succ 0$ be jointly measurable functions of (Z, D) with finite real entries. Suppose, for a fixed $0 \le \epsilon < 1$, almost surely*

$$\|S^{-1/2}(\widehat{S} - S)S^{-1/2}\|_{\mathrm{op}} \le \epsilon.$$

Define the learned readout and its effective centering error by

$$\widehat{B} = (\widehat{C}\widehat{S}^{-1/2})_R \widehat{S}^{-1/2}, \qquad \widehat{X} = \widehat{m}_X + \widehat{B}(U - \widehat{m}_U), \qquad e_m = \widehat{m}_X - m_X - \widehat{B}(\widehat{m}_U - m_U).$$

Let

$$\mathcal{E}_R = \mathbb{E}\|T - T_R\|_F^2, \qquad \zeta = \frac{\|(\widehat{C} - C)S^{-1/2}\|_F + \epsilon\|T\|_F}{1 - \epsilon},$$

$$\Gamma = \mathbb{E}\zeta^2 < \infty, \qquad \eta = \mathbb{E}\|e_m\|^2 < \infty, \qquad \kappa = \frac{1 + \epsilon}{1 - \epsilon}.$$

Then

$$0 \le \mathbb{E}\|X - \widehat{X}\|^2 - R_{\mathrm{ad}} \le \eta + (\kappa - 1)\mathcal{E}_R + 2\sqrt{\kappa}(1 + \sqrt{\kappa})\sqrt{\mathcal{E}_R \Gamma} + (1 + \sqrt{\kappa})^2 \Gamma.$$

The right side is $\eta + (1 + \sqrt{\kappa})^2\Gamma$ if $\operatorname{rank} T \le R$ almost surely. It is $4\sqrt{\mathcal{E}_R \delta} + 4\delta$ when $\widehat{m}_X = m_X$, $\widehat{m}_U = m_U$ and $\widehat{S} = S$ almost surely and one takes $\epsilon = 0$, where $\delta = \mathbb{E}\|(\widehat{C} - C)S^{-1/2}\|_F^2$. No singular-value gap for rank truncation and no absolute eigenvalue floor for S are required.

Proof. Condition on (Z, D). Independence of D leaves the conditional moments of (X, U) unchanged. Completing the square gives, with $\widehat{W} = \widehat{B}S^{1/2}$,

$$\mathbb{E}[\|X - \widehat{X}\|^2 \mid Z, D] = \operatorname{tr}\operatorname{Cov}(X \mid Z) - \|T\|_F^2 + \|\widehat{W} - T\|_F^2 + \|e_m\|^2.$$

This follows by centering both X and U: the centered residual has zero conditional mean, so its cross term with e_m vanishes. Set $E = S^{-1/2}(\widehat{S} - S)S^{-1/2}$, $Q = I + E$, and $\Delta = (\widehat{C} - C)S^{-1/2}$. The definition of $\widehat{B}$ minimizes

$$\operatorname{tr}(B\widehat{S}B^\top) - 2\langle B, \widehat{C}\rangle$$

over matrices of rank at most R. The invertible substitution $W = BS^{1/2}$ changes this objective into $\operatorname{tr}(WQW^\top) - 2\langle W, T + \Delta\rangle$. The unrestricted minimizer is $W_* = (T + \Delta)Q^{-1}$, so $\widehat{W}$ minimizes

$\|(W - W_*)Q^{1/2}\|_F^2$ over the same rank class. All these minima are attained by finite-dimensional SVD; the stipulated Borel tie breaking makes the predictors measurable even at ties.

The spectral bounds $(1-\epsilon)I \preceq Q \preceq (1+\epsilon)I$ and submultiplicativity yield

$$\|W_* - T\|_F = \|(\Delta - TE)Q^{-1}\|_F \le \frac{\|\Delta\|_F + \epsilon\|T\|_F}{1-\epsilon} = \zeta.$$

Use T_R as a feasible competitor in the metric approximation problem:

$$\|\widehat{W} - W_*\|_F \le \frac{\|(\widehat{W} - W_*)Q^{1/2}\|_F}{\sqrt{1-\epsilon}} \le \frac{\|(T_R - W_*)Q^{1/2}\|_F}{\sqrt{1-\epsilon}} \le \sqrt{\kappa}\,\|T_R - W_*\|_F.$$

Writing $a = \|T - T_R\|_F$, the triangle inequality therefore gives $\|\widehat{W} - T\|_F \le \sqrt{\kappa}a + (1+\sqrt{\kappa})\zeta$. Squaring and subtracting a^2 bounds the slope excess by

$$(\kappa - 1)a^2 + 2\sqrt{\kappa}(1+\sqrt{\kappa})a\zeta + (1+\sqrt{\kappa})^2\zeta^2.$$

It is nonnegative because T_R is a best rank-R approximation and $\widehat{W}$ has rank at most R. Conditional regression implies $\|T\|_F^2 \le \operatorname{tr}\operatorname{Cov}(X \mid Z)$, hence $\mathbb{E}a^2 \le V < \infty$. Cauchy–Schwarz bounds $\mathbb{E}(a\zeta) \le \sqrt{\mathcal{E}_R\Gamma}$. The preceding inequalities and $\eta < \infty$ establish integrability of the learned prediction error. Integrating the conditional identity and subtracting R_{ad} proves the bound. The two reductions follow by $a = 0$ or by $\epsilon = \eta = 0$, respectively. □

An estimation-error bound below G_R guarantees improvement over every fixed-main rule of the same rank, including those with arbitrary nonlinear pilot intercepts. Applying this population criterion requires bounds on the moment and relative covariance errors; fitted moments alone do not supply them.

C.2 Risk-difference selection

For nested pilot subsets and fixed total observations, the ideal rank-constrained classes are nested. Moving a main block u_a into the pilot rewrites $b(z) + B_a(z)u_a + B_b(z)u_b$ with intercept $b(z) + B_a(z)u_a$ and main matrix $B_b(z)$, whose rank does not increase. Oracle risk therefore cannot increase with pilot access. Separately fitted moment rules can have different approximation errors, so selecting among them requires their actual risk differences.

A budget decision depends on relative risks. For a finite, nonempty family of fitted readouts indexed by $j \in \mathcal{J}$, let $\mathcal{R}_j$ be their finite target risks and s_j their source-model scores. Fix a reference $0 \in \mathcal{J}$. Define

$$d_j = (\mathcal{R}_j - \mathcal{R}_0) - (s_j - s_0), \qquad t_j = s_j - s_0 + c_j,$$

where c_j estimates the risk-difference error and $c_0 = 0$. This formulation removes any common additive error in the model scores.

Proposition C.2 (Selection from calibrated risk differences)**.** *Suppose the finite real quantities satisfy $|c_j - d_j| \le e_j$ simultaneously for all j, with $e_j \ge 0$ and $e_0 = 0$. Any minimizer $\widehat{j}$ of t_j satisfies*

$$\mathcal{R}_{\widehat{j}} - \min_j \mathcal{R}_j \le \operatorname{osc}_{j\in\mathcal{J}}(d_j - c_j) \le 2\max_j e_j.$$

This holds for any fixed tie rule.

Proof. Write $a_j = d_j - c_j = \mathcal{R}_j - \mathcal{R}_0 - t_j$. For a risk minimizer $j^\star$, the selector obeys $\mathcal{R}_{\widehat{j}} - \mathcal{R}_{j^\star} \le a_{\widehat{j}} - a_{j^\star}$. □

The captured-view experiment estimates c_j on development sequences and tests transfer to other categories. Its sequence-bootstrap intervals describe those cohorts; the proposition specifies the simultaneous error control needed for a population guarantee.

D A GROWING-DIMENSIONAL SEPARATION

This construction proves the growing-dimensional consequence stated in Section 3. The fixed-manifold limits require a separate uniform argument here. Appendix D.1 defines the shared source and target experiment. Appendix D.2 derives fixed-affine risk, using the derivative conditioning established in Appendix D.3. Appendix D.4 supplies orbit separation for global localization. Finally, Appendix D.5 combines localization with posterior covariance bounds to prove pilot-affine risk tending to 3 on the same sequence. The derivative and orbit estimates control different parts of this last step: local estimation and exclusion of distant aliases.

D.1 SOURCE CALIBRATION AND DERIVATIVE PROJECTION BOUNDS

D.1.1 RAW OBSERVATIONS AND THE GEOMETRIC OBJECT

Equip $SO(3)$ with $\langle X, Y\rangle = -\operatorname{tr}(XY)/2$; its distance from the identity is the rotation angle in $[0, \pi]$. The source consists of m independent noisy matrix observations

$$Y_j = g + \Xi_j, \quad j = 1, \dots, m, \qquad g \in SO(3), \tag{19}$$

where every Ξ_j has nine independent standard real Gaussian entries. Equivalently one observes the noisy images of a known orthonormal triad, repeated m times. There are $9m$ scalar measurements. The predictor receives all raw source observations.

Let $U : SO(3) \to O(\mathcal{H})$ be a smooth finite-dimensional real orthogonal representation and f a known unit vector. Write $J_i = dU(X_i)$ for the three orthonormal coordinate rotations and choose $\Lambda \geq 1$ with $\|J_i\| \leq \Lambda$. Put $A_i = J_i/\Lambda$. The target is $U_g f$. We use derivative frames to control the residual of a source-configured projection.

For $q \geq 0$, let $\mathcal{J}_q$ be the span of all words $A_{i_1} \cdots A_{i_k} f$ of length $0 \leq k \leq q$, let $H(q) = \dim \mathcal{J}_q$, and let F_q be the sum of their rank-one forms, including ff^*. Assume

$$\lambda_{H(q)}(F_q) \geq \eta^q \quad (1 \leq q \leq q_*), \qquad 0 < \eta \leq 1. \tag{20}$$

This independently computable condition concerns the representation and template, not the statistical posterior.

D.1.2 HEAT-KERNEL DERIVATIVE BOUNDS

Lemma D.1 (Rotation heat bounds)**.** *Let B_t have the heat law on $SO(3)$ with generator one half the Laplace–Beltrami operator and let k_t be its density with respect to Haar probability. There are fixed $a_0, a_1, C_0, C_1 \geq 1$ such that, for $0 < t \leq 1$,*

$$k_t(R) \leq a_0 t^{-3/2} e^{-d(e,R)^2/(a_1 t)}, \qquad \Pr\{d(e, B_t) > r\} \leq C_0 e^{-r^2/(C_1 t)}.$$

The kernel is strictly positive. If $K = \sum_i J_i^2$, $D_t = e^{tK/2}$ and $C_f(t) = \mathbb{E}(U_{B_t} f)(U_{B_t} f)^*$, then

$$C_f(t) = D_t \left[\sum_{k=0}^{\infty} \frac{(\Lambda^2 t)^k}{k!} \sum_{i_1, \dots, i_k} (A_{i_1} \cdots A_{i_k} f)(A_{i_1} \cdots A_{i_k} f)^* \right] D_t. \tag{21}$$

Consequently, if $a = \Lambda^2 t \leq 1$ and $1 \leq q \leq q_*$, then

$$\lambda_{H(q)}(C_f(t)) \geq (c_h a)^q / q!, \qquad c_h = \eta e^{-3}.$$

Proof. The density bound is the compact three-dimensional specialization of Grigor'yan (1997, Theorem 1.1): the diagonal heat kernel is $O(t^{-3/2})$, and rescaling time for generator $\Delta/2$ and volume to Haar probability only changes constants. Integrating against the rotation-angle density $(2/\pi)\sin^2(\theta/2)$ gives the tail bound. Connected ellipticity gives strict positivity. We next derive the representation-dependent identity and eigenvalue bound.

Ad-invariance makes K commute with every J_i: the structure-constant coefficient tensor is skew in the contracted indices, whereas $J_i J_j + J_j J_i$ is symmetric. Both left and right invariant orthonormal frames have zero divergence and zero self-covariant derivatives under the bi-invariant metric; their

sums of squares equal the same Laplace–Beltrami operator. Therefore the right-invariant Brownian SDE realizes the same heat law. In that realization $V_t = U_{B_t} f$ satisfies $dV_t = \sum_i J_i V_t \, dW_t^i + KV_t \, dt/2$. Ito's formula gives

$$C_f' = \Phi(C_f) + (KC_f + C_f K)/2, \qquad \Phi(C) = \sum_i J_i C J_i^*.$$

The two superoperators commute by Casimir centrality. Exponentiating this finite-dimensional equation yields (21). Its series converges absolutely in trace norm, since the undamped kth trace is bounded by $(3\Lambda^2 t)^k/k!$. For $k \le q$ and $a \le 1$, $a^k/k! \ge a^q/q!$. Thus $C_f(t) \succeq (a^q/q!) D_t F_q D_t$. Since $K \succeq -3\Lambda^2 I$, the least singular value of D_t is at least $e^{-3a/2}$. Applying the variational principle to $F_q^{1/2} D_t^2 F_q^{1/2}$ and (20) proves the eigenvalue bound, using $e^{-3a}\eta^q \ge (\eta e^{-3})^q$. □

Lemma D.2 (Source-configured projection residual)**.** *Put $\alpha = \Lambda^2/m \le 1$. There is a Borel source-only projector P_q onto $U_{\widehat{g}}\mathcal{J}_{q-1}$, of rank $H(q-1)$, satisfying*

$$\sup_g \mathbb{E}\|(I - P_q)U_g f\|^2 \le (54\pi^2\alpha)^q/q!, \qquad 1 \le q \le q_*.$$

Proof. Choose a Borel closest rotation $\widehat{g}$ to $\overline{Y}$ in Frobenius norm, using the measurable minimum theorem on compact $SO(3)$.

Since the true g is feasible,

$$\|\widehat{g} - g\|_F \le 2\|\overline{Y} - g\|_F, \qquad d(\widehat{g}, g) \le \frac{\pi}{\sqrt{2}} \|\sqrt{m}(\overline{Y} - g)\|_F/\sqrt{m}.$$

For a standard nine-dimensional Gaussian vector z, $\mathbb{E}\|z\|^{2q} = \prod_{j=0}^{q-1}(9+2j) \le 9^q q!$. Therefore

$$\mathbb{E}d(\widehat{g}, g)^{2q} \le (9\pi^2/(2m))^q q!$$

uniformly in g. Use the projection onto $U_{\widehat{g}}\mathcal{J}_{q-1}$. A minimum-length geodesic logarithm of $\widehat{g}^{-1}g$ has norm at most $d(\widehat{g}, g)$. The first q Taylor terms of its represented exponential applied to f lie in $\mathcal{J}_{q-1}$. The skew-adjoint integral remainder bounds the residual by $(\sqrt{3}\Lambda d(\widehat{g}, g))^q/q!$. Squaring and applying the Gaussian moment bound gives $(27\pi^2\alpha/2)^q/q!$, which is bounded by the stated expression. A logarithm is needed only pointwise for this inequality, not as part of the measurable rule. □

D.2 Sharp affine prediction risk

D.2.1 Source-configured affine estimators

Use the representation, template, raw source data, derivative frames and constants of Appendix D.1. Write $N = \dim \mathcal{H}$ and add an independent target statistic $Z = U_g f + n^{-1/2}\varepsilon$, $\varepsilon \sim N(0, I_N)$, and set $\alpha = \Lambda^2/m \le 1$.

An allowed linear smoother has form

$$\widehat{v} = A(Y, \omega)Z + c(Y, \omega),$$

where $A \in \mathbb{R}^{N \times N}$ and $c \in \mathbb{R}^N$ are arbitrary Borel functions of the full source data and an independent random seed. They may depend on known budgets, template and representation. They must be chosen without target labels Z. Define

$$\mathcal{R}_{\mathrm{aff}}(n, m, 0) = \inf_{A,c} \sup_g \mathbb{E}\|A(Y, \omega)Z + c(Y, \omega) - U_g f\|^2.$$

Rules of infinite risk can be discarded. This class includes every ridge or kernel-ridge smoother with source-selected features and regularization, as well as source-selected projections. It excludes target-label-dependent tuning and nonlinear fitting of a rotation to target labels. There is no output-rank constraint.

The source-configured projection and heat-frame bounds are Lemmas D.2 and D.1; their constants are independent of growing dimensions and budgets. Given source data Y, classical conditional affine regression gives Bayes risk $\mathrm{tr}\{C(Y)(I + nC(Y))^{-1}\}$ for $C(Y) = \mathrm{Cov}(U_g f \mid Y)$, as in (1). The observation covariance $C + n^{-1}I$ is positive definite even if C is singular; independent randomization does not improve the minimum.

Lemma D.3 (Posterior minorant on an event of probability tending to one)**.** *Give g Haar probability. There are fixed $a \in (0,1]$ and $c_0 \in (0,1]$ such that, for every $B \geq 1$, on the joint true-parameter/data event*

$$E_B = \{\|\sqrt{m}(g_0^T \overline{Y} - I)\|_F \leq B\},$$

the posterior uncentered moment $M(Y) = \mathbb{E}[U_g f f^* U_g^* \mid Y]$ satisfies

$$M(Y) \succeq c_0 e^{-3B^2/2} U_{g_0} C_f(a/m) U_{g_0}^*.$$

The probability of this event is $\pi_B = \Pr\{\chi_9^2 \leq B^2\} \to 1$ as $B \to \infty$, independently of all statistical and representation parameters.

Proof. The sufficient source mean has likelihood proportional to $\exp[-m\|\overline{Y} - g\|_F^2/2]$; centered source residuals are independent ancillary Gaussians. Write $R = g_0^T g$, $x = \|R - I\|_F^2$ and $Z_0 = \sqrt{m}(g_0^T \overline{Y} - I)$. Relative to $R = I$, the likelihood is $\exp[-mx/2 + \sqrt{m}\langle Z_0, R - I\rangle]$. On E_B, Young's inequality bounds it below by $e^{-mx - B^2/2}$ and above by $e^{-mx/4 + B^2}$. For rotation angle θ, $8\theta^2/\pi^2 \leq x \leq 2\theta^2$. The normalizer is at most

$$e^{B^2} \int_0^\pi e^{-2m\theta^2/\pi^2} \frac{\theta^2}{2\pi}\, d\theta \leq C e^{B^2} m^{-3/2}.$$

Thus the normalized density is at least $C^{-1} e^{-3B^2/2} m^{3/2} e^{-2m\theta^2}$. Choose fixed $a \leq \min\{1, (2a_1)^{-1}\}$. Lemma D.1 bounds $k_{a/m}(R)$ above by $a_0 a^{-3/2} m^{3/2} e^{-2m\theta^2}$. Decreasing a fixed c_0 to at most one proves the density minorant. Integrating against the positive rank-one form proves the matrix bound. The event's probability follows from rotational invariance of the nine-dimensional Gaussian source mean. The true g_0 and event need not be observable: the inequality is used only inside a joint Bayes expectation. □

Theorem D.4 (Sharp affine risk under exact source sharing)**.** *Assume the frame condition* (20) *and the two bounds above along a sequence $n \to \infty$. The quantities $N, U, f, \Lambda, m, H, q_*$ may vary, while η and the constants in the heat and projection bounds are fixed. Set $L = \log n$ and let*

$$Q(\log Q + \log(1/\alpha)) = L.$$

Suppose $Q \to \infty$, $Q = o(q_*)$, and, for fixed $a_*, r > 0$, $H(\lfloor xQ \rfloor) = (a_* + o(1))(xQ)^r$ for each fixed $x > 0$. Then

$$\mathcal{R}_{\mathrm{aff}}(n, m, 0) \sim a_* Q^r / n. \tag{22}$$

Proof. Fix $\epsilon \in (0,1)$. For $q = \lceil (1+\epsilon)Q \rceil$, use the projector of Lemma D.2 and output $P_q Z$. Independence gives

$$\mathbb{E}\|P_q Z - v\|^2 = \mathbb{E}\|(I - P_q)v\|^2 + H(q-1)/n.$$

The residual bound is $\exp[-(1+\epsilon)L + o(L)]$: its principal exponent is $q(\log q + \log(1/\alpha))$, with error $O(q + \log(q+1))$. Indeed $Q = o(L)$; fixed-factor perturbation changes the exponent to $(1+\epsilon)L + O_\epsilon(Q)$, and integer rounding adds $O_\epsilon(L/Q + \log Q) = o(L)$. Also $q \leq q_*$ eventually. Thus the upper risk is at most $(a_* + o(1))(1+\epsilon)^r Q^r / n$.

For the lower bound set $q = \lfloor (1-\epsilon)Q \rfloor$ and $k = H(q) \to \infty$. For every posterior second moment M and mean μ, rank-one interlacing yields

$$\lambda_{k-1}(M - \mu\mu^*) \geq \lambda_k(M). \tag{23}$$

Give g Haar probability and choose $B = L^{1/4}$. Lemma D.3 and the heat-frame bound imply on E_B that

$$\lambda_k(M(Y)) \geq \lambda_* := c_0 e^{-3B^2/2} (c_h a \alpha)^q / q!.$$

The same root expansion gives $-\log \lambda_* = (1-\epsilon)L + o(L)$, because $B^2 = o(L)$, and hence $n\lambda_* \to \infty$. By interlacing and conditional affine regression,

$$\mathcal{R}_{\mathrm{aff}} \geq \pi_B \frac{k-1}{n} \frac{n\lambda_*}{1 + n\lambda_*} = (1 - o(1))(k-1)/n.$$

Here $\pi_B \to 1$; the complement contributes nonnegative risk. The lower bound is $(a_* + o(1))(1 - \epsilon)^r Q^r / n$. Taking $n \to \infty$ and then $\epsilon \downarrow 0$ proves the result. □

Scalar-label interpretation. For deterministic $X^*X = nI_N$ with $n \geq N$, labels $y = Xv + \xi$ give sufficient statistic $Z = X^*y/n$. Their orthogonal residual is independent pure noise: averaging over it preserves affinity and reduces squared loss. Since $\|Xw\|^2/n = \|w\|^2$, the sequence and scalar-label minimax risks agree.

D.3 Conditioned derivative growth of generic real templates

D.3.1 Derivative frames of known templates

Specialize Appendix D.1 to the real degree-ℓ irreducible harmonic representation V_ℓ, with $N = 2\ell+1$ and $\Lambda = \ell$. The standard skew-adjoint generators satisfy $\|J_i\| = \ell$; set $A_i = J_i/\ell$ and use its word spaces $\mathcal{J}_q(f)$ and frames $F_q(f)$, including the empty word. All dimensions below are real.

Theorem D.5 (A random known-template class)**.** *Let $\ell \geq 4$, $1 \leq q_* \leq \lfloor \ell/4 \rfloor$, $D = (q_*+1)^2$, and $0 < \delta < 1$. Suppose*

$$N \geq 256D\{D\log 9 + \log(2/\delta)\}. \tag{24}$$

Draw f uniformly from the real unit sphere in V_ℓ. With probability at least $1-\delta$, simultaneously for every integer $0 \leq q \leq q_*$,

$$\dim \mathcal{J}_q(f) = (q+1)^2, \qquad \lambda_{(q+1)^2}(F_q(f)) \geq (1/40)^q.$$

Here λ_j denotes the jth largest eigenvalue. The probability is over a template drawn once and then supplied to the predictor.

The following lemmas establish the uniform frame bounds used in the growing-rotation sequence.

D.3.2 The operator filtration

Equip $\mathrm{End}_{\mathbb{R}}(V_\ell)$ with the normalized Hilbert–Schmidt inner product $\langle S, T\rangle_{\mathrm{av}} = N^{-1}\,\mathrm{tr}(S^*T)$. Conjugation by the orthogonal representation is an orthogonal action on this operator space.

Lemma D.6 (Multiplicity-free operator decomposition)**.** *Under conjugation the real operator space is the orthogonal direct sum*

$$\mathrm{End}_{\mathbb{R}}(V_\ell) = \bigoplus_{L=0}^{2\ell} \mathcal{T}_L, \qquad \dim \mathcal{T}_L = 2L+1,$$

where $\mathcal{T}_L$ is a real degree-L irreducible representation. The span of normalized generator words of length at most $q \leq 2\ell$ is precisely $\mathcal{T}_{\leq q} = \bigoplus_{L=0}^{q} \mathcal{T}_L$ and has dimension $(q+1)^2$.

Let S_q be the operator-space frame formed by all these words, so

$$\mathsf{S}_q = \sum_{k=0}^{q} \sum_{i_1,\dots,i_k} |A_{i_1}\cdots A_{i_k}\rangle\langle A_{i_1}\cdots A_{i_k}|,$$

where bras use $\langle\cdot,\cdot\rangle_{\mathrm{av}}$. For $\ell \geq 4$ and $1 \leq q \leq \ell/4$,

$$\mathsf{S}_q \succeq \frac{1}{5}(3/8)^q \Pi_{\leq q}, \tag{25}$$

with $\Pi_{\leq q}$ the orthogonal operator-space projector.

Proof. The decomposition is the classical Clebsch–Gordan rule for $V_\ell \otimes V_\ell^*$ (Hall, 2015). Its integer-spin summands have real forms and multiplicity one, so their real dimensions are $2L+1$. It remains to identify the word filtration and establish its quantitative frame bound.

The generator triple transforms as degree one under conjugation. The equivariant multiplication map from its $k$fold tensor power has no weights, and hence no highest weight, above k. Every word of length at most q therefore belongs to $\mathcal{T}_{\leq q}$. Conversely write $L_i = iJ_i$ and $L_+ = L_x + iL_y$. In the weight basis,

$$L_+ e_k = \sqrt{\ell(\ell+1) - k(k+1)}\, e_{k+1}.$$

The nonzero operator $(L_+/\ell)^L$ is killed by commutation with L_+ and has commutator weight L under L_z. It is a highest-weight vector in $\mathcal{T}_L$. It is a linear combination of words of length L. That word span is invariant under conjugation, so contains the entire $\mathcal{T}_L$. This proves equality of spans.

Each fixed-length frame in S_q is invariant under conjugation: the adjoint rotation mixes the three generators orthogonally, and its tensor power mixes all ordered words orthogonally. Thus S_q is scalar on each $\mathcal{T}_L$, by multiplicity one and self-adjointness (Hall, 2015) (equivalently apply complex Schur's lemma). For $1 \le L \le q \le \ell/4$, put $B_L = (L_+/\ell)^L$. Restricting its squared Hilbert–Schmidt norm to input weights $-\lfloor \ell/2 \rfloor \le k \le 0$, every ladder factor squared divided by ℓ^2 is at least $3/4$. There are $\lfloor \ell/2 \rfloor + 1 \ge N/5$ such inputs, all within the representation throughout the ladder. Hence

$$\|B_L\|_{\mathrm{av}}^2 \ge \tfrac{1}{5}(3/4)^L.$$

Expanding B_L in $J_x/\ell, J_y/\ell$ gives 2^L ordered words, with coefficients of modulus one. Cauchy–Schwarz on the complexified operator Hilbert space gives $|B_L\rangle\langle B_L| \preceq 2^L \mathsf{S}_q$. Evaluating this inequality at $B_L/\|B_L\|_{\mathrm{av}}$, and using that S_q is scalar on $\mathcal{T}_L$, bounds that scalar below by $\frac{1}{5}(3/8)^L$. The scalar on $\mathcal{T}_0$ is at least one because the identity word has normalized norm one. Taking the smallest bound over $0 \le L \le q$ proves (25). This is an operator-frame statement before choosing f. □

D.3.3 Random evaluation preserves this operator filtration

Lemma D.7 (A random evaluation map)**.** *Let $\mathcal{T}$ be a real D-dimensional subspace of operators on a real N-dimensional irreducible orthogonal representation. Assume $\mathcal{T}$ is invariant under conjugation and contains the identity. Let $T_1, \dots, T_D$ be an orthonormal basis in normalized Hilbert–Schmidt norm. Fix $0 < \delta < 1$. Under the size inequality (24) with these N, D, δ, with probability at least $1 - \delta$ for uniform unit f, the Gram matrix*

$$G_f = (\langle T_a f, T_b f \rangle)_{a,b=1}^D$$

has all eigenvalues between $1/3$ and 3.

Proof. The positive self-adjoint operator $\sum_a T_a^* T_a$ commutes with the representation, because conjugation changes the orthonormal operator basis orthogonally. Irreducibility makes it scalar: each eigenspace of a commuting self-adjoint operator is invariant. Its trace is ND, so

$$\sum_a T_a^* T_a = DI.$$

For a real unit coefficient vector u, write $T(u) = \sum_a u_a T_a$. Then $N^{-1} \operatorname{tr}(T(u)^* T(u)) = 1$ and, by vector Cauchy–Schwarz, $\|T(u)\|_{\mathrm{op}}^2 \le D$.

Draw $g \sim N(0, I_N)$ and let $\widetilde{G}$ be the Gram matrix of $T_a g/\sqrt{N}$. For fixed unit u, the eigenvalues of $T(u)^* T(u)$ satisfy $0 \le \lambda_j \le D$, $\sum_j \lambda_j = N$ and $\sum_j \lambda_j^2 \le ND$. The weighted chi-square inequalities of Laurent & Massart (2000, Lemma 1), with $x = N/(256D)$, therefore give

$$\Pr\{|u^T(\widetilde{G} - I)u| > 1/4\} \le 2e^{-N/(256D)}.$$

Take a $1/4$-net of the real unit sphere in dimension D with at most 9^D points; the disjoint-ball volume argument supplies this size. A union bound and (24) give the deviation bound simultaneously on the net with probability $1 - \delta$. For a symmetric matrix M, approximating a maximizing unit eigenvector by a net point gives $\|M\| \le \sup_{u \text{ in net}} |u^T M u| + \frac{1}{2}\|M\|$. Thus $\|\widetilde{G} - I\| \le 1/2$.

The identity belongs to $\mathcal{T}$ and has normalized norm one, so this same event implies $1/2 \le \|g\|^2/N \le 3/2$. Normalizing $f = g/\|g\|$ therefore puts every eigenvalue of $G_f = (N/\|g\|^2)\widetilde{G}$ between $1/3$ and 3. The zero vector has Gaussian probability zero, and this normalized Gaussian is uniform on the real sphere. □

Proof of Theorem D.5. Apply Lemma D.7 to $\mathcal{T} = \mathcal{T}_{\le q_*}$ of dimension D. On its probability-$1 - \delta$ event, the evaluation map $E_f : T \mapsto Tf$ is injective on this subspace, with squared singular values between $1/3$ and 3. The same lower bound holds on every nested $\mathcal{T}_{\le q}$. Lemma D.6 identifies the image with $\mathcal{J}_q(f)$, proving its rank.

By the definitions of adjoints with the normalized operator-space metric, $F_q(f) = E_f \mathsf{S}_q E_f^*$. For $q \geq 1$, (25) and the nonzero singular values of $E_f \Pi_{\leq q}$ give

$$\lambda_{(q+1)^2}(F_q(f)) \geq \frac{1}{15}(3/8)^q \geq (1/40)^q.$$

The last inequality is $15^{q-1} \geq 1$. For $q = 0$, $F_0 = ff^*$ has its only nonzero eigenvalue equal to one. All statements hold simultaneously on the single evaluation event. □

D.4 Uniform separation and embedding of rotation orbits

Let $U_\ell : SO(3) \to O(V_\ell)$ be the real degree-ℓ irreducible representation, $N = 2\ell+1$, and f a known real unit vector. Rotation angle is $d(e, R) \in [0, \pi]$. A unit-axis infinitesimal rotation has operator norm exactly ℓ, and a rotation of angle θ has character

$$\chi_\ell(\theta) = \sum_{j=-\ell}^{\ell} e^{ij\theta} = \frac{\sin((\ell + 1/2)\theta)}{\sin(\theta/2)},$$

with the continuous value N at zero. The complex weight representation computes the trace of the real orthogonal operator as well.

We use the following consequence of Laurent & Massart (2000, Lemma 1): for $0 \preceq D \preceq 4I$, $\operatorname{tr} D \geq c_* N$, and uniform unit f, $\Pr\{f^* D f < c_*/4\} \leq 2e^{-c_2 N}$, where $c_2 = \min\{c_*/64, 1/10\}$. Apply the weighted chi-square tails to $G^* D G$ and $\|G\|^2$ for $f = G/\|G\|$.

Theorem D.8 (Uniform separation of a generic rotation orbit)**.** *There exist absolute $c, C, c' > 0$ such that, for every integer $\ell \geq 16$, with probability at least $1 - C\ell^3 e^{-c'\ell}$ over a uniform real template f,*

$$c \min\{\ell d(e, R), 1\} \leq \|U_\ell(R) f - f\| \leq 2 \min\{\ell d(e, R), 1\} \quad \textit{for all } R \in SO(3). \tag{26}$$

On this event the stabilizer is trivial and the orbit is a smooth embedded three-dimensional compact submanifold of the unit sphere. The lower probability bound may be vacuous at finite small ℓ; only its asymptotic convergence to one is asserted.

Proof. For a unit axis u, write $A_u = dU_\ell(X_u)/\ell$, so $\|A_u\| = 1$ and $\|A_u - A_v\| = \|u - v\|$. For $r \in [0, \pi\ell]$ define

$$K(u, r) = \frac{e^{rA_u} - I}{\min\{r, 1\}} \quad (r > 0), \qquad K(u, 0) = A_u, \qquad D(u, r) = K(u, r)^* K(u, r).$$

These are real matrices; $\|K\| \leq 2$ by the skew-adjoint integral remainder and unitarity. We claim $\operatorname{tr} D(u, r)/N \geq c_* := 1/5000$ uniformly.

For $0 < r \leq 4$, put $\theta = r/\ell$ and $M = \lfloor \ell/4 \rfloor \geq \ell/8$. Since $j\theta \leq 1$ for $j \leq M$ and $1 - \cos x \geq x^2/4$ for $|x| \leq 1$,

$$\|U_\ell(R) - I\|_F^2 = 4 \sum_{j=1}^{\ell} (1 - \cos(j\theta)) \geq \theta^2 \sum_{j=1}^{M} j^2 \geq \theta^2 \ell^3 / 1536.$$

Divide by $N \min\{r^2, 1\}$, use $N \leq 3\ell$, and obtain a lower bound $1/4608$. The same bound holds at $r = 0$ by continuity. For $r \geq 4$, $\sin(\theta/2) \geq \theta/\pi$, so

$$|\chi_\ell(\theta)|/N \leq \pi/(N\theta) \leq \pi/8, \qquad \operatorname{tr} D/N = 2(1 - \chi_\ell/N) \geq 2(1 - \pi/8) > 1.$$

This proves the claim.

The family D has an operator-norm net of cardinal at most $C_3 \ell^3$ at every fixed positive resolution. Here are explicit regularity bounds ensuring this remains true near the identity. For $r \leq 1$, use

$$K(u, r) = \int_0^1 e^{srA_u} A_u \, ds.$$

Duhamel's formula gives Lipschitz constants at most $3/2$ in u and $1/2$ in r. For $r \geq 1$, $K = e^{rA_u} - I$ has constants at most $\pi\ell$ in u and one in r. The pieces agree at one. Therefore

$$\|D(u, r) - D(v, s)\| \leq 4(\pi\ell + 2)\|u - v\| + 4|r - s|.$$

For any fixed $\delta > 0$, take a Euclidean S^2 net with mesh $\delta/[8(\pi\ell + 2)]$ and a net of $[0, \pi\ell]$ with mesh $\delta/8$. The usual area-packing argument gives at most $C_\delta \ell^3$ pairs; the displayed Lipschitz bound gives a δ operator net. This includes $r = 0$.

Apply this Gaussian quadratic bound to every member of the net with $\delta = c_*/8$ and use a union bound. Except with probability $C\ell^3 e^{-c'\ell}$, every net quadratic form is at least $c_*/4$. Since f has norm one, extension to the full family loses at most δ, so $f^* D(u, r) f \geq c_*/8$ everywhere. Take $c = \sqrt{c_*/8}$. The upper bound is deterministic from $\|K\| \leq 2$.

The lower bound excludes every nonidentity stabilizer. Taking $r \downarrow 0$ shows $\|dU(X_u)f\| \geq c\ell$ for all unit axes, so the orbit map has derivative rank three. A free smooth compact-group orbit is an embedded compact homogeneous manifold, proving the last claim. □

D.5 Pilot adaptation on the growing rotation sequence

D.5.1 Observation model and uniform geometry

We establish uniform bounds for the following growing-dimensional sequence. For every sufficiently large degree ℓ, choose a deterministic unit template f_ℓ from the joint frame and orbit-separation event constructed below. The template is fixed before observations and is known to the estimator. Let $U_\ell : SO(3) \to O(\mathbb{R}^{N_\ell})$ be the real degree-ℓ irreducible representation, $N_\ell = 2\ell + 1$. For an orthonormal basis of $\mathfrak{so}(3)$ under $-\operatorname{tr}(XY)/2$, put $A_i = dU_\ell(X_i)/\ell$. We import exactly the following established properties, with constants independent of ℓ:

$$\|A(t)\| \leq \|t\|, \qquad A(t) = \sum_{i=1}^{3} t_i A_i, \qquad [A_i, A_j] = \ell^{-1} \sum_h c_{ijh} A_h, \tag{27}$$

$$F_j := \sum_{|w| \leq j} (A_w f_\ell)(A_w f_\ell)^T, \quad \operatorname{rank} F_j = (j+1)^2, \quad \lambda_{(j+1)^2}(F_j) \geq \eta^j, \quad 1 \leq j \leq \lfloor \ell^{1/5} \rfloor, \tag{28}$$

$$K_\ell := (\langle A_i f_\ell, A_j f_\ell \rangle)_{ij} \succeq I_3/9, \qquad K_\ell \preceq I_3, \tag{29}$$

$$c \min\{\ell d(e, R), 1\} \leq \|U_\ell(R) f_\ell - f_\ell\| \leq 2 \min\{\ell d(e, R), 1\}, \quad R \in SO(3). \tag{30}$$

Here F_j includes the empty word f_ℓ, $\eta > 0$ and $c > 0$ are fixed, and $d(e, R)$ is the rotation angle. These properties follow directly from the random evaluation and orbit-separation results in the preceding two subsections. To see that they hold simultaneously, set $q_* = \lfloor \ell^{1/5} \rfloor$, $D = (q_* + 1)^2$ and $\delta_\ell = e^{-\ell^{1/5}}$. The size condition in Lemma D.7 holds eventually: its right side is $O(\ell^{4/5} + \ell^{3/5})$, whereas $N_\ell = 2\ell + 1$. On that event the evaluation map $T \mapsto Tf$ has squared singular values in $[1/3, 3]$ on the entire operator filtration through q_*. Theorem D.5 gives (28) with $\eta = 1/40$. Moreover, rotation invariance and the Casimir identity give

$$N_\ell^{-1} \operatorname{tr}\{A(t)^* A(t)\} = \frac{\ell + 1}{3\ell} \|t\|^2.$$

Applying the evaluation lower bound to $A(t)$ gives $t^T K_\ell t \geq \|t\|^2/9$; the generator norm gives the upper bound. Theorem D.8 supplies (30) on an event of probability at least $1 - C\ell^3 e^{-c'\ell}$. A union bound makes the intersection nonempty for every sufficiently large ℓ. Choose one template from it at each degree. Thus one deterministic sequence has all four properties, with constants independent of degree. The generator norm and Casimir identity below belong to the specified irreducible representation.

Set $\ell = \lfloor (\log n)^{10} \rfloor$, $m = \ell^2$, $k = \lfloor n^\beta \rfloor$, $r = n - k$, for integer $n \to \infty$ and fixed $0 < \beta < 1$. Observations, independently conditional on $g \in SO(3)$, are

$$Y_j = g + \Xi_j \quad (1 \leq j \leq m), \qquad Z_0 = U_\ell(g) f_\ell + k^{-1/2} \varepsilon_0, \qquad Z_1 = U_\ell(g) f_\ell + r^{-1/2} \varepsilon_1.$$

All entries of Ξ_j and coordinates of ε_i are independent standard Gaussians. A pilot-affine rule is $A(Y, Z_0, \omega) Z_1 + b(Y, Z_0, \omega)$ with arbitrary Borel coefficients and independent randomization. Its minimax squared coefficient risk is $R_{\mathrm{PA}}(k, r)$. Source processing and pilot processing are unrestricted. Templates and representations are known. The frozen-affine risk $R_{\mathrm{aff}}(n)$ instead restricts dependence on all target labels to be affine, conditional on the source.

At $f = f_\ell$ define $V_j = \operatorname{span}\{A_w f : 1 \le |w| \le j\}$, $V_0 = 0$, and $d_j = \dim V_j$. The normalized Casimir identity is

$$\sum_i A_i^2 = -\frac{\ell+1}{\ell} I. \tag{31}$$

Skew-adjointness and (29) give $d_1 = 3$ and $f \perp V_1$. Equation (31) puts $f \in V_2$, so (28) gives $d_j = (j+1)^2$ for $j \ge 2$ in its range. These spaces agree with nonempty chart derivatives: reordering words changes them by lower-order commutators, which are already included. Let $P_j(g)$ project onto $U_\ell(g)V_j$.

Theorem D.9. *For the above sequence, fix $\beta \in (0,1)$ with $1/\beta \notin \mathbb{N}$ and put $q = \lfloor 1/\beta \rfloor$. Then*

$$nR_{\mathrm{PA}}(k, n-k) \longrightarrow d_q, \qquad d_1 = 3, \quad d_q = (q+1)^2 \ (q \ge 2).$$

The same assertion holds for actual scalar labels under independent-noise fixed orthogonal blocks $X_0^T X_0 = kI_{N_\ell}$ and $X_1^T X_1 = rI_{N_\ell}$, with squared prediction loss under an isotropic test covariate. Both block sizes exceed N_ℓ eventually.

D.5.2 Uniform upper bound without a tubular-radius assumption

Let $\widetilde{g}$ be a Borel global minimizer of $\|Z_0 - U_\ell(g)f\|$, and use

$$\widehat{v} = U_\ell(\widetilde{g})f + P_q(\widetilde{g})\{Z_1 - U_\ell(\widetilde{g})f\}.$$

Existence follows from continuity and compactness; the measurable minimum theorem supplies a Borel selection. The constant-rank projector is Borel, for instance by the Moore–Penrose formula applied to the finite word matrix. No uniqueness of the pilot fit is required. Conditional main-noise variance is exactly d_q/r.

By orthogonality, bounds are uniform in the true g, so take $g = e$. The minimizing property gives $\|U_\ell(\widetilde{g})f - f\| \le 2\|\varepsilon_0\|/\sqrt{k}$. On $\|\varepsilon_0\| \le c\sqrt{k}/4$, (30) implies $\ell d(e, \widetilde{g}) \le 2\|\varepsilon_0\|/(c\sqrt{k}) \le 1/2$. The relative logarithm therefore has a normalized coordinate u with $\|u\| \le 2\|\varepsilon_0\|/(c\sqrt{k})$. The exponential expansion around the fitted point has all its terms of orders $1, \dots, q$ in $U_\ell(\widetilde{g})V_q$. For a skew-adjoint generator, the integral exponential remainder has norm at most $\|u\|^{q+1}/(q+1)!$. Thus the squared residual is at most $C_q\|\varepsilon_0\|^{2q+2}/k^{q+1}$. On the complement it is at most 4, since both coefficients have unit norm and $I - P_q$ is a contraction. Gaussian norm concentration and $N_\ell = o(k)$ bound this complement's probability by $Ce^{-c_1 k}$ eventually. The exact chi-square moments give

$$\sup_g \mathbb{E}_g\|\widehat{v} - U_\ell(g)f\|^2 \le \frac{d_q}{r} + C_q\frac{(N_\ell + 2q)^{q+1}}{k^{q+1}} + Ce^{-c_1 k}. \tag{32}$$

All constants may depend on q, c, β but not on n, ℓ, g. Since N_ℓ is polylogarithmic and $\beta(q+1) > 1$, multiplying the last two terms by n gives zero in the limit. Also $k/n \to 0$. This proves the upper bound in Theorem D.9.

D.5.3 Growing-source posterior localization

Give g the normalized Haar prior and write $C_0 = \operatorname{Cov}(U_\ell(g)f \mid Y, Z_0)$. Conditional affine square completion and left equivariance yield the exact identity, for every finite n,

$$R_{\mathrm{PA}}(k, r) = \mathbb{E}\operatorname{tr}\{C_0(I + rC_0)^{-1}\}. \tag{33}$$

Indeed the posterior mean μ gives optimal coefficients $A = rC_0(I + rC_0)^{-1}$, $b = (I - A)\mu$. This conditional Bayes risk lower-bounds the worst-case risk of any Borel randomized rule. Simultaneous left rotation of the source and both targets preserves the experiment, transforms C_0 orthogonally, and makes this rule equivariant. Its frequentist risk is constant in g, attaining the Bayes lower bound. Compactness gives all required moments. Identity (33) uses no fixed-dimension limit.

For posterior analysis also take true $g = e$, allowed by the same invariance of covariance eigenvalues. Put $\overline{Y} = I + W/\ell$, where W has nine independent standard Gaussian entries, and $Z_0 = f + \varepsilon/\sqrt{k}$. Let $E_\ell = (A_1 f, A_2 f, A_3 f)$ and $a = E_\ell^T \varepsilon$. The coordinates of a have covariance K_ℓ, whose eigenvalues lie in $[\kappa, 1]$ for a fixed $\kappa > 0$ and large ℓ. For fixed L, define

$$\mathcal{E}_n(L) = \{\|a\| \le L,\ \|\varepsilon\| \le \sqrt{N_\ell} + \log n,\ \|W\|_F \le \log n\}.$$

Then $\liminf_n \Pr(\mathcal{E}_n(L)) \geq 1 - \epsilon_L$, where $\epsilon_L \to 0$ as $L \to \infty$.

Choose fixed $0 < \delta < \min\{1, \pi\}/2$ and the chart

$$R(u) = \exp\Big(\sum_i u_i X_i/\ell\Big), \qquad h_\ell(u) = U_\ell(R(u))f = e^{A(u)}f, \quad \|u\| < \delta.$$

Haar density is $c_H \ell^{-3} J(u/\ell)du$, where J is the smooth positive exponential-coordinate Jacobian with $J(0) = 1$, bounded above/below on these balls independently of ℓ. The chart is injective because rotation logarithms are unique there; (30) also implies a trivial stabilizer. The chart on $SO(3)$ and its image orbit are thus one-to-one. Set $T = \sqrt{k}\, u$ in the chart and $T = 0$ on its complement; this defines a posterior random variable everywhere.

Lemma D.10 (Uniform local moments with a moving Gaussian mean)**.** *For every fixed polynomial P in three variables and fixed L, uniformly over observations in $\mathcal{E}_n(L)$,*

$$\mathbb{E}[P(T) \mid Y, Z_0] - \mathbb{E}[P(G_\ell)] \longrightarrow 0, \qquad G_\ell \sim N(K_\ell^{-1}a, K_\ell^{-1}).$$

Moreover the posterior mass outside the chart is at most $C_L \ell^3 k^{3/2} e^{-c_2 k}$ eventually. All fixed polynomial moments of the chart coordinate are uniformly bounded on these events.

Proof. Relative to $R = e$, the source log likelihood is

$$\ell^2(\operatorname{tr} R - 3) + \ell\langle W, R - I\rangle_F.$$

On $R = R(u)$ the first term is nonpositive and $O(\|u\|^2)$; the second is $O(\|W\|_F\|u\|)$. Both vanish uniformly on bounded t-sets with $u = t/\sqrt{k}$ and observations in $\mathcal{E}_n(L)$.

The expansion $h_\ell(u) - f = E_\ell u + R_2(u)$, with $\|R_2(u)\| \leq \|u\|^2/2$, gives pilot log likelihood

$$\sqrt{k}\langle \varepsilon, h_\ell(t/\sqrt{k}) - f\rangle - \frac{k}{2}\|h_\ell(t/\sqrt{k}) - f\|^2 = a^T t - \tfrac{1}{2}t^T K_\ell t + o(1)$$

uniformly on bounded sets, since $(\sqrt{N_\ell} + \log n)/\sqrt{k} \to 0$. The Haar factor $J(t/(\ell\sqrt{k})) \to 1$ uniformly there.

For the whole chart, the chord bound and the same remainder bound make the combined log likelihood at most

$$L\|t\| + \frac{\|\varepsilon\|}{2\sqrt{k}}\|t\|^2 - \frac{c^2}{2}\|t\|^2 + C\frac{\|W\|_F}{\sqrt{k}}\|t\| \leq C_L - c_3\|t\|^2.$$

The bounded-set expansion on $\|t\| \leq 1$ gives an unscaled normalizer at least $c_L \ell^{-3} k^{-3/2}$. Outside the chart, the chord is at least $c\delta$ and the pilot log ratio is at most $-c_4 k$. The global source log ratio is at most $C\ell\|W\|_F = o(k)$, yielding outside mass at most $C_L \ell^3 k^{3/2} e^{-c_2 k}$.

For fixed L, Gaussian means and covariances range over a compact set. The envelope $C_L(1 + \|t\|^{\deg P})e^{-c_3\|t\|^2}$ now gives uniform convergence of numerator and denominator, as in Lemma A.1. The outside contribution is $P(0)$ times the exponentially small mass. This proves the stated fixed-degree moment limits and bounds. □

D.5.4 A uniform lower bound for joint derivative layers

For fixed q, choose any orthonormal column blocks B_j spanning $W_j = V_j \cap V_{j-1}^{\perp}$, and put $B = (B_1, \ldots, B_q)$. Their total number of columns is the fixed d_q once ℓ is sufficiently large. Define the homogeneous vector polynomials

$$p_j(t) = \frac{1}{j!}B_j^T A(t)^j f, \qquad p(t) = (p_1(t), \ldots, p_q(t)).$$

All their coefficient norms are bounded independently of ℓ, since q and the number of generator words are fixed and every normalized generator has norm at most one.

Lemma D.11 (Nondegenerate joint polynomial covariance)**.** *For each fixed q, L, there is $b_{q,L} > 0$, independent of ℓ and of the orthonormal block choices, such that for $\|a\| \leq L$,*

$$\operatorname{Cov}(p(G_\ell)) \succeq b_{q,L} I_{d_q}, \qquad G_\ell \sim N(K_\ell^{-1}a, K_\ell^{-1}).$$

Proof. For $v \in W_1$, the tangent Gram (29) gives $\sum_i |\langle v, A_i f\rangle|^2 \geq \kappa\|v\|^2$. For $v \in W_2$, the first-order terms vanish but the empty word need not. By (31),

$$|\langle v, f\rangle|^2 \leq 3 \sum_i |\langle v, A_i^2 f\rangle|^2.$$

Thus the full frame inequality (28), restricted to V_2, implies $\sum_{|w|=2} |\langle v, A_w f\rangle|^2 \geq \eta^2\|v\|^2/4$. For $j \geq 3$ and $v \in W_j$, all shorter words, including $f \in V_2$, vanish in the inner products. Hence

$$\sum_{|w|=j} |\langle v, A_w f\rangle|^2 \geq \gamma_j \|v\|^2, \quad \gamma_1 = \kappa,\ \gamma_2 = \eta^2/4,\ \gamma_j = \eta^j\ (j \geq 3). \tag{34}$$

In (34), v belongs to V_j, the range of F_j, so the smallest nonzero eigenvalue applies.

When paired with $v \in W_j$, swapping two adjacent generators contributes a length-$(j-1)$ word by (27), which vanishes. Therefore all permutations of each multiplicity type α give the same scalar, and the polynomial $\langle v, A(t)^j f\rangle/j!$ has coefficient $\langle v, A_{w_\alpha} f\rangle/\alpha!$. Its squared coefficient norm is at least $1/(j!)^2$ times the left side of (34). For a linear combination $\sum_j z_j^T p_j(t)$, coefficients at distinct total degrees occupy disjoint entries. Its nonconstant coefficient norm squared is consequently at least $\min_{j \leq q}\{\gamma_j/(j!)^2\} \sum_j \|z_j\|^2$.

The covariance matrix of all nonconstant monomials of degrees $1, \ldots, q$ under a nonsingular Gaussian is positive definite: zero variance would make a nonconstant polynomial constant on all of $\mathbb{R}^3$, since the Gaussian has positive density. These covariance matrices vary continuously in mean and covariance by finite Gaussian moment formulas. For $\|a\| \leq L$ and $\kappa I \preceq K_\ell \preceq I$, their parameters form a compact set, giving a uniform positive smallest eigenvalue. Combining this fact with the coefficient bound proves the result, including cross-degree covariance. Summing isolated layer variances would not suffice. □

D.5.5 Covariance compression and the class lower bound

Let S_k be block diagonal with blocks $k^{j/2} I_{\dim W_j}$. On the chart, all lower-order terms vanish after projection to W_j, and the exponential remainder gives

$$k^{j/2} B_j^T \{h_\ell(T/\sqrt{k}) - f\} = p_j(T) + \rho_j(T), \qquad \|\rho_j(T)\| \leq C_j \|T\|^{j+1}/\sqrt{k}.$$

Lemma D.10 bounds the conditional second moment of the rightmost error by $C_{j,L}/k$. Off the chart, the squared norm of the scaled left side is at most $4k^j$, and its posterior contribution is at most $C_L k^{q+3/2} \ell^3 e^{-c_2 k} = o(1)$. As the coefficient norms of the polynomials are uniformly bounded and their number is fixed, moment convergence implies, uniformly on $\mathcal{E}_n(L)$,

$$S_k B^T C_0 B S_k = \mathrm{Cov}(p(G_\ell)) + o(1)$$

in operator norm. Covariance is unchanged by subtracting the fixed f. Lemma D.11 yields, for sufficiently large n depending on fixed L,

$$B^T C_0 B \succeq \frac{b_{q,L}}{2} S_k^{-2} \succeq \frac{b_{q,L}}{2} k^{-q} I.$$

Compression interlacing therefore gives $\lambda_{d_q}(C_0) \geq b_{q,L} k^{-q}/2$ on $\mathcal{E}_n(L)$. By (33) and nonnegativity of all other eigenvalue contributions,

$$nR_{\mathrm{PA}} \geq \Pr(\mathcal{E}_n(L)) \frac{nd_q}{r} \frac{r b_{q,L} k^{-q}/2}{1 + r b_{q,L} k^{-q}/2}.$$

Since $\beta q < 1$, we have $r/k^q \to \infty$. Taking $n \to \infty$ first gives $\liminf_n nR_{\mathrm{PA}} \geq (1 - \epsilon_L) d_q$. Then $L \to \infty$ proves the lower bound d_q. Together with (32) this proves the sequence assertion in Theorem D.9.

For actual labels with $X_i^T X_i = n_i I$, $Z_i = X_i^T y_i/n_i$ has exactly the corresponding Gaussian law. Orthogonal residual label components contain only independent Gaussian noise, independent of both sufficient statistics and the parameter. Allowing them cannot beat the sequence minimax bound because arbitrary independent randomization is already allowed; averaging main-block residual noise preserves affinity and reduces squared loss. Conversely the displayed rule is implemented from the sufficient statistics. Isotropic prediction loss is coefficient loss. Finally $k, r \gg N_\ell$ gives existence of the required fixed orthogonal block designs eventually. This proves the regression assertion.

Same-sequence removal of the frozen cost. The frame profile holds through $q_* = \lfloor \ell^{1/5} \rfloor \asymp (\log n)^2$. With $m = \ell^2$, $\alpha = 1$ and $Q \log Q = \log n$, we have $Q = o(q_*)$. Theorem D.4 therefore gives $R_{\mathrm{aff}} \sim Q^2/n$ on these same predetermined templates, source observations and target labels. Combining it with Theorem D.9 yields, for each fixed noncritical $\beta \in (0, 1)$,

$$\frac{R_{\mathrm{aff}}(n)}{R_{\mathrm{PA}}(\lfloor n^\beta \rfloor, n - \lfloor n^\beta \rfloor)} \sim \frac{Q^2}{d_q} \longrightarrow \infty.$$

For $\beta > 1/2$, $d_q = 3$: a vanishing pilot fraction removes a diverging normalized frozen cost. The scalar-label interpretation follows from Appendix D.2; its design condition holds since $N_\ell^2/n \to 0$.

E Finite-budget readouts and risk certification

E.1 Finite-budget rule

We construct a finite-sample pilot-affine rule and certify an upper bound on its worst-case risk. Fix $n > k > 0$, $r = n - k$, a known compact parameter set $M \subset \mathbb{R}^2$, and independent observations $Z_0 = v + G_0/\sqrt{k}$, $Z_1 = v + G_1/\sqrt{r}$, with standard Gaussian G_i. A fixed finite rectangular partition $\{C_j\}$ covers $\mathbb{R}^2$, including unbounded edge rectangles. For each cell store a real matrix A_j and vector b_j, and use $\widehat{v} = A_j Z_1 + b_j$ if $Z_0 \in C_j$. Boundary ties can be assigned arbitrarily. This is Borel and affine in the entire main block.

One way to construct the table is to choose a finitely supported prior on M, compute the mean m_j and covariance S_j of V conditional on $Z_0 \in C_j$ whenever that event has positive prior-predictive mass, and set

$$A_j = rS_j(I + rS_j)^{-1}, \qquad b_j = (I - A_j)m_j.$$

Assign arbitrary finite coefficients to zero-mass cells. Prior optimization designs the rule offline. The upper certificate below applies directly to its stored binary coefficients and is independent of optimizer convergence.

E.2 Exact risk and continuum bound

Write $D_j = I - A_j$. At a fixed parameter v the normalized risk is exactly

$$F(v) = n\mathbb{E}_v\|\widehat{v} - v\|^2 = \sum_j p_j(v) g_j(v), \qquad g_j(v) = n\|b_j - D_j v\|^2 + \frac{n}{r}\|A_j\|_F^2. \tag{35}$$

For $C_j = (a_x, b_x] \times (a_y, b_y]$, its probability is

$$p_j(v) = \prod_{h \in \{x,y\}} [\Phi(\sqrt{k}(b_h - v_h)) - \Phi(\sqrt{k}(a_h - v_h))].$$

Thus main noise and pilot noise are integrated analytically, with no simulation or Gaussian quadrature in this identity.

Theorem E.1. *Let $v(\theta)$ be a 2π-periodic C^2 curve. Suppose uniformly in θ, j,*

$$\|v'\| \le V_1, \quad \|v''\| \le V_2, \quad \|D_j\|_{\mathrm{op}} \le d_0, \quad \|b_j - D_j v\| \le B_0, \quad \|A_j\|_F^2 \le a_0.$$

Define

$$G_0 = nB_0^2 + na_0/r, \quad G_1 = 2nB_0 d_0 V_1, \quad G_2 = 2n(d_0^2 V_1^2 + B_0 d_0 V_2),$$
$$H = G_2 + 2G_1\sqrt{k}V_1 + G_0(\sqrt{2}kV_1^2 + \sqrt{k}V_2).$$

For any equally spaced L phases θ_i, including any fixed phase shift,

$$\sup_\theta F(v(\theta)) \le \max_i F(v(\theta_i)) + \frac{H}{8}(2\pi/L)^2.$$

Proof. Write $g(z, \theta) = g_j(v(\theta))$ for $z \in C_j$; cells are independent of θ. Direct differentiation gives $|g| \le G_0$, $|g_\theta| \le G_1$, $|g_{\theta\theta}| \le G_2$. The Gaussian density score for varying θ is

$$s = k(z - v)^T v', \qquad s_\theta = k(z - v)^T v'' - k\|v'\|^2.$$

The second derivative of its integral is $F'' = \mathbb{E}[g_{\theta\theta} + 2g_\theta s + g(s^2 + s_\theta)]$. Differentiation under the Lebesgue integral is valid twice: the curve and its first two derivatives are bounded, the coefficients are a finite bounded table, and on the compact phase domain each integrand derivative is bounded by a constant times $(1 + \|z\|^2)\exp(-k\|z\|^2/4)$, which is integrable. Cell boundaries have Lebesgue measure zero and never move. Under the true pilot law, $s = \sqrt{k}G^T v'$, so $\mathbb{E}|s| \le \sqrt{k}V_1$. If $v' \neq 0$, writing its normalized Gaussian projection as W gives

$$s^2 + s_\theta = k\|v'\|^2(W^2 - 1) + \sqrt{k}G^T v''.$$

Cauchy–Schwarz and $\mathbb{E}(W^2 - 1)^2 = 2$ bound its absolute mean by $\sqrt{2}kV_1^2 + \sqrt{k}V_2$. At $v' = 0$ the same bound follows directly from the remaining term. Hence $|F''| \le H$. A periodic continuous function attains its maximum; at a maximizing phase its derivative is zero. Taylor's integral remainder to a nearest grid point, at distance at most π/L on the circle, gives the result. □

E.3 Nonconvex-curve certificate

The curve is

$$v(\theta) = (\tfrac{23}{20}\cos\theta + \tfrac{3}{20}\cos 3\theta, \tfrac{17}{20}\sin\theta + \tfrac{3}{20}\sin 3\theta).$$

Here $\|v\| \le 1.3$, $V_1 = \sqrt{1.6^2 + 1.3^2}$ and $V_2 = \sqrt{2.5^2 + 2.2^2}$ are valid conservative bounds. The exploratory configuration fixes $n = 1024, k = 16, L = 16384$ and 102 intervals on each pilot axis. The saved endpoints and coefficients are interpreted as exact dyadic real numbers. The verifier checks every entry of A, D has absolute value below 1.000001 and each component of b below 2. Consequently $d_0 = 2.000002$, $a_0 = 4.000009$ and $B_0 = 5.600003$ suffice: $\|b\| < 3$ and $\|v\| \le 1.3$.

The polynomial $g_j(v)$ has six features $(v_x^2, v_x v_y, v_y^2, v_x, v_y, 1)$. Its coefficients, in this order, are

$$\begin{aligned} &n(D^T D)_{11}, \quad 2n(D^T D)_{12}, \quad n(D^T D)_{22}, \\ &-2n(D^T b)_1, \quad -2n(D^T b)_2, \quad n\|b\|^2 + n\|A\|_F^2/r. \end{aligned}$$

Arb enclosures (Johansson, 2017) at 100-bit precision verify every saved coefficient against this exact formula to absolute error 10^{-8}. The stored coefficients have magnitude less than 10000.

At every one of the L exact phases $(2i+1)\pi/L$, Arb trigonometric and error-function enclosures verify each of the six floating features and each of the 204 one-dimensional cell probabilities to absolute error 10^{-12}. Unbounded edge CDF values are exactly zero and one. This includes trigonometric and CDF evaluation error, not just summation error.

To bound floating-point summation error, set $m = 102$, $e_p = e_f = 10^{-12}$, $e_c = 10^{-8}$. Replacing the two exact probability vectors, each of sum one, changes their product in ℓ_1 by at most $2me_p + m^2 e_p^2$. Coefficient and feature replacement therefore contributes at most

$$6(4 + e_f)\{e_c(1 + me_p)^2 + 10000(2me_p + m^2 e_p^2)\} + 6e_f\, 10000 < 0.00005.$$

The two length-m dot products, multiplication by a feature, and final six-term sum, in IEEE double round-to-nearest arithmetic, have error at most

$$\gamma_{2m+16}\, 6(4 + e_f)m^2(1 + e_p)^2\, 10000 < 0.00007, \quad \gamma_q = qu/(1 - qu), \quad u = 2^{-53}.$$

This is the standard absolute dot-product bound (Higham, 2002), which also allows fused multiply-add. Intermediate magnitudes exclude overflow; accumulating underflow errors adds less than 10^{-280} here. Thus the fixed margin 0.001 safely covers both input and arithmetic error. The verifier independently recomputes grid risk directly from A, b.

Applying Theorem E.1 with this margin yields a continuum upper bound for the stored rule. A numerically verified upper below 1.81 improves on both all-label OLS, with normalized risk 2, and main-only OLS, with normalized risk $2n/r = 128/63$. The certificate establishes a finite-budget improvement on the known curve by directly bounding the stored rule, independently of optimizer convergence.

The partition contains 10,404 cells, including 404 unbounded edge cells. Independent confirmation uses 64 fixed phases with 128 pilot/main pairs each, yielding conditional mean risk $1.711 \pm .010$ and actual loss $1.736 \pm .022$; these averages are distinct from the continuous worst-case certificate.

F Simulation protocols and results

The experiments below test the derivative-risk staircase, curvature and normal-error laws, quadratic recovery with known or learned geometry, and source-accuracy requirements for configuration gains. Each subsection specifies the observation model, shared inputs, replication unit and uncertainty calculation alongside the corresponding results.

F.1 Budget exponents on a harmonic curve

We test the risk staircase on $h(t) = (\cos t, \sin t, \cos 2t, \sin 2t)$, whose first three derivative spaces have dimensions $1, 2, 3$. The grid uses $n = 10^j$ for $j = 4, 6, 8, 10, 12, 14, 16$ and $\beta \in \{.28, .30, .36, .40, .45, .55, .65\}$, with $k = \lfloor n^\beta \rfloor$ and $r = n - k$. Each cell contains 512 independent Gaussian pilot observations at each of 16 equally spaced phases, totaling 401,408 pilots. These simulations generate Gaussian sufficient statistics directly: n controls observation precision. Main-block noise is integrated analytically.

Both readouts use the same pilot projection $p = h(\widehat{t})$. Derivative projection retains $q = \lfloor 1/\beta \rfloor$ directions. The quadratic readout is $p + Ta + Ha^2/2 - H/(2r)$, where $T = h'(\widehat{t})/\sqrt{5}$, $H = h''(\widehat{t})/5$, and $a = T^\top(Z_1 - p)$. The comparison at $\beta = .40$ was specified before simulation. Risks are averaged equally over the fixed phases; standard errors reflect independent pilots within each phase. Figure 2a compares the seven exponents at $n = 10^6$ and 10^{16}.

At $\beta = .40$, as n increases from 10^4 to 10^{16}, normalized affine risk decreases from $2.0122 \pm .0003$ to $2.000016 \pm .000001$, while quadratic risk decreases from $1.0275 \pm .0008$ to $1.000016 \pm .000001$ (mean $\pm$ SE). The three derivative-projection levels are visible across the exponent grid. At the smallest pilot budget ($n = 10^4, \beta = .28, k = 13$), occasional distant pilot fits instead produce risk 15.353 ± 7.152; increasing precision localizes the pilot and reveals the third-order level.

Pilot projection uses a 256-point periodic search with Newton refinement, checked against a denser grid and stationary-root solutions.

F.2 Curvature and normal-error experiments

Critical-curvature protocol. We preselect 16 equally spaced circle phases, those 16 plus the four curvature zeros for the nonconvex curve, and a 4×4 torus grid. Each phase has 32 independent pilots and one main block per pilot. The 27 configurations use $n = 1024, 4096, 16384$ and $k = c\sqrt{n}$, $c = 1/2, 1, 2$. The confirmation uses 14,976 pilots independent of exploration data.

Posterior calculations use a uniform prior on each angle (uniform phase, not arclength, on the nonconvex curve). Circular factors use Bessel moments; the curve uses full-period quadrature with refinement and shifted-grid diagnostics. Nearest-point fitting on the curve compares stationary polynomial roots and checks a dense grid.

Normal-offset protocol. For each $n \in \{4096, 16384\}$, set $k = \sqrt{n}$ and test $\varrho \in \{.5, 1, 2\}$. Circle base points use phases $j\pi/8$, $j = 0, \ldots, 15$; the nonconvex curve adds its four curvature-zero phases. At each base point $h(t)$, use both offsets $u = \pm\varrho\,(-h_2'(t), h_1'(t))/\|h'(t)\|$. Torus base angles range over $\{0, \pi/2\}^2$. At $x = (x_1, x_2, x_3, x_4)$, its unit normal axes are $e_1 = (x_1, x_2, 0, 0)$ and $e_2 = (0, 0, 2x_3, 2x_4)$; test $u = \pm\varrho e_1, \pm\varrho e_2$ separately. Every target is $x + u/\sqrt{n}$, so all offsets lie on the radius boundary. The circle, curve and torus have $J = 32, 40, 16$ targets per setting, each with 32 independent pilot/main pairs, totaling 16,896 pairs. Seed 9962026 and separate model, budget, radius and block indices define the random streams.

Given a pilot-fitted affine rule (A, b) at target v_j, integrate main noise to obtain $R_{j\ell} = n\{\|Av_j + b - v_j\|^2 + \|A\|_F^2/(n-k)\}$. Report $J^{-1}\sum_j \overline{R}_j$ with SE $J^{-1}\sqrt{\sum_j s_j^2/32}$, where s_j^2 is the sample variance across pilots; realized-loss summaries use the same weights. Dotted lines average the pointwise limiting risk derived after (17) over these same targets. Both torus curvature eigen-directions are tested, including a direction attaining the worst-offset limit; the displayed equal-direction average evaluates the full specified test set.

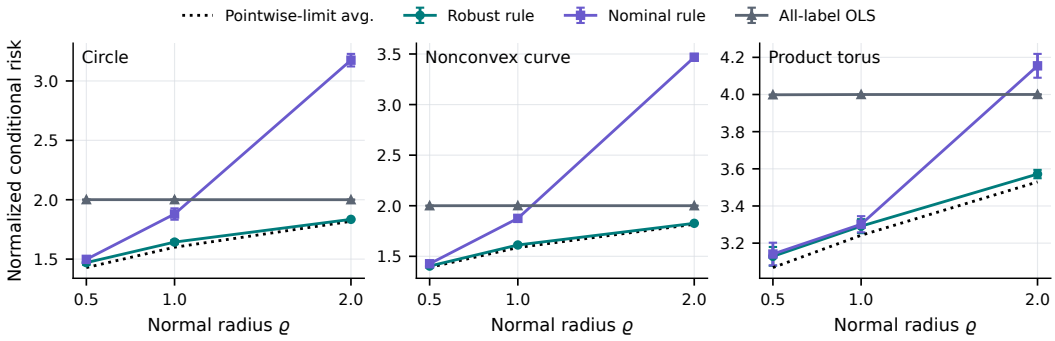

Figure 5: Normal model error at $n = 16384, k = 128$. Conditional fixed-target risks with pilot-sampling SEs. The horizontal axis shows ϱ. The dotted curve averages the pointwise limit over the tested offsets; it need not equal the worst-direction minimax limit.

At $n = 16384, k = 128, \varrho = 2$, the largest estimated target risks are 1.894, 3.798 and 1.900 for circle, torus and curve, respectively; the corresponding same-grid theoretical peaks are 1.818, 3.529 and 1.836. These finite-grid maxima complement the equal-target risk curves in Figure 5; they do not bound the continuous supremum.

Independent high-repetition confirmation. An independent confirmation uses a new seed, 2,048 pilots per target and two budgets with $n/k^2 = 1$: 16 targets each for circle and torus, and 64 equally spaced phases plus four curvature zeros for the curve. All 409,600 pilot sufficient statistics are retained; main noise is integrated analytically. Table 4 reports means and sampled peaks, whose grid and sampling errors are distinct from a continuous supremum. Independent spectral evaluation and sampled main-noise losses check the conditional-risk calculation. Means report pilot-sampling SEs; peaks are maxima of the target means. Same-grid theoretical means are 1.333 (circle), 1.312 (68-point curve) and 3.000 (torus); the curve mean of 1.273 in the main text uses 20 points.

Table 4: Independent fixed-grid confirmation with 2,048 pilots per target.

Model	n	Empirical mean	Empirical peak	Theory peak
Circle	16384	1.347 ± 0.005	1.392	1.333
Circle	262144	1.335 ± 0.004	1.395	1.333
Curve	16384	1.336 ± 0.002	1.586	1.522
Curve	262144	1.316 ± 0.002	1.546	1.522
Torus	16384	3.052 ± 0.008	3.109	3.000
Torus	262144	3.024 ± 0.008	3.084	3.000

F.3 Quadratic recovery with learned geometry

F.3.1 Data, source cost, and held-out tasks

The protocol was fixed before generating observations on CPU. The eight-dimensional construction uses a fixed 8×4 orthonormal matrix U, obtained by QR of a standard Gaussian matrix with seed 100101. The true coefficients are

$$B = [c, U_1/\sqrt{2}, U_2/\sqrt{2}, bU_3/\sqrt{2}, bU_4/\sqrt{2}], \quad c = \text{linspace}(-.15, .15, 8), \quad b \in \{1, .3\}.$$

The feature vector is $\phi(t) = (1, \sqrt{2}\cos t, \sqrt{2}\sin t, \sqrt{2}\cos 2t, \sqrt{2}\sin 2t)^T$, so $\|\phi(t)\|^2 = 5$. Methods learn all coefficients from $Y_j = B\phi(2\pi j/m) + .02\epsilon_j$, with $m \in \{128, 512, 2048\}$ and independent standard Gaussian vectors ϵ_j. The design obeys $\Phi^T\Phi = mI_5$, so $W = Y^T\Phi/m$ and the pointwise source variance is $\nu = 5(.02)^2/m$. The known feature family and known noise variance are part of the experiment. Oracle controls additionally receive B; the other rules receive W and target observations.

Each cell uses 128 independent source cohorts and 16 target phases $2\pi(j + .37)/16$, which do not coincide with source phases. For $n = 4096, 16384$, target statistics have precisions $k = \sqrt{n}$ and $r = n-k$, independently of the source and each other. Target observations are generated as Gaussian sufficient statistics. A source cohort costs $8m$ responses at source noise .02 and is shared across 16 tasks. All methods receive the same source allocation; target noise has unit standard deviation.

Random streams use seed 10120260914 followed by the shape, budget, source-size, and cohort indices and a stream index. Streams 0, 1, and 2 generate source, pilot, and main noise. The 1,536 cohorts and 24,576 target pairs are new observations.

F.3.2 Rules and numerical fitting

Let p minimize distance from Z_0 to the learned curve numerically, let $T = h'(\widehat{t})/\|h'(\widehat{t})\|$, and let

$$H = \frac{h''}{\|h'\|^2} - \frac{h'(h'^T h'')}{\|h'\|^4}, \quad P = TT^T, \quad Q = I - P, \quad K = HH^T/2.$$

All derivatives use the learned coefficients unless the method is explicitly an oracle. The nominal affine rule has $\mu = p + H/(2k)$, $S = (r/k^2)K$ on the normal space, and output $\mu + [P + S(I + S)^{-1}Q](Z_1 - \mu)$. The source-noise-aware affine rule replaces S by $S + r\nu Q$. The nominal quadratic rule is (7); its source-noise-aware version mixes it with Z_1 using weight $r\nu/(1 + r\nu)$. The quadratic weight agrees with Theorem 4.1 for this constant-norm feature family. The experiment uses numerical projections without the theorem's regularity guard; the source-noise-aware affine rule is an empirical extension.

The remaining rules are full-target OLS $\bar{Z}$; projection q of $\bar{Z}$ on the learned curve; joint nonlinear output $(q + n\nu\bar{Z})/(1 + n\nu)$; and oracle nominal affine and quadratic outputs. This joint formula follows by profiling the Gaussian quadratic objective in B at each phase: since $\|\phi(t)\|^2 = 5$, its residual weight is independent of phase. The phase minimization is numerical. The oracle rules are diagnostic and have more information than learned methods.

All fits use a fixed 1,024-point full-period grid, then 24 Newton iterations clipped to .05 radians per step. A 512-point grid confirms numerical stability on the same observations.

F.3.3 Risk estimates and paired comparisons

For realized loss $n\|\widehat{x} - x\|^2$, average the 16 tasks within each source cohort, then report the mean and SE across 128 independent cohorts. Paired 95% intervals use mean ± 1.96 SE and are pointwise across cells. Table 5 gives all 12 settings: source-aware affine (A), source-aware quadratic (Q) and joint nonlinear (J), using the same observations. Every interval favors Q over A and J over Q.

Table 5: Learned-geometry risks and paired differences across all settings.

b	n	m	Affine	Quadratic	Joint	$Q - A$ [95% CI]	$Q - J$ [95% CI]
1	4096	128	1.677	1.431	1.407	-0.246 $[-0.280, -0.212]$	0.024 $[0.009, 0.039]$
1	4096	512	1.368	1.123	1.089	-0.245 $[-0.280, -0.210]$	0.034 $[0.020, 0.047]$
1	4096	2048	1.277	1.012	0.991	-0.265 $[-0.312, -0.219]$	0.020 $[0.007, 0.033]$
1	16384	128	2.573	2.397	2.383	-0.176 $[-0.206, -0.145]$	0.015 $[0.005, 0.024]$
1	16384	512	1.652	1.415	1.399	-0.237 $[-0.275, -0.198]$	0.016 $[0.006, 0.027]$
1	16384	2048	1.343	1.105	1.094	-0.238 $[-0.274, -0.202]$	0.011 $[0.001, 0.020]$
0.3	4096	128	1.847	1.489	1.449	-0.358 $[-0.401, -0.316]$	0.040 $[0.023, 0.057]$
0.3	4096	512	1.593	1.232	1.186	-0.361 $[-0.408, -0.315]$	0.046 $[0.029, 0.063]$
0.3	4096	2048	1.481	1.081	1.032	-0.400 $[-0.450, -0.350]$	0.049 $[0.033, 0.066]$
0.3	16384	128	2.724	2.432	2.402	-0.292 $[-0.329, -0.256]$	0.030 $[0.017, 0.042]$
0.3	16384	512	1.740	1.418	1.403	-0.322 $[-0.360, -0.283]$	0.016 $[0.004, 0.027]$
0.3	16384	2048	1.512	1.124	1.111	-0.388 $[-0.428, -0.348]$	0.013 $[0.002, 0.025]$

Finite-grid readout check. The finite search in Appendix B.1.3 uses $G_k = \lceil 4\pi\sqrt{10k} \rceil$ points (318 or 450), with guards $\|W\|_F \le 6$ and $\sigma_{\min}(W) \ge .05$. Reusing all 1,536 source cohorts and 24,576 target pairs, every guard accepts and mean normalized risks differ from numerical quadratic fitting by at most .001017. Quadratic still improves on affine in all 12 cells, and joint nonlinear remains better

than quadratic. At $b = .3, n = 4096, m = 512$, the paired increase is .001017 [.000270, .001764]; intervals average tasks within source cohorts as above.

F.4 Matched-input quadratic confirmation

The protocol was fixed before generating 32,768 independent Gaussian pilot/main pairs on CPU. Four known models are the unit circle; the planar curve $(1.15\cos t + .15\cos 3t, .85\sin t + .15\sin 3t)$; $(2\cos t, \sin t, .4\sin 2t)$; and the product of circles of radii $1, .7$. Curves use 16 equally spaced phases and the product uses a 4×4 grid. Each of 16 cells has 128 independent repetitions at each phase, with $n \in \{4096, 16384\}$ and $k \in \{\sqrt{n}, \sqrt{n}/4\}$. These are sufficient statistics, not generated raw scalar target labels.

At the same pilot projection p, let $T = [e_1, \ldots, e_d]$ have orthonormal tangent columns, $P = TT^T$, and $H_i = \mathrm{II}_p(e_i, e_i)$. Curves have $d = 1$; on the product torus we use the two circle tangents, for which the mixed second fundamental forms vanish. Write $H = \sum_i H_i$, $K = \frac{1}{2}\sum_i H_i H_i^T$, $S = rK/k^2$, $A = P + S(I + S)^{-1}$, and $\mu = p + H/(2k)$. Compare $\mu + A(Z - \mu)$ with $p + Tc + \frac{1}{2}\sum_i H_i c_i^2 - H/(2r)$, where $c = T^T(Z - p)$, separately for $Z = Z_1$ and $Z = \bar{Z}$. Both quadratic versions retain the same r correction. Their finite risks are measured directly; the aggregate version is connected to the main-block theorem by the full-aggregate argument in Appendix A.3. Controls delete exactly the quadratic term or exactly the centering term, and include both OLS inputs and projection of $\bar{Z}$ (11 rules total).

Circle and product projections are exact block normalizations. Other projections use a 4,096-point full-period grid and 24 Newton steps clipped to .05 radians; a 2,048-point rerun is a stability diagnostic. Seed streams are $[20260914102, \text{model}, \text{budget}, \text{pilot}, \text{stream}]$, with separate pilot and main streams.

For each paired normalized loss difference, the mean averages the fixed phases equally. With within-phase sample variances s_j^2, the standard error is $\sqrt{\sum_{j=1}^{16} s_j^2/128}/16$. All 32 prespecified quadratic-minus-affine comparisons have ordinary and Bonferroni-32 normal-approximation intervals; coverage is approximate. The largest upper endpoint among the 16 square-root-budget Bonferroni intervals is $-.167$. Among the 16 small-pilot contrasts, eight Bonferroni intervals are below zero and eight cross zero. Six small-pilot means are positive, corresponding to three settings on both inputs. The quadratic term improves all 32 mean comparisons against its deletion control. Full-target projection has lower mean in all 32, with 29 ordinary paired intervals excluding zero. These are fixed-grid average risks; all method losses, including OLS comparisons, are available with the saved predictions.

Table 6: Matched-input quadratic-minus-affine normalized risks. Positive values favor affine.

		$k = \sqrt{n}$		$k = \sqrt{n}/4$	
Model	n	Main	Aggregate	Main	Aggregate
circle	4096	-0.346	-0.343	-0.556	-0.533
circle	16384	-0.297	-0.296	-0.842	-0.838
wavy	4096	-0.326	-0.324	2.359	2.381
wavy	16384	-0.346	-0.347	-0.118	-0.119
nonhomogeneous	4096	-0.224	-0.223	0.595	0.617
nonhomogeneous	16384	-0.235	-0.233	-0.633	-0.625
torus	4096	-0.759	-0.748	9.991	10.115
torus	16384	-0.828	-0.826	-1.190	-1.160

F.5 Testing the conditional-moment gain criterion

We isolate the effect of source estimation in a model with known conditional means and input covariance. Let Z have two equiprobable states, $U \mid Z \sim N(0, I_2)$, and $X = T_Z U + \varepsilon$, with independent $\varepsilon \sim N(0, \frac{1}{4}I_2)$. Set $T_0 = \mathrm{diag}(1, b)$ and $T_1 = \mathrm{diag}(b, 1)$, using $b = 0$ or .35 and rank cap one. Each state has m independent source pairs; their average XU^T estimates T_Z. The best fixed rank-one matrix is $[\mathbb{E}T_Z]_1$, even when its intercept may depend arbitrarily on the pilot. Its excess risk over the oracle configured rule is $G_1 = 1 - (1 + b)^2/4$, and the oracle configured risk is $1/2 + b^2$.

Gaussian fourth moments give $\delta = \{3(1 + b^2) + 1\}/m$ and $\mathcal{E}_1 = b^2$, so the sufficient gain bound is available exactly. Using root seed 20260926 and separate streams for each model, sample size and cohort, we use 512 independent source cohorts at each $m \in \{4, 8, 16, 32, 64, 128, 256, 512, 1024, 4096\}$ and integrate test risk analytically. Table 7 reports representative budgets. The first tested budgets with positive guaranteed gains are $m = 32$ and 128 for the two models; the measured gains are $.6276 \pm .0040$ and $.5104 \pm .0010$. Positive gains can occur earlier, showing that the criterion supplies a conservative source-accuracy requirement. Across all 10,240 cohorts, no realized moment-error bound was violated. This experiment tests the gain criterion with exact means and input covariance; the captured-view experiments separately evaluate source-fitted readouts.

Table 7: Source accuracy and configuration gain.

b	m per state	Guaranteed mean gain	Measured mean gain $\pm$ SE
0	8	-1.2500	$.2423 \pm .0203$
0	32	.2500	$.6276 \pm .0040$
0	128	.6250	$.7201 \pm .0009$
.35	8	-2.6738	$-.0296 \pm .0227$
.35	32	$-.5188$	$.4002 \pm .0048$
.35	128	.1493	$.5104 \pm .0010$

The guaranteed column is $G_1 - 4\sqrt{\mathcal{E}_1\delta} - 4\delta$; a negative value leaves the sign of the actual gain undetermined. Standard errors describe variation across source cohorts. Total source size is $2m$.

G Captured-view protocols, results and ablations

Matched fixed controls share the configured readout's source model, observations and rank cap. We first specify the collection protocols, then vary source data, capacity and target allocation.

G.1 Controlled object collections

G.1.1 COIL-20: source models and matched readouts

On COIL-20 (Nene et al., 1996a), all methods share object identity and source views, while target pose is hidden. They predict unobserved pixels from the same pilot and main observations; captured images receive no added noise or synthetic transformations.

Views and observations. The original processed archive contains 20 objects with 72 poses each. We average each 4×4 block of the 128×128 grayscale image to form a 32×32 response, and divide by 255 without imagewise normalization. For every object, even pose indices form the 36-view source set; indices $1 \bmod 4$ form the 18-view validation set and indices $3 \bmod 4$ the 18-view test set. Thus 360 captured test views are excluded from source fitting and parameter selection. Three masks, fixed before evaluation, uniformly select 16 pilot positions and 64 disjoint main positions. The other 944 positions are targets. The same masks apply to all objects and views.

Learning and configuring the response model. Let $\phi_h(\theta) = (1, \cos\theta, \sin\theta, \ldots, \cos h\theta, \sin h\theta)$. Source least squares fits the coefficient map $W\phi_h(\theta)$, with known source angles. A 720-angle uniform quadrature and isotropic working residual variance define a response prior. The variance is the source leave-one-view-out mean squared residual, computed with the exact linear-model leverage formula, lower-bounded by 255^{-2}. Validation selects a multiplier from $\{0.5, 1, 2, 4\}$ and degree $h \in \{3, 6, 9\}$. This variance describes source-model prediction uncertainty; no instrument-noise estimate enters the experiment.

Given the 16 pilot responses, posterior quadrature weights give the feature mean $\bar{\phi}(z)$ and covariance $V_\phi(z)$. For coefficient submatrices W_X, W_U at target and main positions,

$$m_X = W_X\bar{\phi}, \quad m_U = W_U\bar{\phi}, \quad C = W_X V_\phi W_U^T, \quad S = W_U V_\phi W_U^T + \sigma^2 I.$$

The rank-R configured rule is $m_X + [CS^{-1/2}]_R S^{-1/2}(U - m_U)$. The fixed-main control uses the corresponding optimal rank-R matrix from $\mathbb{E}C$ and $\mathbb{E}S$ under the fitted prior and retains the

same nonlinear m_X, m_U. Source-prior averages use 4,096 stratified antithetic pilot draws per object, mask and candidate. Thin QR of W_X computes the full-image rank-constrained objective in feature coordinates exactly. Both rules export a matrix of rank at most R before observing main values.

Validation independently selects shared settings across the 20 objects and three masks for configured, fixed-main, pilot-only and all-input nonlinear rules. Configured and fixed-main both select degree nine and multiplier one at both ranks; consequently their pilot means coincide. Pilot-only selects multiplier 0.5, and the all-input nonlinear reference selects multiplier four. The latter updates quadrature weights using all 80 input positions and belongs to a larger nonlinear-label class. Offline source data comprise 720 complete reference views, and validation uses another 360 views; these costs are separate from the 80 observed positions per target task.

Evaluation and results. We evaluate rank caps two and eight on all 20 objects and three masks. Model settings are selected on validation views before test evaluation. For object j, let $L_{j,a}$ average target-pixel MSE over its 18 test poses and three masks for method a. Reported relative gain against control b is $20^{-1}\sum_j(1 - L_{j,\mathrm{cfg}}/L_{j,b})$; it differs from a ratio of pooled losses. Descriptive 95% intervals use 10,000 object-cluster bootstrap resamples, keeping poses and masks together (Efron, 1979). Table 8 reports mean MSE ($\times 10^3$) and relative gains on the same 360 test views at both ranks.

Table 8: COIL-20 reconstruction at two readout ranks.

Quantity	Primary cap 2	Secondary cap 8
Configured MSE	3.280	3.185
Fixed-main MSE	3.777	3.426
Pilot-only MSE	4.060	4.060
All-input nonlinear MSE	3.142	3.142
Gain over fixed main	14.83% $[10.54, 19.10]$	5.86% $[1.56, 9.39]$
Objects improved over fixed main	19/20	17/20

The smaller configuration gain at rank eight indicates that a larger fixed matrix recovers more useful directions.

Validation-based readout selection We select between the configured and fixed matrices using source validation. For each object, aggregate the matched-fixed minus configured loss across masks for each validation pose. Following the block-resampling approach for dependent data (Lahiri, 2003), take the 2.5th percentile of 10,000 circular block-bootstrap means, with block length three. A positive value selects configuration; otherwise use the matched fixed matrix. This offline choice exports one of the original matrices, so the online rank cap and pilot means are preserved. The rule and threshold were fixed before evaluating new budget cells; the 36-source/64-main anchor reuses the initial COIL-20 evaluation.

G.1.2 COIL-100: transfer across captured objects

We evaluate the fixed readout and validation criterion on COIL-100 (Nene et al., 1996b). The transfer protocol was specified before accessing this collection. Fourier degree nine, the source leave-one-out variance multiplier one, rank caps two and eight, and the validation selection threshold transfer directly from COIL-20. Pilot-only and all-input nonlinear controls retain their previously selected multipliers 0.5 and four. No hyperparameter is selected using COIL-100 losses.

Collection and source-only identity screening. COIL-100 has 100 objects with 72 color views each. Pillow's fixed RGB-to-grayscale conversion precedes the same 4×4 block averaging and division by 255. We retain the COIL-20 pose-index split (36 source, 18 validation, 18 test), three masks, 16 pilot and 64 main pixels, and 944 targets. Object identity and source angles are known; target pose is hidden.

The COIL collections share some objects. Before any validation or test evaluation, we inspected only pose-zero source atlases and fixed a list of 33 clear or possible counterparts, conservatively including

ambiguous cars, cups, wooden shapes and lidded pots. The primary endpoint retains all 100 objects; a prespecified secondary analysis reports the other 67. The saved atlases and mapping document visual similarity rather than independently verified object provenance. Both analyses use exactly the same algorithm and thresholds.

Evaluation. Each object's source views independently determine its response fit, residual variance and 4,096-draw prior integration. The validation rule and all readout choices are fixed before test evaluation. Gains use the COIL-20 object-cluster bootstrap and rank conventions.

Table 9 reports per-object gains (percent, 95% intervals) for configuration and validation selection. Selection retains two negative test gains at each rank; its mean gain is lower than unselected configuration at rank two and higher at rank eight.

Table 9: Configuration and validation-based selection on COIL-100.

	Rank cap 2	Rank cap 8
All 100 objects		
Configured vs fixed	16.30 [13.81,18.66]	7.83 [4.87,10.48]
Selected vs fixed	15.54 [13.18,17.84]	9.21 [7.71,10.82]
Source-screened 67-object subset		
Configured vs fixed	17.44 [14.36,20.47]	9.12 [5.65,12.11]
Objects improved	64/67	60/67
Selected vs fixed	16.01 [12.82,19.14]	9.53 [7.54,11.65]

G.2 Natural captures: models, resolution and readout capacity

Data and confirmation design. We use original RGB captures from the official CO3D single-sequence release and the MVImgNet distribution in SuperGaussian (Shen et al., 2024), selecting sequences by category, name and frame availability before evaluating losses. Development uses eight CO3D sequences from apple, backpack, ball and banana, and twelve MVImgNet sequences from bag, vessel, telephone and pillow. Confirmation uses disjoint categories and sequences: CO3D baseballbat, baseballglove, bench, bicycle, book, bottle, bowl and broccoli (16 sequences); MVImgNet pot, helmet, guitar, earphone, computer keyboard, cap, camera and basket (24 sequences). Thus confirmation evaluates transfer of the calibration settings to new categories, with each sequence's response fitted from its own source frames.

From 72 equally spaced CO3D frames per sequence, we remove constant images and exact duplicates in decoded RGB or grayscale inputs, retaining original time coordinates. The resulting 36–72 frames per sequence retain all categories. Even retained indices form source, $1 \bmod 4$ validation and $3 \bmod 4$ test. The source/validation/test counts are 220/108/108 for development and 439/216/216 for confirmation, with no cross-split content overlap. MVImgNet uses the same index rule: three sequences per category with at least 28 frames, 725 frames total and 169 test frames. BOX downsampling produces 32×32 grayscale frames; the RGB extension preserves aspect ratio. Three fixed masks allocate 16 pilot, 64 main and 944 target pixels. Predictors observe sequence identity and sampled pixels, without target frame positions, camera parameters or object masks.

Conditional residual model. For normalized ordered frame position $t \in [0, 1]$, fit $f(t)$ with features $1, \cos(\pi t), \ldots, \cos(9\pi t)$ using source frames. Let E be their leave-one-out residual matrix and $s^2 = \max\{\mathrm{mean}(E^2), (1/255)^2\}$. We use $\Omega = \alpha E^T E/m + \eta s^2 I$ and a uniform prior on 720 midpoint positions. For pilot, main and target pixel sets P, M, T, the component conditional means and covariances are

$$\mu_A(t, z) = f_A(t) + \Omega_{AP}\Omega_{PP}^{-1}(z - f_P(t)), \qquad V_{AB} = \Omega_{AB} - \Omega_{AP}\Omega_{PP}^{-1}\Omega_{PB}, \quad A, B \in \{M, T\}.$$

Posterior weights are proportional to the Gaussian pilot likelihood with covariance Ω_{PP}. Total expectation and covariance yield the conditional means, $C(z)$ and $S(z)$ in (9). A thin QR factorization computes target moments in the span of the fitted responses and residuals; the final singular-value

truncation enforces the chosen rank cap. The fixed control averages C and S over 4,096 source-predictive pilots and shares the same conditional means.

The grid $\alpha \in \{.1, .5, 1\}$, $\eta \in \{.05, .25, 1\}$ is selected using development validation losses. Configured and fixed methods select independently and choose the same settings: $(.5, 1)$ for CO3D and $(.1, 1)$ for MVImgNet. These settings remain fixed throughout confirmation. For the rank-two confirmation, new validation frames serve the predeclared readout-selection gate, using blocks of three frames and a positive lower 2.5th-percentile gain endpoint. Main-table intervals use 10,000 whole-sequence bootstrap resamples and describe variation within the selected cohorts.

In the 32×32 grayscale control, mean per-sequence configuration gains are 4.38% on CO3D and 10.25% on MVImgNet. Reweighting the same sequence-level gains equally across categories gives 4.50% and 10.25%, respectively; CO3D has unequal sequence counts per category. Table 3 uses sequence weighting. The validation gate yields 4.03% $[2.63, 5.48]$ on CO3D and 10.22% $[7.42, 13.29]$ on MVImgNet.

A post-confirmation component deletion holds the selected η fixed and sets $\alpha = 0$. Calibration reduces MSE relative to this deletion by 8.89% on CO3D and 2.52% on MVImgNet, supporting the role of residual correlation beyond scalar variance adjustment.

Resolution and color controls. On the same 24 MVImgNet sequences, rank-two configuration gains at 16/64 are 8.24% $[5.94, 10.61]$ for 32×32 RGB, 7.69% $[5.66, 9.77]$ for 64×64 grayscale and 7.15% $[5.13, 9.26]$ for 64×64 RGB. Together, these square-resized controls distinguish color and resolution effects; the evaluation below also preserves aspect ratio. At rank eight on 64×64 RGB, source-selected 64/16 increases MSE by 4.13% relative to 16/64; development calibration reduces MSE relative to this source choice by 4.01% $[2.81, 5.33]$. This comparison separates transfer of configuration gains from the accuracy of budget ranking.

Aspect-preserving RGB evaluation. We resize original MVImgNet captures to long edges 256 and 512 with BOX filtering, rounding the short edge to the nearest pixel. There is no crop, stretch or padding. At long edge 512, 22 sequences are 288×512, one is 230×512 and one is 512×288. The same 24 sequences and frame splits enter every setting. Each of three masks maps its 80 original locations through normalized coordinates. At 512, a second setting appends 240 fixed random locations, giving 320 observations and approximately the same sampling density as 256/80. Each observed RGB location supplies all three channels, and rank constrains the joint-channel readout. MSE averages all unobserved channels, then frames, masks and sequences. Table 10 uses MSE $\times 10^3$, rank two and a pilot fraction of $1/5$ for configured and fixed readouts. Linear and nonlinear references use all observed RGB values; they are unrestricted by the rank-two cap.

Source coefficients and residuals are refitted at each resolution with the same degree nine and $(\alpha, \eta) = (.1, 1)$. For $\Omega = FF^T + \nu I$, Gaussian conditioning uses $H = (I + F_P^T F_P/\nu)^{-1}$ in residual coefficient space. Both conditional means and cross moments lie in the span of the fitted responses and residuals. Target Gram matrices evaluate full-image squared loss without forming a target-by-target covariance or dropping pixels. Fixed readouts average 4,096 source-predictive pilots; source risks use 512 antithetic pilots. The Gaussian sufficient statistics give the same posterior law without generating every observed channel separately.

Quality references at larger resolutions. Table 10 gives quality references for the aspect-preserving RGB extension. The all-input nonlinear reference computes its posterior mean from all observed RGB values. The linear reference uses source sample means and covariances with the same ηs^2 diagonal regularizer.

Table 10: Quality references on aspect-preserving RGB frames.

Long edge	Locations	Configured	Fixed	Linear	Nonlinear
256	80	8.723	9.191	8.250	8.775
512	80	10.086	10.602	9.464	10.167
512	320	9.923	10.237	8.447	9.557

Readout capacity. Table 11 varies the rank at 16 pilot and 64 main grayscale pixels. Source fits, residual parameters, masks and targets remain fixed; all-input references share the source model and observations.

Table 11: Readout capacity on 32×32 grayscale captures.

Predictor (MSE $\times 10^3$)	CO3D	MVImgNet
Pilot conditional mean	10.324	3.665
All-input linear	8.309	3.165
All-input nonlinear	8.495	3.076
Configured, rank one	9.453	3.172
Configured, rank two	9.125	3.070
Configured, rank four	8.747	3.000
Configured, rank eight	8.369	2.950
Matched fixed, rank two	9.526	3.326

Selection of reconstruction examples. Figure 3 selects the lower-median object/sequence by aggregate gain, breaking ties by identifier, and its lower-median test frame under mask zero. Observed pixels are copied identically and excluded from losses. Difference maps show fixed minus configured absolute error, averaged across RGB channels, using shared RdBu colors under $\operatorname{asinh}(\Delta/.001)$ for $\Delta \in [-.1, .1]$; ticks report Δ. All metrics use unclipped predictions.

G.3 Source data and model capacity

On COIL-20, take 12, 18 or 36 source views at pose-index spacings six, four or two, and retain the original validation/test split. The smaller source libraries are nested within the 36-view library. Keep 16 pilot pixels and use the first 16, 32 or 64 main positions; all settings predict the same 944 targets, excluding unused reserved inputs.

Validation selects each method's settings from the original grids, restricted to $2h + 1 <$ source count for a defined leave-one-out fit. The allowed degrees are therefore $\{3\}$, $\{3, 6\}$ and $\{3, 6, 9\}$. Table 12 reports mean per-object relative MSE reductions (%) for the 64-main slice. Matched fixed controls share the configured model and pilot means; independent controls select their own model. The validation-selected readout is compared with the matched control.

Table 12: Source-view sensitivity with 16 pilot and 64 main pixels.

	Rank cap 2			Rank cap 8		
Source	Matched	Independent	Selected	Matched	Independent	Selected
12	3.41	3.41	3.29	−0.53	−0.49	0.30
18	13.12	11.97	13.39	4.96	4.96	5.90
36	14.83	14.83	14.81	5.86	5.86	6.92

Fixed-capacity source control. The preceding sweep improves source precision and permits richer models together. To isolate precision, fix $h = 3$, variance multiplier one, the 16 pilot pixels and each rank cap. Table 13 reports the 64-main slice of the same source/main grid. More source views reduce both methods' errors, but the rank-eight configured model remains worse than its matched fixed control. Thus a larger relative configuration gain in the preceding sweep cannot be attributed solely to more accurate fitting.

MSE columns are multiplied by 10^3; gains average paired object-level relative reductions.

G.4 Allocating a fixed target-observation budget

All allocations partition the same observed locations and retain the same target pixels. Source-model scores and validation choices are frozen before test evaluation. Gains average object/sequence-level

Table 13: Fixed-capacity source control ($h = 3$, 16 pilot, 64 main).

Sources	Rank	Configured MSE	Fixed MSE	Gain (%) [95%]
12	2	10.963	11.248	3.41 [1.21,5.47]
12	8	10.696	10.672	-0.53 [-2.43,0.87]
18	2	9.995	10.170	2.45 [-0.07,4.72]
18	8	9.761	9.585	-2.43 [-4.97,-0.45]
36	2	9.517	9.646	2.11 [-0.66,4.59]
36	8	9.331	9.082	-3.33 [-6.02,-1.07]

relative MSE reductions; intervals use 10,000 whole-unit bootstrap resamples. Excess risk is measured relative to each unit's best tested allocation.

G.4.1 Fixed-total allocation on COIL-20

For each original mask, concatenate its 16 pilot and 64 main positions and split this same ordered set at $k \in \{4, 8, 16, 32, 48, 64\}$. Thus the observed set and the 944 target pixels are identical across allocations. We fix 36 source views, degree nine, variance multiplier one and rank two. For each object and mask, 256 stratified antithetic draws from the fitted 720-angle model approximate the expected conditional risk

$$\widehat{\mathcal{R}}(k) = |T|^{-1}\mathbb{E}_Z\big[\mathrm{tr}\{W_T V(Z) W_T^T\} - \|[C(Z)S(Z)^{-1/2}]_2\|_F^2\big].$$

Here $V(Z)$ is the conditional feature covariance and T indexes targets. The fixed control averages covariance over the same source draws. All-input nonlinear prediction is allocation-invariant (maximum MSE discrepancy below 10^{-12}). These scores test the conditional-covariance mechanism on the actual pixel design.

Figure 4 shows the measured curves. Predicted and measured rankings have mean within-object Spearman correlation 0.703 [0.606,0.797]. Source risk selects $k = 64$ for every object, improving on $k = 16$ by 4.90% [0.39,10.57]. Validation selects the common $k = 48$; the source choice gains 0.47% [-0.42,1.91] over it, leaving superiority to validation unresolved.

G.4.2 Fixed-total allocation on natural grayscale captures

Figure 4 uses pilot counts $8, 16, 32, 48, 64$ on the same 16 CO3D and 24 MVImgNet sequences. Rank caps $1, 2, 4, 8$ are evaluated with unchanged source fits, residual parameters, masks and targets. Source risks average 512 antithetic pilots.

At rank two, source risk selects 64/16 throughout. Relative to 16/64, gains are 7.35% $[4.57, 10.03]$ for CO3D and 2.23% $[0.23, 4.36]$ for MVImgNet. Mean within-sequence ranking correlations are 0.844 $[0.656, 0.969]$ and 0.429 $[0.129, 0.712]$. Source-selected excess over each sequence's best tested allocation is 0.52% and 1.30%; relative to per-sequence validation selection, gains are 0.30% $[-0.09, 0.92]$ and -0.57% $[-1.63, 0.35]$.

G.4.3 Development-calibrated RGB allocation

For each setting and rank $R \in \{2, 8\}$, candidate pilot fractions are $1/5, 2/5, 4/5$. Let ρ_{jk} and V_{jk} be the source-predicted and validation risks for development sequence j at fraction k, with reference $k_0 = 1/5$ and $q_j = \rho_{jk_0} > 0$. We estimate the normalized discrepancy

$$c_{jk} = \frac{V_{jk} - V_{jk_0} - (\rho_{jk} - \rho_{jk_0})}{q_j}.$$

Using the twelve development sequences, choose $\lambda \in \{0, .25, .5, .75, 1, 2\}$ by leave-one-sequence-out validation: estimate $\bar{c}_k$ from the other eleven and minimize the held-out validation risk divided by reference risk, with ties toward smaller λ. Refit $\bar{c}_k$ on all twelve and freeze it. Evaluation sequences use

$$\widehat{k} = \arg\min_k \{(\rho_k - \rho_{k_0})/\rho_{k_0} + \lambda \bar{c}_k\}.$$

At ranks two/eight, the selected weights are $2/.75$, $2/.5$ and $0/.25$ for 256/80, 512/80 and 512/320. Evaluation-sequence validation losses are used only for the validation-selection baseline; all methods retain the same source models and frame splits.

At rank eight, calibration reduces mean excess risk from 4.66% to $.20\%$, 5.65% to $.48\%$, and 3.58% to $.37\%$ across those settings. Gains over target-sequence validation are $.83\%$ $[.18, 1.65]$, $.23\%$ $[-.67, 1.04]$ and $.31\%$ $[-.16, .88]$. Rank eight chooses $1/5$ except for two sequences at 512/80, which choose $2/5$ and $4/5$; rank two chooses $1/5, 1/5, 4/5$ across settings. These are empirical cross-category results; Proposition C.2 gives the error condition for a guarantee.

Development-fixed control. A post-hoc control selects a single fraction per setting and rank by mean development validation MSE, with ties toward smaller fractions, using the same frozen candidate losses. Its choices coincide with calibration in five of six setting–rank pairs, producing identical predictions. At rank eight and 512/80, it chooses $1/5$ for all sequences and has MSE 9.552×10^{-3} versus calibrated 9.596×10^{-3}; the calibrated relative gain is $-.32\%$ $[-.94, .00]$. Thus most calibration benefit comes from correcting a shared ranking bias. Pilot-dependent matrix configuration remains isolated by the matched comparisons in Table 3.

G.5 Readout implementation and resource costs

We implement the readouts in NumPy and SciPy (Harris et al., 2020; Virtanen et al., 2020). Factor the configured matrix as DA, with $A \in \mathbb{R}^{2\times 64}$ and $b = m_X - DAm_U$. The acquisition side receives 128 coefficients, accumulates $s = AU$, and returns two scalars for host decoding $b + Ds$. The nonlinear baseline also streams: writing its degree-nine features as $\phi(\theta) = (1, \psi(\theta))$, the likelihood depends on all 80 pixels through $t = W_{O,1:}^T y_O \in \mathbb{R}^{18}$. The constant-feature inner product and $\|y_O\|^2$ cancel in normalized posterior weights, leaving 18 accumulators. Both implementations reproduce the original predictions within 10^{-12} over 1,080 test tasks. The nonlinear variance multiplier is four, its original validation-selected setting.

Table 14: Acquisition-side resources for one 80-pixel task.

Quantity	Configured rank two	Nonlinear sufficient statistic
Main-stage accumulators	2	18
Main-stage MACs per pixel	2	18
Accumulation MACs per task	128	1,440
Coefficient scalars at acquisition side	128	1,440
Upstream scalars per task	$16 + 2$	18
New downstream coefficients per task	128	0

Source fitting and decoding run on the host. The configured device buffers 16 pilot values while the host computes its matrix; the nonlinear device preloads its coefficients.

On an Intel(R) Core(TM) Ultra 7 270K Plus CPU (one thread, float64), median source fitting, leave-one-out variance estimation and static-plan construction take 1.320 ms per object. Median pilot configuration and decoding take 436.2 and 1.20 microseconds per task; nonlinear decoding takes 23.13 microseconds. These CPU timings favor all-input nonlinear prediction; hardware-stream latency remains unmeasured.

Numerical checks. Independent centered-feature, Cholesky and dense-conditioning checks agree within 2×10^{-12} and verify rank caps. Increasing prior integration from 4,096 to 32,768 pilots changes mean gains from 14.826% to 14.812% (COIL-20), 4.378% to 4.376% (CO3D) and 10.249% to 10.251% (grayscale MVImgNet), with unchanged win counts.