

The Statistical Price of Unknown Geometry in In-Context Regression

Jufeng Li

q3529416243@gmail.com

China University of Petroleum, East China

Wenjie Li

China University of Petroleum, East China

Research Article

Keywords: Bayesian prediction, unknown geometry, label feedback, prediction actions, subspace uncertainty, weak identification

Posted Date: October 12th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11317710/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

The Statistical Price of Unknown Geometry in In-Context Regression

Abstract

Unlabelled reference data can reveal the geometry of a regression task, but estimating that geometry and choosing a predictive action are different statistical problems. We study this distinction in Gaussian regression with one labelled observation, a Haar-distributed covariance subspace, a known spectrum and a coefficient prior that transforms coherently with the input coordinates. We establish dimension-uniform finite-sample bounds for the extra Bayes risk caused by unknown geometry, together with a sharp fixed-contrast constant and a complete-regression weak-identification limit. We then optimize over all predictors linear in the observed label, allowing their coefficients to depend on the context input, query and references. The remaining risk is exactly the cost of selecting a same-spectrum geometric action without label feedback, rather than a comparison with a particular estimator. This cost is strictly positive at every finite reference count. Its fourth-order weak-contrast coefficient is positive at intermediate reference information and vanishes at the two information endpoints, even though both optimized total risks decrease with additional references. An observable correction based on reference-posterior moments reduces Bayes approximation regret from fourth to sixth order at fixed dimension. Frozen two-dimensional calculations illustrate the mechanism under equal information. The results identify when label feedback improves geometric prediction and separate information limits from avoidable decision error.

Keywords: Bayesian prediction, unknown geometry, label feedback, prediction actions, subspace uncertainty, weak identification

MSC 2020: 62C10; 62J05; 62H25.

1 Introduction

Unlabelled reference observations are often used to select a covariance metric before a scarce labelled context is used for prediction. This two-stage workflow raises a statistical question: when is geometry recovery enough for prediction, and when should the observed label also influence the geometric action? A label can inform both the regression task and the geometry in which that task is expressed. Optimizing a reconstruction loss for the latent subspace does not, by itself, answer either predictive question.

We give a decision-theoretic analysis in a Gaussian one-label regression experiment. The unknown covariance subspace has a Haar prior, its spectrum is known, and the coefficient prior changes coherently with the input coordinates. The geometry oracle knows the covariance but must still learn the regression task from the same label. The excess risk of the complete-observation Bayes predictor over this oracle therefore isolates the price of missing geometry information. On top of that price, a separate optimized comparison measures the error caused by requiring the predictive coefficient to be chosen without the label.

Predictive information limits.

We establish dimension-uniform finite-sample upper and lower bounds for the geometry information cost, retaining the input, query and label in the lower-bound experiment (Theorem 5). These bounds identify the reference-information crossover and the normalization imposed by one-label prediction. Exact fixed-dimension asymptotics and a symmetric Gaussian matrix-channel limit describe the same risk at finer resolution (Theorems 8 and 9). The distinction matters when allocating reference observations: subspace-estimation accuracy and response-prediction accuracy operate on different scales.

A comparison with an optimized decision class.

We compare full Bayes prediction with *all* label-linear predictors, whose coefficients may use the context input, query and references. The comparison is consequently stronger than a comparison with PCA or a particular shrinkage rule. A normalized-action representation connects these predictors to same-spectrum rank-constrained metrics: the rank constraint can represent the Bayes scalar prediction, but selecting the action without the label incurs a strictly positive finite-reference cost (Theorems 4, 6 and 7). The relevant restriction is the information used to choose the action, rather than its rank alone.

When and how feedback helps.

In weak geometric contrast, the optimized decision-class cost is fourth order. Its explicit covariance coefficient is positive at finite intermediate reference information and vanishes at the two information endpoints (Theorem 10). In this weak-contrast transition, intermediate reference information makes label feedback valuable even as both optimized total risks decrease. A reference-posterior moment correction uses this feedback constructively and has sixth-order Bayes approximation regret, uniformly over reference counts at fixed dimension (Theorem 12). The complete frozen two-dimensional experiment illustrates the cancellation under equal information, without selecting favourable cells.

Position within statistical methodology.

Conditional-mean orthogonality and minimum excess risk (Xu and Raginsky 2022) provide the risk decomposition. Bayesian information inequalities (Wahl 2023; Tsang 2020; Chaumette et al. 2017; Gill and Levit 1995), principal-subspace estimation (Cai et al. 2013) and posterior subspace decisions (Besson et al. 2011; Hoff 2009) provide

tools for analysing geometry uncertainty. Gaussian-mixture regression (Cohn et al. 1996; Sung 2004; Flåm et al. 2012), posterior covariance response (Giordano et al. 2018)¹, and nonlinear-versus-linear MMSE theory (Akyol et al. 2012; Guo et al. 2011) supply the relevant decision mechanisms. Our analysis connects these mechanisms through the actual one-label regression loss: an optimized geometric decision class, its finite-reference strictness, and a fourth-order full-risk transfer with an observable higher-order correction. Together, the results provide a predictive decision framework for geometry adaptation in the one-label regression experiment.

The paper separates three questions: how much geometry information is missing, which observations a decision rule may use, and how closely an implementable rule approximates the resulting Bayes action. Sections 2–5 establish the model and these risk comparisons; Section 6 gives numerical verification; Sections 7–8 explain their use in context-based learning and their statistical scope. Complete technical proofs are in the appendices.

2 Problem setup

Fix integers $d \geq 2$, $1 \leq r < d$, known $\lambda > 0$, and $m \geq 0$. The parameter Π is a Haar-uniform rank- r orthogonal projector. Write

$$\Sigma = I + \lambda\Pi, \quad \alpha = \lambda/(1 + \lambda), \quad A = I - \alpha\Pi, \quad L = I + (\sqrt{1 + \lambda} - 1)\Pi.$$

Given Π , $X, q, Z_1, \dots, Z_m$ are independent $N(0, \Sigma)$, independently of $w \sim N(0, A/d)$. Let $y = X^T w$, $Y_* = q^T w$. Only $O = (X, y, q, Z)$ and the known parameters are visible to a predictor. $S_Z = \sum Z_i Z_i^T$ is sufficient. Truth is never an input to prediction or selection. A useful coupling, used only after fixed-data differentiation, is

$$X = L\xi, \quad q = L\nu, \quad w = L^{-1}\eta/\sqrt{d}, \quad y = \xi^T \eta/\sqrt{d},$$

with independent standard Gaussian innovations, independent of Π ; references use independent innovations.

Proposition 1 (Projection identity) *Put $Q_\Pi = X^T A_\Pi X$ and $g_\Pi(D) = yq^T A_\Pi X/Q_\Pi$, $D = (X, y, q)$. Then $g_\Pi = \mathbb{E}[Y_* | O, \Pi]$, $\bar{g} = \mathbb{E}[g_\Pi | O]$, and for all observable finite-risk f ,*

$$R(f) = \frac{d-1}{d} + V_m(\lambda) + \mathbb{E}(f - \bar{g})^2, \quad V_m = \mathbb{E}(g_\Pi - \bar{g})^2, \quad 0 \leq V_m \leq 1/d.$$

Proof Gaussian conditioning gives mean AXy/Q and covariance $C = (A - AXX^T A/Q)/d$ for w . Independence of references and w conditional on geometry leaves this unchanged. Query averaging yields $\text{tr}(\Sigma C) = (d-1)/d$. Conditional-mean orthogonality twice proves the decomposition. Write $\xi = se$, e uniform on the sphere, $a = e^T \eta/\sqrt{d}$ and $b = e^T \nu$. Then $g_\Pi = ab$, $\mathbb{E}g_\Pi^2 = 1/d$, and comparison with the zero predictor proves the cap. $\square$

¹For further covariance derivative identities, see the preprints by Dytso and Cardone (2021), <https://arxiv.org/abs/2105.05106>, and Dytso, Poor and Shamai (2021), <https://arxiv.org/abs/2104.01883>. These versions are cited as preprints, not as accepted journal articles.

The full density at fixed visible data is

$$p_{\Pi}(O) = (2\pi)^{-(m+2)d/2} (1+\lambda)^{-r(m+2)/2} \times \exp\{-\tfrac{1}{2} \text{tr}[A(S_Z + XX^T + qq^T)]\} \left(\frac{d}{2\pi Q_{\Pi}}\right)^{1/2} \exp\{-dy^2/(2Q_{\Pi})\}.$$

In particular, the conditional label variance is Q_{Π}/d . For any smooth parameter curve with A' and $Q' = X^T A' X$, the full score is

$$-\tfrac{1}{2} \text{tr}[A'(S_Z + XX^T + qq^T)] - \frac{Q'}{2Q} + \frac{dy^2 Q'}{2Q^2}.$$

The determinant is constant along a geometry curve, not along a spike-strength curve.

For $d = 2, r = 1$ let $t = \frac{1}{2} \log(1 + \lambda)$, $\kappa = e^t$. The symmetric covariance has eigenvalues κ, κ^{-1} . Then $\Sigma_{\text{here}} = \kappa \Sigma_{\text{sym}}$ and $A_{\text{here}} = A_{\text{sym}}/\kappa$. Map X, q, Z from the symmetric two-dimensional parametrization by multiplication by $\sqrt{\kappa}$, and w by division by $\sqrt{\kappa}$. Both labels and targets are unchanged, and the task prior becomes exactly $A_{\text{here}}/2$. This known invertible observation map proves equality of Bayes risks, not just agreement of a formal covariance.

2.1 The prediction metric

For fixed data let $N = q^T A X$, $Q = X^T A X$. Along any curve of geometry,

$$\dot{g} = y(N'/Q - NQ'/Q^2), \quad \ddot{g} = y(N''/Q - NQ''/Q^2 - 2N'Q'/Q^2 + 2N(Q')^2/Q^3).$$

Whitening is substituted only after these derivatives.

Lemma 2 (Spherical fourth moments and a bilinear prediction form) *Let e be uniform on S^{d-1} , $a \sim N(0, 1/d)$, and $v \sim N(0, I)$ independently. For symmetric H, K define $h_H = av^T(I - ee^T)He$. Then*

$$\mathbb{E} h_H h_K = \frac{d \text{tr}(HK) - \text{tr } H \text{tr } K}{d^2(d+2)}. \quad (\text{M})$$

Proof Condition on e . Integrating a, v gives $d^{-1}\{e^T H K e - (e^T H e)(e^T K e)\}$. Gaussian polar decomposition shows $\mathbb{E} e_i e_j = \delta_{ij}/d$ and $\mathbb{E} e_i e_j e_k e_l = (\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})/[d(d+2)]$: indeed, a standard Gaussian equals an independent radius times e , its fourth moment is the numerator, and the radius has fourth moment $d(d+2)$. Substitution proves (M). As a second route, diagonalise H for the squared case. The squared coordinates of e have the Dirichlet($1/2, \dots, 1/2$) distribution, giving $\mathbb{E} e_i^4 = 3/[d(d+2)]$, $\mathbb{E} e_i^2 e_j^2 = 1/[d(d+2)]$ for $i \neq j$. Expanding the diagonal quadratic form proves the same squared identity; polarization gives the bilinear version. Neither route assumes a prediction/projector reduction. $\square$

At $\Pi_0 = \text{diag}(I_r, 0)$ choose the cross-plane generator $G_{ij} = e_j e_i^T - e_i e_j^T$, $i \leq r < j$, and the curve $\Pi_s = e^{sG_{ij}} \Pi_0 e^{-sG_{ij}}$. Put $\delta_\lambda = \lambda/\sqrt{1+\lambda}$. Then $\dot{\Pi} = E_{ij} + E_{ji}$,

$A' = -\alpha(E_{ij} + E_{ji})$, and $LA'L = -\delta_\lambda(E_{ij} + E_{ji})$. The raw derivative becomes $\dot{g} = a\nu^T(I - ee^T)(LA'L)e$. For one reference row the score is $\alpha Z_i Z_j = \delta_\lambda \zeta_i \zeta_j$. Therefore, in the $p = r(d - r)$ cross-plane coordinates,

$$I_Z = \delta_\lambda^2 I_p, \quad J_g = \mathbb{E}[\nabla g \nabla g^T] = \frac{2\delta_\lambda^2}{d(d+2)} I_p. \quad (\text{T})$$

Independence and odd Gaussian moments show that off-diagonal Fisher entries vanish; (M) gives the prediction entries. This computes the local ratio J_g/I_Z , not by itself an asymptotic theorem.

For positive definite M independent of D conditional on true geometry, write $B = LML$. Integrating a and the query Gaussian proves the exact finite metric identity

$$\mathbb{E}[(g_M - g_\Pi)^2 \mid \Pi, M] = \frac{1}{d} \mathbb{E}_e \left[\frac{e^T B^2 e}{(e^T B e)^2} - 1 \right]. \quad (\text{F})$$

It is generally not a scalar multiple of a Frobenius projector distance at finite spike. For nearby equal-rank projectors at fixed λ , the tangent expansion of (F) is $\delta_\lambda^2 \left\| \hat{\Pi} - \Pi \right\|_F^2 / [d(d+2)] + o(\left\| \hat{\Pi} - \Pi \right\|_F^2)$. A D -selected metric does not meet the independence condition in (F). For $M = I$, if $B_* = e^T \Pi e \sim \text{Beta}(r/2, (d-r)/2)$,

$$\mathbb{E}(g_I - g_\Pi)^2 = \frac{\lambda^2}{d} \mathbb{E} \frac{B_*(1 - B_*)}{(1 + \lambda B_*)^2}.$$

Its weak-spike coefficient is $r(d-r)/[d^2(d+2)]$.

2.2 Why the coefficient prior matters

The main prior comes from a coordinated change of variables: $X = L\xi$, $q = L\nu$, $w = L^{-1}\eta/\sqrt{d}$. Then $y = \xi^T \eta/\sqrt{d}$ and $Y_* = \nu^T \eta/\sqrt{d}$ are unchanged by that coordinate transformation.

Remark 3 (A geometry-insensitive task prior) *Keep the same inputs but replace only the coefficient prior by $w \mid \Pi \sim N(0, I/(d+r\lambda))$. The target variance is still one. Gaussian conditioning now gives*

$$g_\Pi(D) = y \frac{q^T X}{\|X\|^2},$$

independent of Π , so the geometry-oracle excess V_m is zero for every m . The oracle still does not know w . Its risk is instead

$$R_{\text{oracle}}^{\text{iso}} = 1 - \frac{1 + \lambda(1 + \lambda)\mathbb{E}[T/(1 + \lambda T)]}{d + r\lambda}, \quad T \sim \text{Beta}(r/2, (d-r)/2).$$

Indeed the conditional coefficient covariance is $[I - XX^T/\|X\|^2]/(d+r\lambda)$, and averaging the query gives $1 - \mathbb{E}[X^T \Sigma X/\|X\|^2]/(d+r\lambda)$. Writing $X = LRe$ gives the displayed Beta expression, since $X^T \Sigma X/\|X\|^2 = [1 + (2\lambda + \lambda^2)T]/(1 + \lambda T)$. This comparison isolates the role of the task prior: coordinated input-coefficient geometry produces a prediction information cost, whereas an isotropic coefficient prior makes the conditional mean geometry-insensitive. Each experiment has its own oracle benchmark.

2.3 An exact prediction representation

Write $g_S = yq^T(I - \alpha S)X/[X^T(I - \alpha S)X]$ for any $0 \preceq S \preceq I$, where $\alpha = \lambda/(1 + \lambda)$.

Theorem 4 (The prediction-action ball) *For $d \geq 2$, $1 \leq r < d$, $0 < \alpha < 1$ and any unit e ,*

$$\left\{ \frac{(I - \alpha Q)e}{e^T(I - \alpha Q)e} : Q^2 = Q = Q^T, \text{tr } Q = r \right\} = \left\{ e - \alpha v : v \perp e, \|v\| \leq (2\sqrt{1 - \alpha})^{-1} \right\}.$$

There is a Borel measurable projector selection for every vector in this ball. Consequently, a full-observation rank- r metric $Q(O)$ exists with $g_{Q(O)} = \bar{g}$ almost surely. If $M_Z = \mathbb{E}[\Pi | Z]$, a selection $Q(Z, X)$ satisfies $g_{Q(Z, X)} = g_{M_Z}$ pointwise for every label and query, at fixed Z, X .

Appendix K proves the full image, measurable selection and two representations. The radius is consistent with the classical Kantorovich turning-angle inequality (Rao 2007); filling the entire image and preserving the one-label normalized action are the relevant steps here. The represented object is the posterior scalar conditional mean. The selected metric is prediction-equivalent for the observed context, with geometric identification and predictive uncertainty remaining separate tasks. Its information dependence distinguishes $Q(O)$ and $Q(Z, X)$ from reference-only rules $Q(Z)$.

3 Finite-sample information limits

Write $b = 1 + \lambda$, $\alpha = \lambda/b$, $p = r(d - r)$, and $h = \lambda^2/b$.

Theorem 5 (Uniform bounded-spike result) *For every finite $\Lambda > 0$, simultaneously for all integers $d \geq 2$, $1 \leq r < d$, $m \geq 0$, and $0 < \lambda \leq \Lambda$,*

$$c_\Lambda \frac{p}{d(d+2)} \min \left\{ \frac{h}{d}, \frac{1}{m+1} \right\} \leq V_m \leq C_\Lambda \frac{p}{d(d+2)} \min \left\{ \frac{h}{d}, \frac{1}{m+1} \right\}.$$

Explicit constants are

$$K_\Lambda = 2(1 + \Lambda) + \frac{11\Lambda^2}{4(1 + \Lambda)}, \quad c_\Lambda = \frac{2}{K_\Lambda + 1}, \quad C_\Lambda = 20480(1 + \Lambda)^4.$$

The lower bound concerns every square-integrable scalar decision based on the complete observation O , not just reference-only subspace decisions. No growing-dimensional limit is used in this theorem.

The reference scale $B = p \min\{h/d, 1/(m+1)\}/[d(d+2)]$ is a rate benchmark, not an exact Bayes-risk curve. The transition is at mh of order d . In particular its prediction normalization is essential: the cap $V_m \leq 1/d$ rules out an unnormalized dimension-uniform projector rate.

The lower proof constructs the full-observation witness

$$C_O = \sum_a \{D_a^2 g + (D_a g) D_a \log p_\Pi(O)\}, \quad J_* = \mathbb{E} \sum_a (D_a g)^2 = \frac{2ph}{d(d+2)}.$$

An explicit contraction gives $H = \mathbb{E}C_D^2 \leq K_\Lambda d J_*$. Haar integration by parts then yields

$$V_m \geq \frac{J_*^2}{H + mhJ_*} \geq \frac{2ph}{d(d+2)(K_\Lambda d + mh)}.$$

The label score is part of C_D . The argument applies to every finite-risk predictor using O , not only to reference-based projector estimators. The complete algebra and Fubini justification are in Appendix A.

For the upper bound use the isotropic prediction unless both $m \geq d$ and $mh \geq d$; on that branch use the top- r empirical covariance projector from Z only. Independent-projector prediction comparison, an elementary Gaussian covariance bound, and the global Ky–Fan inequality prove the upper rate and cover $m = 0$, $m < d$, edge ranks, and the crossover without hidden signal assumptions. See Appendix B.

3.1 The cost of choosing geometry before reading the label

In this decision comparison put $s = \|X\|^2$, $e = X/\sqrt{s}$. Let $W = (X, q, Z)$ and $\mathcal{L} = \{y\beta(W) : \beta \text{ measurable, finite risk}\}$. This class already permits query- and input-dependent coefficients; it is larger than the class of reference-only metric rules. Write

$$\tau = \mathcal{L}(\Pi \mid W), \quad v_P = X^T A_P X / d, \quad c_P = q^T A_P X / d.$$

The two posteriors are explicitly

$$\tau(dP) \propto e^{\alpha \operatorname{tr}[P(S_Z + X X^T + q q^T)]/2} dH(P), \quad \tau_y(dP) \propto v_P^{-1/2} e^{-y^2/(2v_P)} \tau(dP). \quad (1)$$

The label may change the geometric action through its amplitude, despite the response being noiseless. Selection timing means a restriction on observable-variable dependence; it does not require queries to arrive earlier than labels. A label-blind geometry still uses the label in its final prediction.

Theorem 6 (Exact optimized comparator) *The best label-linear predictor is*

$$f_{\text{lin}} = y\beta(W), \quad \beta = \frac{\mathbb{E}_\tau c_P}{\mathbb{E}_\tau v_P} = \frac{q^T (I - \alpha M_W) X}{X^T (I - \alpha M_W) X}, \quad M_W = \mathbb{E}(P \mid W).$$

Define $N_m = R(f_{\text{lin}}) - R(\bar{g})$. Then

$$N_m = \mathbb{E}_W \mathbb{E}_{y|W} \left[y^2 \{ \mathbb{E}_{\tau_y} (c_P / v_P) - \beta \}^2 \right], \quad (2)$$

$$\inf_{Q(W)} R(g_Q) = R(f_{\text{lin}}), \quad \inf_{Q(O)} R(g_Q) = R(\bar{g}). \quad (3)$$

The infima in (3) are over measurable rank- r projectors and are attained. Thus N_m is exactly the optimal label-blind versus label-adaptive geometry cost, and a lower bound on the optimized reference-only cost.

The ratio of posterior moments in the theorem is essential: the optimal coefficient is $\mathbb{E}_\tau c / \mathbb{E}_\tau v$, not $\mathbb{E}_\tau (c/v)$. These coincide only under additional cancellation. The reference-only optimum need not equal the label-blind optimum.

Theorem 7 (Strictness in the actual Haar model) *For all $d \geq 2$, $1 \leq r < d$, $\lambda > 0$, and every finite $m \geq 0$, $N_m > 0$.*

Appendix E gives the conditional normal equation, an exact variance-level cancellation criterion, and a positive-probability Haar-model witness. Consequently,

$$R(f_{\text{lin}}) = R_{\text{oracle}} + V_m + N_m, \quad R(f) = R(f_{\text{lin}}) + \mathbb{E}(f - f_{\text{lin}})^2 \quad (f \in \mathcal{L}).$$

The information cost V_m and the timing cost N_m answer different questions. Restricting a metric to the same rank and spectrum does not itself force a predictive loss; restricting when that metric may use the label does.

4 Sharp asymptotics

This result fixes d, r, λ before m tends to infinity.

Theorem 8 (Fixed-spike predictive asymptotics) *For every fixed d, r as above and fixed $\lambda > 0$,*

$$\lim_{m \rightarrow \infty} mV_m(\lambda) = C(d, r) := \frac{2r(d-r)}{d(d+2)}.$$

For $(d, r) = (2, 1)$ the constant is $1/4$; for $(4, 2)$ it is $1/3$. The Fisher information per reference tangent direction is h , while the prediction sensitivity is $2h/[d(d+2)]$. Their ratio suggests the constant but is not itself a proof. Appendix C sandwiches the actual complete-observation Bayes risk between a full-context information lower bound and the attainable independent PCA prediction error. A global fourth-moment bound controls expectation convergence, with no posterior BvM assumption.

5 Weak-identification transition

This section fixes dimension and rank and identifies how reference information translates into complete regression risk. The resulting channel also distinguishes reference-only geometry decisions from context-adaptive prediction.

Theorem 9 (Full regression-risk limit) *For every sequence $\lambda_j \downarrow 0$, $m_j \rightarrow \infty$ with $m_j \lambda_j^2 \rightarrow \rho \in (0, \infty)$,*

$$\frac{V_{m_j}(\lambda_j)}{\lambda_j^2} \rightarrow \mathcal{F}_{d,r}(\rho) := \frac{1}{d(d+2)} \mathbb{E} \|\Pi - M_\rho(Y)\|_F^2, \quad Y = \sqrt{\rho}\Pi + G, \quad M_\rho(Y) = \mathbb{E}[\Pi \mid Y]. \quad (\text{L})$$

Here G is symmetric; its upper-triangular entries are mutually independent with $G_{ii} \sim N(0, 2)$ and $G_{ij} \sim N(0, 1)$ for $i < j$. The endpoint $\rho = 0$ holds for $m_j \lambda_j^2 \rightarrow 0$, even if m_j is bounded, with $\mathcal{F}_{d,r}(0) = r(d-r)/[d^2(d+2)]$.

The posterior is taken over the Haar projector and uses

$$M_\rho(Y) = \frac{\int P \exp\{\sqrt{\rho} \operatorname{tr}(PY)/2\} d\mu(P)}{\int \exp\{\sqrt{\rho} \operatorname{tr}(PY)/2\} d\mu(P)}.$$

The trace direction is ancillary for P and may be removed; diagonal entries of the centered noise then have covariance $2(\delta_{ij} - 1/d)$ and are not independent. The posterior matrix mean retains subspace uncertainty. Projecting it back to rank r yields a reference-only decision (Besson et al. 2011); Theorem 19 quantifies its predictive cost, and Theorem 4 describes the additional freedom supplied by context adaptation. At zero signal, its Frobenius risk is p/d , whereas a fixed rank- r projector has risk $2p/d$.

Appendix D proves joint convergence, a full-orbit auxiliary-posterior perturbation bound, fixed-data Taylor expansion, cancellation of the radial denominator, and square-risk expectation convergence through an explicit fourth-moment envelope. It also verifies the zero-information endpoint and exact two-dimensional normalization. The result is a fixed-dimensional continuous transition indexed by ρ .

5.1 The additional value of label-adaptive geometry

The quadratic scale of V_m quantifies the information left in the reference experiment. A different, smaller scale quantifies how much of that information an optimized label-linear predictor leaves unused. For a probability μ on projectors and a unit e , set

$$b_\mu(e) = \operatorname{Cov}_\mu((I - ee^T)Pe, e^T Pe), \quad \|b_\mu(e)\| \leq 1/4.$$

The covariance couples transverse prediction action with the variance of the observed label.

Theorem 10 (Complete-regression quartic transfer) *Fix d, r . On every path $\lambda \downarrow 0$ with $m\lambda^2 \rightarrow \rho \in [0, \infty)$,*

$$\frac{N_m(\lambda)}{\lambda^4} \rightarrow C_{d,r}(\rho) := \frac{3}{2d} \mathbb{E}_{Y,e} \|b_{\mu_Y}(e)\|^2. \quad (4)$$

Here P is Haar, $Y = \sqrt{\rho}P + G$, and G is symmetric with independent diagonal $N(0, 2)$ and upper off-diagonal $N(0, 1)$ entries. The direction e is independent uniform on the sphere, and

$$\mu_Y(dP) \propto \exp\{\sqrt{\rho} \operatorname{tr}(PY)/2\} dH(P).$$

Moreover C is continuous on $[0, \infty)$, $C(0) = 0$, $C(\rho) > 0$ for every finite $\rho > 0$, and $C(\rho) \rightarrow 0$ as $\rho \rightarrow \infty$. One has

$$C(\rho) \leq \frac{3}{32d}, \quad C(\rho) \leq \frac{3(d+1)}{8\rho} \quad (\rho > 0).$$

The coefficient uses the full regression risk, not a substitution of a subspace-estimation error. The proof first removes the truly optimal label-linear component in the actual Gaussian mixture, then transfers the resulting squared residual through the joint reference experiment. Polynomial Gaussian envelopes control the radial terms and justify passage to expectations; see Appendix E.

Corollary 11 (An information effect, not a worsening risk) *For sufficiently small fixed λ , some finite increase in reference count increases N_m above N_0 , whereas $N_m \rightarrow 0$ at infinite reference count. Both optimized total risks are nonetheless nonincreasing in m .*

The zero-information endpoint is a statement about the normalized coefficient: $N_0(\lambda) > 0$ at every fixed positive spike, although its quartic weak-spike coefficient is zero. Thus reference data can increase the value of responding to the label while improving both competing total risks. The zero/positive/zero endpoints quantify this decision-class effect; they do not assert a unique maximum or a discontinuous phase transition. At fixed positive spike, $N_m \rightarrow 0$ as references identify the geometry.

5.2 A finite-reference posterior-moment predictor

Put $s = \|X\|^2$, $b_D = q^T X/s$, $a = dy^2/s$, and, for a candidate P ,

$$z_P = X^T P X/s, \quad k_P = q^T P X/s - b_D z_P, \quad S_P = \frac{1}{2} \{\text{tr}[P(XX^T + qq^T)] + (1-a)z_P\}.$$

Theorem 12 (Complete-context moment correction) *Let μ_Z be the actual reference posterior, $M_Z = \mathbb{E}_{\mu_Z} P$, and*

$$f_{\text{corr}} = g_{M_Z} - \alpha^2 y \text{Cov}_{\mu_Z}(k_P, z_P + S_P).$$

For every fixed d , finite $\Lambda > 0$, all $m \geq 0$ and $0 < \lambda \leq \Lambda$, its actual Bayes excess obeys

$$R(f_{\text{corr}}) - R(\bar{g}) \leq C_{d,\Lambda} \alpha^6, \quad C_{d,\Lambda} = \frac{196(1+\Lambda)^{15}}{d} \sum_{j=0}^8 \binom{8}{j} 2^j (3d/2)_j.$$

Here $(x)_j$ is the rising factorial. The soft metric alone has $R(g_{M_Z}) - R(\bar{g}) = O_{d,\Lambda}(\alpha^4)$ uniformly in m .

A direct implementation is: compute the reference first and second projection moments; form the displayed linear functionals z, k, S of candidate projections at the observed data; contract their covariance; and subtract the second-order term from the soft prediction. No true subspace, coefficient or query answer is used. The covariance response is classical (Giordano et al. 2018); Appendix K gives the complete fixed-data likelihood expansion and polynomial square-risk remainder. The displayed constant makes the absolute, fixed-dimensional regret guarantee explicit for every reference count. The information cost V_m and this algorithmic approximation error remain separate. In two dimensions, the reference moments use scaled Bessel functions.

For every fixed $0 < \rho < \infty$, Theorems 10 and 12 imply

$$\frac{R(f_{\text{corr}}) - R(\bar{g})}{N_m} \longrightarrow 0 \quad (\lambda \downarrow 0, m\lambda^2 \rightarrow \rho).$$

Hence this observable construction asymptotically surpasses the best label-linear decision, not merely a particular soft baseline. Its form is

$$f_{\text{corr}} = \ell(W)y + \frac{\alpha^2 d}{2\|X\|^2} \text{Cov}_{\mu_Z}(k, z)y^3.$$

It is a general observable predictor, not automatically a same-spectrum metric representative for every label amplitude. At $m = 0$ the Haar covariance $\text{Cov}(k, z)$ is zero, so the zero-reference improvement also reflects input and query feedback rather than only label nonlinearity.

The correction is a constructive way to exploit the label’s geometry feedback using reference-posterior moments. Its guarantee concerns population Bayes regret at fixed dimension, uniformly over reference counts. At weak contrast it cancels the leading discrepancy from the full posterior action. A reference-only hard-versus-soft comparison and selective-feedback variants are recorded in the appendices; they do not substitute for the optimized timing comparator.

6 Numerical verification

The calculations distinguish a proved asymptotic coefficient, fixed-visible-data conditional checks, and saved random-task algorithmic losses. Only the last is a Monte Carlo prediction experiment. The complete regression posterior is an evaluation benchmark, never an input to the compared observable rules.

6.1 The timing coefficient in a tractable subspace experiment

For $d = 2, r = 1$, the matrix-channel normalization gives circle signal-to-noise ratio $\rho/4$. With independent standard $G_2 \in \mathbb{R}^2$, set

$$R = \|(\rho/4)e_1 + (\sqrt{\rho}/2)G_2\|, \quad \Delta(R) = \frac{I_2(R)}{I_0(R)} - \left(\frac{I_1(R)}{I_0(R)}\right)^2.$$

Direct posterior moments and uniform input-angle averaging yield

$$C_{2,1}(\rho) = \frac{3}{512}\mathbb{E}\Delta(R)^2, \quad C_{2,1}(\rho) \sim \frac{3\rho^2}{65536} \quad (\rho \downarrow 0).$$

Indeed $\text{Var}(e_\phi^T P e_\phi) = \{1 - a_1^2 + (a_2 - a_1^2)\cos 4\phi\}/8$, whose spherical derivative is $4b$. Its squared derivative average gives $\mathbb{E}_e \|b\|^2 = \Delta^2/128$. For the endpoint, $\Delta = -R^2/8 + O(R^4)$ and $\mathbb{E}R^4 = \rho^2/2 + O(\rho^3)$; the global bound $|\Delta| \leq CR^2$ supplies Gaussian domination.

An independent implementation integrates posterior projector moments and input angles without calling the Bessel moment formula, at concentrations 0, 0.1, 0.5, 2, 8, 32. The routes agree within 2×10^{-12} . Separate original-model conditional calculations at fixed visible X, q, Z use angular/label orders (64, 32), (128, 64), (256, 128) to check the optimized linear coefficient, positive conditional gap and same-spectrum action. These fixtures are conditional diagnostics, not typical unconditional risk samples.

6.2 An observable correction under equal information

We reuse all 6144 tasks in the frozen experiment: $\lambda \in \{.0625, .125, .25\}$, $\rho \in \{0, .25, 1, 4\}$, $m = \rho/\lambda^2$, and 512 independently seeded tasks per cell. Every method

ρ	$C_{2,1}(\rho)$	$ C_{128} - C_{64} $
0.25	$2.39842e - 06$	$2.54e - 21$
1	$2.47869e - 05$	$1.36e - 20$
4	$1.22466e - 04$	$7.66e - 16$
16	$1.51378e - 04$	$2.77e - 10$
64	$2.26028e - 05$	$2.36e - 10$

Table 1 Deterministic quadrature of the proved quartic timing coefficient. Gaussian quadrature orders were fixed at 32, 64, 128. The final column is a refinement diagnostic, not a rigorous interval error bound or Monte Carlo standard error.

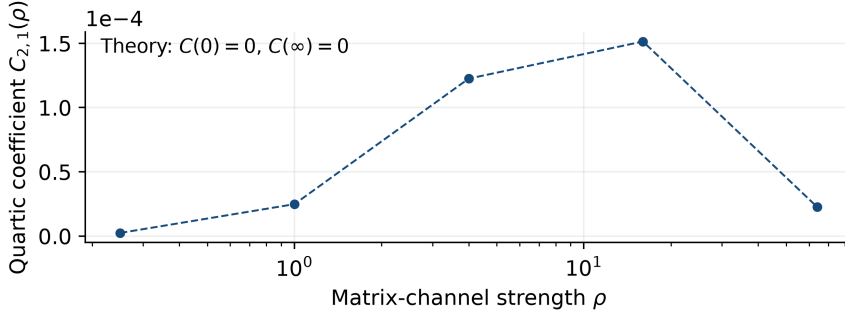


Fig. 1 Evaluated asymptotic coefficient $C_{2,1}(\rho)$, not a finite-spike N_m or V_m curve. Lines connect the fixed evaluation points as a visual guide. The zero endpoints and finite-positive strictness are proved; no unique peak is claimed.

receives the same X, y, q, S_Z, m, λ . Methods are isotropic prediction, reference PCA, the soft reference mean and the moment correction. All twelve cells, including the strong isotropic rule at $m = 0$, are retained. Tables and figures use the complete frozen arrays and taskwise paired standard errors; no grid cell was selected or tuned for the presentation.

Conditional-mean orthogonality gives

$$\mathbb{E}[(f - Y_*)^2 - (\bar{g} - Y_*)^2 \mid O] = (f - \bar{g})^2.$$

Thus the conditionalized loss estimates the same population algorithmic regret A_f , with target noise integrated out. Table 2 reports this quantity and the paired conditionalized difference, each with its own Monte Carlo standard error. Actual target losses and their differently noisy paired differences are retained separately.

The correction has smaller empirical regret than the soft rule in every frozen cell. The relative separation is most pronounced at smaller spectral contrast, consistent with cancelling the soft predictor’s second-order discrepancy. Its benefit is closeness to the full predictive decision, rather than better recovery of the latent projector. Across

λ	ρ	m	A_{soft} (SE)	A_{corr} (SE)	$\Delta_{\text{corr-soft}}$ (SE)
0.0625	0	0	$3.71e-07$ ($1.7e-07$)	$2.28e-10$ ($6.9e-11$)	$-3.71e-07$ ($1.7e-07$)
0.0625	0.25	64	$2.81e-07$ ($9.1e-08$)	$2.95e-10$ ($1.1e-10$)	$-2.81e-07$ ($9.0e-08$)
0.0625	1	256	$3.91e-07$ ($1.8e-07$)	$5.44e-10$ ($3.2e-10$)	$-3.91e-07$ ($1.8e-07$)
0.0625	4	1024	$1.25e-07$ ($4.4e-08$)	$2.18e-10$ ($1.2e-10$)	$-1.25e-07$ ($4.4e-08$)
0.125	0	0	$9.15e-06$ ($3.6e-06$)	$2.07e-08$ ($7.0e-09$)	$-9.13e-06$ ($3.6e-06$)
0.125	0.25	16	$3.91e-06$ ($1.6e-06$)	$1.56e-08$ ($7.7e-09$)	$-3.90e-06$ ($1.6e-06$)
0.125	1	64	$4.04e-06$ ($1.2e-06$)	$1.35e-08$ ($5.0e-09$)	$-4.02e-06$ ($1.2e-06$)
0.125	4	256	$1.71e-06$ ($3.9e-07$)	$8.47e-09$ ($2.5e-09$)	$-1.70e-06$ ($3.9e-07$)
0.25	0	0	$7.43e-05$ ($4.0e-05$)	$3.14e-07$ ($6.5e-08$)	$-7.40e-05$ ($3.9e-05$)
0.25	0.25	4	$7.73e-05$ ($2.0e-05$)	$5.82e-07$ ($1.7e-07$)	$-7.67e-05$ ($2.0e-05$)
0.25	1	16	$3.91e-05$ ($7.2e-06$)	$3.72e-07$ ($7.9e-08$)	$-3.87e-05$ ($7.1e-06$)
0.25	4	64	$2.82e-05$ ($6.1e-06$)	$5.03e-07$ ($1.3e-07$)	$-2.77e-05$ ($6.0e-06$)

Table 2 Algorithmic regret estimates and paired corrected-minus-soft differences, with Monte Carlo SE in parentheses. The displayed difference is the mean of taskwise conditionalized loss differences, and its SE is computed from those differences. These are not multiplicative improvements in total prediction risk.

all twelve cells, the ratio of the corrected to soft empirical Bayes approximation regret ranges from 0.00061 to 0.0179. This descriptive ratio is not a confidence interval: the paired taskwise differences and their standard errors quantify the stochastic comparison. The mechanism is the removal of the leading discrepancy from the complete predictive action, rather than a change to the irreducible geometry information cost. The complete-method means are in Appendix L; Figure 2 displays every method and cell with the SE of the plotted quantity.

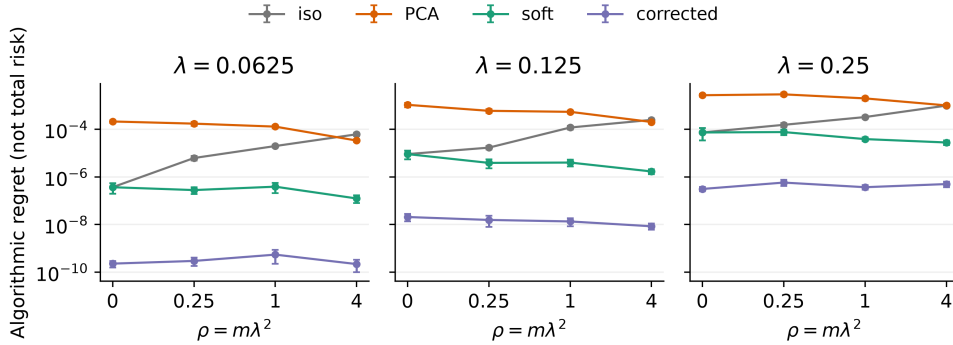


Fig. 2 Conditionalized algorithmic regret on the entire frozen grid. Bars show Monte Carlo SE, not posterior integration error. Isotropic and soft predictions coincide at zero reference count.

Precision and reproducibility.

Saved full-posterior means and variances use 256- and 512-direction periodic grids. Every task satisfies the frozen $2 \times 10^{-9}(1 + |\text{quantity}|)$ refinement check, with maximum

mean change 1.38×10^{-14} . Refinement stability is distinct from a certified quadrature error. Independent checks verify the mixture normal equation, cancellation, cubic projection, action representation, channel normalization and raw-loss arithmetic; deliberately incorrect linear-ratio and channel-strength implementations serve as negative controls. Execution dates, collection lists and return codes are reported separately in the audit materials. Visible sufficient statistics, evaluator-only truth, actual predictions and taskwise losses are supplied. The numerical risk illustration is two-dimensional; higher-dimensional results in this paper are analytic.

7 Connection to in-context learning

In-context regression infers a task from a labelled context and predicts a query, possibly using unlabelled inputs from the test environment. For any learned or explicit predictor with the stated observation boundary,

$$R_{\text{model}} = R_{\text{oracle}} + V_m + A_m, \quad A_m = \mathbb{E}(f_{\text{model}}(O) - \bar{g}(O))^2.$$

These terms distinguish insufficient test information from an ineffective use of that information. References principally address V_m ; changing the decision rule addresses A_m . Within the model, a rule linear in its sole observed label necessarily satisfies $A_m \geq N_m$, even if its coefficient uses all reference, context-input and query information.

Decision	Information selecting its action	Predictive interpretation
Reference geometry plugin	Z only	Recovers geometry before task prediction; softening retains reference uncertainty
Optimized label-linear rule	$W = (X, q, Z)$	Best coefficient chosen without label amplitude; regret N_m remains
Label-adaptive action	$O = (X, y, q, Z)$	Can represent the complete Bayes scalar prediction
Moment correction	Reference moments and O	Observable approximation with sixth-order regret at fixed dimension

Table 3 Observation permissions and the corresponding predictive decisions. The information cost V_m remains even for the complete Bayes action.

The decision-class result suggests a specific adaptation target. The label amplitude should be allowed to change the geometric action when it helps identify the latent task geometry. A fixed-rank representation can still be used: Theorem 6 shows that the timing of its selection, rather than its rank alone, determines attainability. The moment correction implements the leading feedback through a cubic label term

while using only observable posterior moments. These conclusions provide a statistical benchmark for the information and actions available to a context-processing predictor.

The coordinated coefficient prior describes a common latent task under changing input coordinates. The geometry-insensitive prior in Remark 3 delineates why task–input alignment is essential to the information price. The benchmark separates the Bayes information cost, the decision-class cost and a particular learner’s residual error, providing distinct targets for evaluating learned predictors.

8 Discussion and limitations

From geometry reconstruction to predictive action.

The response requires a normalized precision action along the observed context direction. This smaller decision object explains why a same-spectrum rank-constrained metric can reproduce the Bayes scalar prediction without identifying the latent subspace. The optimized comparator then isolates the consequence of choosing that action without label feedback. This distinction is useful for evaluating two-stage adaptation procedures: a geometry estimate should be judged by the predictive action it induces, with its observation permissions stated explicitly.

A reference-information regime for feedback.

The finite-sample bound locates the main information bottleneck around $mh \asymp d$. At fixed dimension, the matrix channel resolves this transition, while the quartic timing coefficient identifies the additional value of label feedback within it. The two scales answer different questions: missing geometry contributes at quadratic weak contrast, and restricting label feedback contributes at fourth order. The moment correction targets the latter error and is particularly natural for weak contrast, where its cancellation has a sixth-order guarantee and a direct two-dimensional implementation. The empirical separation is an improvement in Bayes approximation regret; its effect on total prediction error must also account for the oracle and information terms.

Relation to existing decisions and information bounds.

Group information inequalities (Wahl 2023; Chaumette et al. 2017; Gill and Levit 1995), minimum excess risk (Xu and Raginsky 2022), subspace estimation (Cai et al. 2013) and predictive projection (Besson et al. 2011; Piironen et al. 2020) establish the surrounding framework. Complete-observation score contraction and explicit Gaussian/Beta moments yield the predictive normalization with controlled dimension and rank dependence. The action-ball construction uses elementary projection geometry consistent with classical turning-angle bounds (Rao 2007). Its exact use in the optimized decision-class comparison is the relevant consequence.

Gaussian-mixture regression (Cohn et al. 1996; Sung 2004; Flåm et al. 2012), estimator-linearity criteria (Akyol et al. 2012), covariance differentiation (Giordano et al. 2018) and action-based value of information (De Lara and Gossner 2020) explain the feedback mechanism. Intermediate-information nonlinear advantages also occur in additive Gaussian MMSE theory (Guo et al. 2011, Propositions 4, 15–16). The present results quantify that mechanism for the complete regression experiment after

removing its truly optimal label-linear component. The finite-reference strictness and the covariance-risk normalization distinguish the precise statement from a qualitative comparison of curve shapes.

Scope of the conclusions.

The finite-sample information bound is uniform in dimension and rank for bounded contrast. The sharp constants, channel limit and quartic expansion fix dimension and rank; their limits are not interchanged. The experiment has one noiseless label, a known binary spectrum and a coordinated coefficient prior. The prior comparison in Remark 3 shows why this task–input relation matters. The action representation concerns the observed scalar conditional mean, rather than a reusable geometry estimate or the full predictive distribution. These distinctions also identify natural extensions, including several labelled context directions and approximate computation of higher-dimensional reference moments.

Data availability

The anonymized supplementary reproducibility archive contains the complete frozen visible sufficient statistics, separately stored evaluation truth, predictions, taskwise losses, quadrature diagnostics and executable source code. No human-participant, proprietary or restricted data are used. Numerical posterior integration in this paper is two-dimensional; the audit archive separately preserves the unsuccessful historical higher-dimensional integration and its unchanged diagnostic criteria.

AI-assisted preparation.

AI-assisted tools were used for analytic drafting, code preparation and internal cross-checking. They are not credited as authors or external reviewers. Human authors retain responsibility for final mathematical verification, approval and submission.

Appendix A Complete-context curvature and the information lower bound

A.1 Fixed-observation differentiation and complete-context curvature

Use all ambient coordinate-plane generators $G_{ij} = e_j e_i^T - e_i e_j^T$, $i < j$, acting by $\Pi \mapsto e^{tG_{ij}} \Pi e^{-tG_{ij}}$. Write D_{ij} for this derivative, holding X, y, q, Z fixed. Haar measure is invariant under each flow. At $\Pi_0 = \text{diag}(I_r, 0)$ only cross-block generators are nonzero. For them, Π'_{ij} has entries 1 in positions $(i, j), (j, i)$ and $\Pi''_{ij} = 2(e_j e_j^T - e_i e_i^T)$. Consequently

$$\sum_{i < j} D_{ij}^2 \Pi = 2(rI - d\Pi), \quad A' = -\alpha \Pi', \quad A'' = -\alpha \Pi''.$$

These are ambient Casimir identities, not derivatives along a moving-data path. For $Q = X^T A X$, $N = q^T A X$, differentiating $g = yN/Q$ gives

$$\begin{aligned} D_a g &= y(N'_a/Q - NQ'_a/Q^2), \\ D_a^2 g &= y(N''_a/Q - NQ''_a/Q^2 - 2N'_a Q'_a/Q^2 + 2N(Q'_a)^2/Q^3). \end{aligned}$$

The full context $D = (X, y, q)$ score is

$$S_a^D = -\frac{1}{2}(X^T A'_a X + q^T A'_a q) - \frac{Q'_a}{2Q} + \frac{dy^2 Q'_a}{2Q^2}. \quad (\text{U1})$$

The last two terms must not be omitted.

Only after these derivatives are taken, write under the true geometry

$$X = LRe, \quad q = L\nu, \quad y = Ru, \quad L = \Sigma^{1/2},$$

where e is uniform on the unit sphere, $R^2 \sim \chi_d^2$, $\nu \sim N(0, I_d)$, and $u \sim N(0, 1/d)$ are independent. To see the last independence, take $w = L^{-1}\eta/\sqrt{d}$ with independent standard Gaussian η , and put $u = e^T \eta/\sqrt{d}$: its conditional law given e is constant. This is a representation for computing expectations of previously derived quantities. Set

$$T = \|e_A\|^2, \quad z = 1 - T, \quad v_A = e_A^T \nu_A, \quad v_B = e_B^T \nu_B, \quad v = v_A + v_B, \quad k = zv_A - Tv_B,$$

$$M = \|\nu_A\|^2 v_B + \|\nu_B\|^2 v_A - 2v v_A v_B, \quad c = 2\alpha(d + \lambda s).$$

Here $T \sim \text{Beta}(r/2, s/2)$. Substitution into the fixed-data formulas yields, for $i \in A, j \in B$,

$$D_{ij} g = -\sqrt{h} u(\nu_i e_j + \nu_j e_i - 2v e_i e_j), \quad S_{ij}^D = \sqrt{h} \{(R^2 + 1 - du^2) e_i e_j + \nu_i \nu_j\}. \quad (\text{U2})$$

For clarity, the label-only part of the score is $\sqrt{h}(1 - du^2)e_i e_j$. Contracting all cross-block terms in the quotient second derivative gives

$$\sum_a D_a^2 g = u\{c - 4h(1 - 2T)\}k.$$

For example $L(\sum A''_a)L$ has diagonal blocks $2\lambda s I_r$ and $-2\alpha r I_s$; the remaining two quotient terms sum to $-4hu(1 - 2T)k$. The score product contributes $-hu\{(R^2 + 1 - du^2)(1 - 2T)k + M\}$. Thus, with $C_D = \sum_a (D_a^2 g + (D_a g)S_a^D)$, we obtain the exact identity

$$C_D = u\{[c - h(R^2 + 5 - du^2)(1 - 2T)]k - hM\}. \quad (\text{U3})$$

No coordinate-dependent data generation is differentiated in obtaining (U3).

A.2 Exact contractions and a dimension-uniform H estimate

Conditional on e , the independent block Gaussian moments give

$$\begin{aligned}\mathbb{E}_\nu k^2 &= Tz, \\ \mathbb{E}_\nu kM &= Tz\ell_k, \quad \ell_k = s - r - 2 + 4T, \\ \mathbb{E}_\nu M^2 &= M_2(T) = r(r+2)z + s(s+2)T - 4(d+1)Tz.\end{aligned}\tag{U4}$$

Here is a second explicit route for the cubic contraction, rather than an assumed Wick constant. For $0 < T < 1$, decompose $\|\nu_A\|^2 = v_A^2/T + U_A$ and $\|\nu_B\|^2 = v_B^2/z + U_B$, where $U_A \sim \chi_{r-1}^2$, $U_B \sim \chi_{s-1}^2$ are independent of $v_A \sim N(0, T)$ and $v_B \sim N(0, z)$ (a zero-degree chi-square is zero). Then

$$M = v_B[(T^{-1} - 2)v_A^2 + U_A] + v_A[(z^{-1} - 2)v_B^2 + U_B].$$

Its cross-product expectation is zero by oddness. The first squared bracket has expectation $r(r+2) - 4(r+2)T + 12T^2$; the second is the corresponding s, z expression. Multiplication by z, T and addition proves M_2 . Direct fourth moments similarly give $\mathbb{E}kM = Tz(s - r - 2 + 4T)$. Cauchy-Schwarz gives $M_2 - Tz\ell_k^2 \geq 0$. For $A_R = R^2 + 5 - du^2$,

$$d\mathbb{E}u^2 = 1, \quad d\mathbb{E}(u^2 A_R) = d + 2, \quad d\mathbb{E}(u^2 A_R^2) = (d+2)^2 + 2d + 6.\tag{U5}$$

These follow from $\mathbb{E}R^2 = d$, $\mathbb{E}R^4 = d(d+2)$ and Gaussian moments $\mathbb{E}u^4 = 3/d^2$, $\mathbb{E}u^6 = 15/d^3$. Completing the square in (U3), using $(d+2)(1-2T) + \ell_k = 2(s-dT)$ and $c - 2h(s-dT) = 2\alpha d(1+\lambda T)$, now proves

Lemma 13 (Exact full-context moment) *For every d, r, λ in the model,*

$$H = \mathbb{E}C_D^2 = \frac{1}{d}\mathbb{E}_T \left[4\alpha^2 d^2 Tz(1+\lambda T)^2 + h^2(2d+6)Tz(1-2T)^2 + h^2\{M_2(T) - Tz\ell_k(T)^2\} \right].\tag{U6}$$

Also $J_* := \mathbb{E}\sum_a (D_{ag})^2 = 2ph/[d(d+2)]$.

Proof Formula (U6) follows from (U3)–(U5). For J_* , the Gaussian mean square of each factor in (U2) is $e_i^2 + e_j^2 - 4e_i^2 e_j^2$. Summing over cross-block pairs gives $\mathbb{E}_e[sT + r(1-T) - 4T(1-T)] = 2p/(d+2)$; multiplying by $h\mathbb{E}u^2 = h/d$ gives the claim. $\square$

The Beta moments required to evaluate (U6) are explicitly

$$\mathbb{E}Tz = \frac{p}{d(d+2)}, \quad \mathbb{E}T^2z = \frac{p(r+2)}{d(d+2)(d+4)}, \quad \mathbb{E}T^3z = \frac{p(r+2)(r+4)}{d(d+2)(d+4)(d+6)}.$$

Higher polynomial moments are $(r/2)_j/(d/2)_j$. In particular $\mathbb{E}M_2 = p[1 + 4/(d(d+2))]$. There is no inverse radial moment in (U2)–(U6): all quantities are polynomials

in u, v, R^2 with bounded sphere factors. Thus the displayed expectations and the integrations below are finite even at $d = 2$. Since $1 + \lambda T \leq b$, $|1 - 2T| \leq 1$, and $M_2 - Tz\ell_k^2 \leq M_2$, the three nonnegative terms in (U6), divided by $dJ_* = 2ph/(d+2)$, are respectively at most

$$2b, \quad \frac{h(2d+6)}{2d^2} \leq \frac{5}{4}h, \quad h \left(\frac{1}{2} + \frac{1}{d} + \frac{2}{d^2} \right) \leq \frac{3}{2}h.$$

Hence

$$H \leq [2b + 11h/4]dJ_* \leq K_\Lambda dJ_*. \quad (\text{U7})$$

For $d = 2, r = 1$, (U6) reduces exactly to $H = h + 37h^2/32$. Under $t = \log(1 + \lambda)/2$, $h = (e^t - e^{-t})^2$, recovering the earlier two-dimensional constant.

A.3 Haar integration by parts with all visible information

For one reference sample, (U1) without the label and with that row alone gives cross-block score $\sqrt{h}\xi_i\xi_j$ for standard Gaussian ξ . Its mean is zero and its covariance is hI_p . Independent rows therefore give conditional covariance mhI_p . Put $C_O = C_D + \sum_a (D_ag)S_a^Z$. Conditioning on (Π, D) proves

$$\mathbb{E}C_O^2 = H + mhJ_*. \quad (\text{U8})$$

For any square-integrable decision $f(O)$, Haar invariance applied to $D_a[(f - g)(D_ag)p_\Pi(O)]$ gives

$$\mathbb{E}[(f - g)C_O] = J_*. \quad (\text{U9})$$

This derivative holds observations fixed, so $D_af = 0$. To justify it, for $X \neq 0$ the functions of Π and the strictly positive likelihood are smooth on the compact orbit; $f(O)$ is fixed during that integral. Haar has no boundary. The absolute integrated products are finite by (U2)–(U8), Cauchy–Schwarz and $\mathbb{E}f^2 < \infty$; Fubini is therefore valid. The event $X = 0$ is null. Finite target squared risk implies $\mathbb{E}f^2 < \infty$ because $\mathbb{E}Y_*^2 = 1$. No differentiability of f in its visible arguments is needed. Taking $f = \bar{g}$ and using Cauchy–Schwarz proves

$$V_m \geq \frac{J_*^2}{H + mhJ_*} \geq \frac{2ph}{d(d+2)(K_\Lambda d + mh)}. \quad (\text{U10})$$

Identity (U9) also bounds $\mathbb{E}(f - g)^2$ for every finite-risk full-observation predictor. This is not an inference from a PCA lower bound with reduced information. With $x = mh/d$ and $a = h/d$, the ratio of (U10) to the theorem's benchmark is $2 \max\{1, (m+1)a\}/(K_\Lambda + x) \geq 2/(K_\Lambda + 1)$, since $\max\{1, (m+1)a\} \geq \max\{1, x\}$ and $K_\Lambda + x \leq (K_\Lambda + 1) \max\{1, x\}$. This includes $m = 0$.

A.4 An independent Gaussian Hermite contraction

The block-norm calculation of (U4) admits a different all-dimensional derivation. Set

$$\mu = rz + sT - 4Tz, \quad \ell = s - r - 2 + 4T, \quad v = v_A + v_B.$$

Then v and $k = zv_A - Tv_B$ are independent Gaussian linear forms of variances $1, Tz$. The degree-one Hermite component of M is $M_1 = (s - 2z)v_A + (r - 2T)v_B = \mu v + \ell k$. The remaining component is

$$M_3 = [\|\nu_A\|^2 - r - 2(v_A^2 - T)]v_B + [\|\nu_B\|^2 - s - 2(v_B^2 - z)]v_A.$$

It is orthogonal to every linear form. The two summands' cross expectation vanishes by blockwise odd parity. Their variances sum to

$$z(2r - 8T + 8T^2) + T(2s - 8z + 8z^2) = 2\mu.$$

Consequently $\mathbb{E}kM = Tz\ell$ and $\mathbb{E}M^2 = Tz\ell^2 + \mu^2 + 2\mu$. Expanding the latter gives (U4), and proves the stronger residual identity $M_2 - Tz\ell^2 = \mu^2 + 2\mu \geq 0$. This is an analytic identity for arbitrary positive block sizes, not an inference from a symbolic grid.

A.5 A posteriori-conditional rederivation and its classical status

This is a second derivation of the information step, not a second information inequality. For a fixed observation O with $X \neq 0$, let $\pi_O(P)$ be the positive smooth posterior density with respect to Haar measure. Put

$$j(O) = \mathbb{E}_{\pi_O} \sum_a (D_a g)^2, \quad C_O = \pi_O^{-1} \sum_a D_a (\pi_O D_a g), \quad k(O) = \mathbb{E}_{\pi_O} C_O^2.$$

The normalizer of π_O is constant in P . Thus this C_O is exactly the witness in (U8), including the label score. Haar integration by parts at this fixed O gives

$$\mathbb{E}_{\pi_O} C_O = 0, \quad \mathbb{E}_{\pi_O} (g C_O) = -j(O).$$

Cauchy–Schwarz therefore yields $\text{Var}_{\pi_O}(g) \geq j(O)^2/k(O)$. If $k = 0$, the second identity forces $j = 0$ and the ratio is defined as zero. Averaging and applying weighted Cauchy–Schwarz once more gives

$$V_m \geq \mathbb{E}[j(O)^2/k(O)] \geq \frac{(\mathbb{E}j(O))^2}{\mathbb{E}k(O)} = \frac{J_*^2}{H + mhJ_*}.$$

Smoothness on the compact orbit and the integrability established above justify both identities and the subsequent averaging. Equivalently one may first condition on D and use its posterior as the prior for the reference experiment; that prior already contains the complete label and query score. This observation locates the lower-bound mechanism within classical posterior-weighted divergence/covariance inequalities (Chaumette et al. 2017; Renaux et al. 2008; Gill and Levit 1995). Its availability for a fixed-data target is not, on its own, a methodological innovation.

Appendix B Independent prediction comparison and finite-sample upper bound

B.1 Global independent-projector prediction comparison

For a rank- r projector Q depending only on Z , put $B_M = LA_Q L = I - \alpha L(Q - \Pi)L$. Conditional on (Π, Q) , the whitened context is independent of Q , and the quotient representation gives

$$\mathbb{E}_D[(g_Q - g_\Pi)^2 \mid \Pi, Q] = \frac{1}{d} \mathbb{E}_e \frac{\|(I - ee^T)B_M e\|^2}{(e^T B_M e)^2}. \quad (\text{U11})$$

Indeed, $g_Q - g_\Pi = u\nu^T[B_M e/(e^T B_M e) - e]$; average over u, ν . The eigenvalues of B_M lie in $[1/b, b]$. For $B_0 = B_M - (\text{tr } B_M/d)I$,

$$\mathbb{E}_e \|(I - ee^T)B_0 e\|^2 = \frac{\|B_0\|_F^2}{d+2}.$$

This follows by subtracting the spherical fourth moment of $(e^T B_0 e)^2$ from the second moment of $\|B_0 e\|^2$, with $\text{tr } B_0 = 0$. Thus

$$\begin{aligned} \mathbb{E}_D[(g_Q - g_\Pi)^2 \mid \Pi, Q] &\leq \frac{b^2 \|B_0\|_F^2}{d(d+2)} \\ &\leq \frac{b^2 \alpha^2 b^2 \|Q - \Pi\|_F^2}{d(d+2)} = \frac{b^3 h \|Q - \Pi\|_F^2}{d(d+2)}. \end{aligned} \quad (\text{U12})$$

Centering minimizes Frobenius distance to scalar matrices, and $\|L\|_{\text{op}}^2 = b$. This proof fails without the conditional independence if Q is selected using D ; we do not do so. The isotropic decision $f_0 = yq^T X / \|X\|^2$ has exact error

$$\mathbb{E}(f_0 - g_\Pi)^2 = \frac{\lambda^2}{d} \mathbb{E}_T \frac{T(1-T)}{(1+\lambda T)^2} \leq \frac{bhp}{d^2(d+2)}. \quad (\text{U13})$$

It follows either by direct substitution $B_M = \Sigma$ into (U11), or by the two-block numerator $\|(I - ee^T)\Sigma e\|^2 = \lambda^2 T(1-T)$. The reference sample count can be zero.

B.2 A self-contained covariance and PCA upper bound

We provide elementary, nonoptimal constants rather than citing a sparse-PCA theorem outside its signal assumptions. Let $S_0 = m^{-1} \sum_{i=1}^m \xi_i \xi_i^T$ for independent standard Gaussian rows, $m \geq d$. For a unit vector v , $mv^T S_0 v \sim \chi_m^2$. The Gaussian-square moment generating function gives

$$\log \mathbb{E} e^{z(\chi_m^2 - m)} \leq \frac{mz^2}{1-2z} \quad (0 < z < 1/2), \quad \log \mathbb{E} e^{-z(\chi_m^2 - m)} \leq mz^2 \quad (z > 0).$$

For the first inequality, expand $-\log(1 - 2z) - 2z$ in its convergent series; for the second use $\log(1 + 2z) \geq 2z - 2z^2$. Chernoff with $z = \sqrt{t/m}/(1 + 2\sqrt{t/m})$ in the upper tail and $z = \sqrt{t/m}$ in the lower tail yields

$$\Pr\{|v^T(S_0 - I)v| \geq 2\sqrt{t/m} + 2t/m\} \leq 2e^{-t}.$$

A maximal $1/4$ -separated set on the unit sphere is a $1/4$ -net with at most 9^d points: disjoint radius- $1/8$ balls fit in the radius- $9/8$ ball. For a symmetric E , approximation of an extremal unit vector gives $\|E\|_{\text{op}} \leq 2 \max_{v \text{ in net}} |v^T E v|$. Union bounding, with $\ell = d \log 9$, proves

$$\Pr\left\{\|S_0 - I\|_{\text{op}} \geq 4\sqrt{(\ell + t)/m} + 4(\ell + t)/m\right\} \leq 2e^{-t}.$$

The square of the threshold is at most $g(t) = 32[(\ell + t)/m + (\ell + t)^2/m^2]$. Tail integration, including the interval $[0, g(0)]$, gives

$$\mathbb{E}\|S_0 - I\|_{\text{op}}^2 \leq g(0) + 2 \int_0^\infty e^{-t} g'(t) dt = 32 \left[\frac{\ell + 2}{m} + \frac{(\ell + 2)^2}{m^2} \right] \leq 640d/m, \quad (\text{U14})$$

using $\ell + 2 \leq 4d$ and $m \geq d$. For true sample covariance $S = LS_0L$, $\mathbb{E}\|S - \Sigma\|_{\text{op}}^2 \leq 640b^2d/m$. Let $\hat{Q}$ be any measurable top- r spectral projector of S (tie-breaking fixed). The variational maximization of $\text{tr}(SQ)$ gives, for $k_0 = \min(r, s)$,

$$\begin{aligned} \frac{\lambda}{2} \left\| \hat{Q} - \Pi \right\|_F^2 &= \text{tr}[\Sigma(\Pi - \hat{Q})] \\ &\leq \text{tr}[(S - \Sigma)(\hat{Q} - \Pi)] \leq \|S - \Sigma\|_{\text{op}} \sqrt{2k_0} \left\| \hat{Q} - \Pi \right\|_F. \end{aligned}$$

The rank of a projector difference is at most $2k_0$. This inequality is global, not contingent on a small-error event. Since $k_0d \leq 2p$, (U14) gives

$$\mathbb{E} \left\| \hat{Q} - \Pi \right\|_F^2 \leq \frac{10240b^2p}{m\lambda^2} = \frac{10240bp}{mh} \quad (m \geq d). \quad (\text{U15})$$

Equations (U12) and (U15) imply prediction error at most $10240b^4p/[d(d+2)m]$.

Completion of Theorem 5 Use the legal, known-parameter selector: PCA only if $m \geq d$ and $mh \geq d$, otherwise isotropic. Its inputs never include the true projector or target. Bayes orthogonality implies $V_m \leq \mathbb{E}(f - g_\Pi)^2$ for either branch. On the PCA branch the minimum in the benchmark is $1/(m+1)$ and $(m+1)/m \leq 2$, giving constant $20480b^4$. On the isotropic branch, divide (U13) by the benchmark. The ratio is at most $b \max\{1, (m+1)h/d\}$. If $mh < d$, then $(m+1)h/d < 1 + h/2$; if $m < d$, integer $m+1 \leq d$ gives $(m+1)h/d \leq h$. In either case the ratio is at most $b(1+h) \leq (1+\Lambda)^2$. The theorem's C_Λ dominates this and the PCA constant. Together with (U10) this proves the result for all m , including $m = 0, 1$ and $m < d$, edge ranks and balanced ranks. $\square$

The constants depend on a finite spike ceiling. No growing-dimension asymptotic interchange follows from this finite-sample theorem.

Appendix C Proof of the sharp fixed-spike limit

Proof Full-observation lower bound. Take all $k = d(d-1)/2$ generators $G_{ij} = e_j e_i^T - e_i e_j^T$, $i < j$, an orthonormal basis for the inner product $-\text{tr}(GH)/2$ on the rotation Lie algebra. Let $\mathcal{D}_a$ differentiate $\Pi \mapsto e^{sG_a} \Pi e^{-sG_a}$ at fixed visible data. Haar measure on projectors is invariant under each flow, so for smooth u , $\int \mathcal{D}_a u d\mu = 0$ (differentiate the invariant integral; the manifold is compact and has no boundary). Apply this to $(f(O) - g_\Pi(D)) \mathcal{D}_a g_\Pi(D) p_\Pi(O)$ and then sum. As in a weighted information identity, the derivative of $-g$ is essential. We obtain

$$\mathbb{E}[(f - g)C_O] = J_*, \quad C_O = \sum_a \{\mathcal{D}_a^2 g + (\mathcal{D}_a g) \mathcal{D}_a \log p_\Pi(O)\}, \quad J_* = \mathbb{E} \sum_a (\mathcal{D}_a g)^2. \quad (\text{IBP})$$

To justify integration over O , for fixed d, λ the first two matrix derivatives of A are uniformly bounded over all projectors and generators, and $X^T A X \geq \|X\|^2 / (1 + \lambda)$. Quotient differentiation bounds both derivatives of g by $C_{d,\lambda} B_0$. The context score, including the conditional label term, is bounded by $C_{d,\lambda}(1 + \|\xi\|^2 + \|\nu\|^2 + \|\eta\|^2)$; reference scores are Gaussian quadratic polynomials. All summands of C_O are square integrable, using the radial cancellation (E). Cauchy–Schwarz with the stipulated finite L^2 error $f - g$ gives absolute integrability, permitting Fubini in (IBP). Thus the identity holds for arbitrary finite-risk observable predictors, not merely a Z -selected subspace.

At Π_0 , only the $p = r(d-r)$ cross-plane generators are active. Formula (T) yields $J_* = 2p\delta_\lambda^2/[d(d+2)]$. This sum is rotation invariant: conjugation acts orthogonally on the full generator basis, so its squared-gradient norm is unchanged. Let C_D denote C_O without the reference scores and let $H = \mathbb{E} C_D^2 < \infty$. At Π_0 , the covariance of the reference-score vector for one row is δ_λ^2 on the active generators and zero on commuting generators. The prediction gradient likewise vanishes on the commuting generators. Conjugation rotates both vectors and this covariance together. Hence for every Π, D the conditional variance of the reference contribution to C_O is $m\delta_\lambda^2 \sum_a (\mathcal{D}_a g)^2$. Conditional independence and zero mean remove its cross term with C_D . Consequently

$$\mathbb{E} C_O^2 = H + m\delta_\lambda^2 J_*.$$

Choosing $f = \bar{g}$ in (IBP) and using Cauchy–Schwarz gives the genuine full-information bound

$$V_m \geq \frac{J_*^2}{H + m\delta_\lambda^2 J_*}, \quad \liminf m V_m \geq J_*/\delta_\lambda^2 = C(d, r). \quad (\text{LB})$$

The exact H is now evaluated in Lemma 13; finiteness alone suffices for this fixed-parameter limit.

Attainable upper bound and expectation convergence. Let $\hat{\Pi}$ be the leading rank- r empirical covariance projector, based only on Z , and use $g_{\hat{\Pi}}$ with the known λ . Fix true Π_0 by rotation, and write $E_m = S_Z/m - \Sigma$. For $\|E_m\|_{\text{op}} < \lambda/4$, a circle of radius $\lambda/2$ around $1 + \lambda$ separates the two spectral clusters. Its resolvent $R_0(z) = (zI - \Sigma)^{-1}$ has norm at most $2/\lambda$. Expanding $(zI - \Sigma - E_m)^{-1} = R_0 + R_0 E_m R_0 + \dots$ and integrating around the circle gives

$$\hat{\Pi} - \Pi_0 = \lambda^{-1} \{\Pi_0 E_m (I - \Pi_0) + (I - \Pi_0) E_m \Pi_0\} + O_{d,\lambda}(\|E_m\|_F^2). \quad (\text{Proj})$$

The first term follows by residues: two equal-eigenvalue blocks have a double pole with zero residue, and the cross block has residue $1/\lambda$. The geometric-series tail is bounded in Frobenius

norm by $8\sqrt{d} \|E_m\|_{\text{op}}^2 / \lambda^2$. This supplies the local perturbation expansion, not an appeal to BvM. The cross block of $\sqrt{m}E_m$ has limiting independent $N(0, 1 + \lambda)$ entries. Indeed its single-row entries are $Z_i Z_j$, of these variances and zero cross covariances; Gaussian third moments and the characteristic-function expansion used in Appendix D give the multivariate CLT. It is conditionally independent of D . Fixed-data Taylor expansion of the metric ratio in the entries of $\hat{\Pi} - \Pi_0$, together with (Proj), proves

$$\sqrt{m}(g_{\hat{\Pi}} - g_{\Pi_0}) \Rightarrow \sum_{i \leq r < j} \mathcal{D}_{ij} g(D) N_{ij} / \lambda, \quad N_{ij} \stackrel{\text{iid}}{\sim} N(0, 1 + \lambda),$$

with these Gaussian variables independent of D . The Taylor remainder is bounded on the good event by $C_{d,\lambda} B_0 \|E_m\|_F^2$. The good event has probability tending to one and $\sqrt{m} \|E_m\|_F^2 = O_P(m^{-1/2})$. We now control expectations globally. On the good event the projector expansion implies $\|\hat{\Pi} - \Pi_0\|_F \leq C_{d,\lambda} \|E_m\|_F$. On its complement, $\|\hat{\Pi} - \Pi_0\|_F \leq 2\sqrt{r}$ and $\|E_m\|_F \geq \lambda/4$ give the same inequality with a larger fixed constant. The metric ratio has a global Lipschitz bound $|g_{P_1} - g_{P_2}| \leq C_\lambda B_0 \|P_1 - P_2\|_F$: interpolate between projectors, along which $A \succeq I/(1 + \lambda)$, and differentiate the observed quotient. Thus $m(g_{\hat{\Pi}} - g)^2 \leq C_{d,\lambda} B_0^2 m \|E_m\|_F^2$ everywhere. For each centered entry of $ZZ^T - \Sigma$, independence gives

$$\mathbb{E} \left(m^{-1} \sum_{i=1}^m U_i \right)^4 = \{m \mathbb{E} U_1^4 + 3m(m-1)(\mathbb{E} U_1^2)^2\} / m^4 = O(m^{-2}).$$

There are finitely many entries, so $\mathbb{E} \|E_m\|_F^4 = O_{d,\lambda}(m^{-2})$. Conditional independence from B_0 and its finite fourth moment prove $\sup_m \mathbb{E} [\{m(g_{\hat{\Pi}} - g)^2\}^2] < \infty$. This explicit second-moment bound for the scaled loss supplies uniform integrability. The limiting mean equals $(1 + \lambda) \sum J_g / \lambda^2 = 2p/[d(d+2)]$. Bayes optimality gives $V_m \leq \mathbb{E}(g_{\hat{\Pi}} - g)^2$; sandwiching with (LB) proves the theorem. No data-driven fixed-parameter formula or restricted PCA lower bound has been substituted for the full Bayes risk. $\square$

Appendix D Complete-observation weak-spike risk transfer

Proof We supply all the risk-transfer steps. Let $b = 1 + \lambda_0$ for a fixed bound $0 \leq \lambda \leq \lambda_0$ and use the coupling in Section 2. Put $B_0(D) = \|y\| \|q\| / \|X\|$ and $W_0 = 1 + \|\xi\|^2 + \|\nu\|^2 + \|\eta\|^2$. Since L has eigenvalues between 1 and $\sqrt{b}$,

$$B_0(D_\lambda) \leq \sqrt{b/d} \|\eta\| \|\nu\|, \quad y^2 / \|X\|^2 \leq \|\eta\|^2 / d. \quad (\text{E})$$

Thus radial denominators are canceled by the label coupling; these envelopes have every finite moment.

1. Joint reference experiment. For $m \geq 1$ set $T_m = (S_Z - mI) / \sqrt{m}$, $c_m = \sqrt{m}\alpha$. Given Π , the centered row matrix $ZZ^T - \Sigma$ has covariance

$$\text{Cov}(Z_i Z_j, Z_k Z_l) = \Sigma_{ik} \Sigma_{jl} + \Sigma_{il} \Sigma_{jk}.$$

For a fixed symmetric test matrix C , its linear functional has variance $2\text{tr}(C\Sigma C\Sigma)$, tending uniformly over all projectors to $2\text{tr} C^2$. Its third absolute moment is uniformly bounded: $|Z^T C Z - \text{tr} C \Sigma| \leq \|C\|_{\text{op}} b(\|\zeta\|^2 + d)$ suffices. The scalar characteristic function at $u/\sqrt{m}$ therefore equals $1 - u^2 \text{Var}(Z^T C Z) / (2m) + O_{d,C,\lambda_0,u}(m^{-3/2})$, uniformly in Π . Raising to the m th power proves uniform conditional convergence of the centered sum; applying this to every linear functional identifies the Gaussian matrix covariance displayed above at $\Sigma = I$.

The mean is $\sqrt{m}\lambda\Pi \rightarrow \sqrt{\rho}\Pi$ uniformly. The null context D_0 depends only on independent (ξ, ν, η) , and is independent of Π and the reference innovations. Multiplying the conditional characteristic function by bounded functions of (Π, D_0) proves

$$(\Pi, D_\lambda, T_m) \Rightarrow (\Pi, D_0, \sqrt{\rho}\Pi + W), \quad (\text{J})$$

where W is independent of (Π, D_0) ; replacing D_0 by its almost-surely convergent D_λ is justified by continuity of the coupling.

2. Exact reference posterior and normalization. At fixed Z , the only geometry-dependent likelihood term is $\alpha \text{tr}(\Pi S_Z)/2$. Since $\text{tr} \Pi = r$, the reference posterior has density proportional to $\exp\{c_m \text{tr}(\Pi T_m)/2\}$ relative to Haar measure. Define the continuous map

$$B(c, T) = \frac{\int \Pi \exp\{c \text{tr}(\Pi T)/2\} d\mu(\Pi)}{\int \exp\{c \text{tr}(\Pi T)/2\} d\mu(\Pi)}.$$

Continuity follows from compactness and positive denominator, uniformly on each compact set of (c, T) . Exactly $M_Z = B(c_m, T_m)$, so (J) gives $M_Z \Rightarrow B(\sqrt{\rho}, Y)$. In Frobenius-orthonormal symmetric coordinates, W has independent $N(0, 2)$ coordinates. Thus its density is proportional to $\exp(-\text{tr} W^2/4)$ and the channel likelihood is $\exp\{\sqrt{\rho} \text{tr}(\Pi Y)/2 - \rho r/4\}$; its posterior mean is exactly $B(\sqrt{\rho}, Y)$. Trace can be removed: $Y_0 = Y - (\text{tr} Y/d)I = \sqrt{\rho}(\Pi - rI/d) + W_0^{\text{mat}}$. The trace Gaussian is independent of the traceless noise and has direction-independent mean $\sqrt{\rho}r$; dropping it leaves the same posterior. On the traceless symmetric space the Frobenius coordinates still have variance two. Centered diagonal entries have covariance $2(\delta_{ij} - 1/d)$ and zero sum; off-diagonal entries retain variance one and independence from them. They are not independent $N(0, 2)$ entries.

3. Full label/query posterior tilt. Ignoring only parameter-independent determinant terms, the auxiliary log likelihood relative to $\lambda = 0$ at fixed D is

$$h_\lambda(\Pi, D) = \frac{\alpha}{2} \text{tr}[\Pi(XX^T + qq^T)] - \frac{1}{2} \log\{Q_\Pi / \|X\|^2\} - \frac{dy^2}{2} \{Q_\Pi^{-1} - \|X\|^{-2}\}.$$

For all projectors, $\|X\|^2/b \leq Q_\Pi \leq \|X\|^2$, $\alpha \leq \lambda$, and $0 \leq Q_\Pi^{-1} - \|X\|^{-2} \leq \lambda/\|X\|^2$. Together with (E), these give $\sup_\Pi |h_\lambda| \leq b\lambda W_0$. For any reference posterior Q interpolate $dQ_a \propto e^{ah} dQ$, $0 \leq a \leq 1$. Differentiating its mean gives $\partial_a \mathbb{E}_{Q_a} \Pi = \mathbb{E}_{Q_a} [\Pi(h - \mathbb{E}_{Q_a} h)]$, with Frobenius norm at most $2\sqrt{r} \|h\|_\infty$. Consequently the complete-observation mean satisfies

$$\|M_O - M_Z\|_F \leq 2\sqrt{r} b\lambda W_0. \quad (\text{P})$$

This is global $O(\lambda)$ over the entire Grassmann manifold and every reference posterior. It requires polynomial moments only, not an exponential moment of W_0 .

4. Fixed-data expansion. Define the traceless symmetric matrix

$$T_D = y \left\{ \frac{qX^T + Xq^T}{2\|X\|^2} - \frac{(q^T X)XX^T}{\|X\|^4} \right\}, \quad h_H(D) = \text{tr}(T_D H).$$

At fixed data $\partial_\lambda A|_0 = -\Pi$, so $\partial_\lambda g|_0 = -h_\Pi(D)$. For $N = q^T AX$ and $Q = X^T AX$, $|N|, |N'| \leq \|q\| \|X\|$, $|N''| \leq 2\|q\| \|X\|$, $Q \geq \|X\|^2/b$, $|Q'| \leq \|X\|^2$, $|Q''| \leq 2\|X\|^2$. The four quotient-rule terms bound $|\partial_\lambda^2 g|$ by $(2b + 4b^2 + 2b^3)B_0$. Taylor's theorem hence gives the uniform bound

$$\sup_\Pi |g_{\lambda, \Pi}(D) - g_I(D) + \lambda h_\Pi(D)| \leq K\lambda^2 B_0(D), \quad K = b + 2b^2 + b^3. \quad (\text{Tay})$$

Also $\|T_D\|_F \leq B_0(D)$ (in the X direction its expression is a symmetric product with the orthogonal component of q). Under the coupling, $T_{D_\lambda} \rightarrow T_{D_0}$ almost surely and in every finite L^p , by continuity off $X = 0$ and (E).

The needed nuclear-norm control is explicit, not a Frobenius-to-nuclear inequality with hidden rank factors. Write $e_X = X/\|X\|$ and $q_\perp = (I - e_X e_X^T)q$. Then

$$T_D = \frac{y}{2\|X\|}(q_\perp e_X^T + e_X q_\perp^T), \quad \|T_D\|_* = \frac{|y| \|q_\perp\|}{\|X\|} \leq B_0.$$

Its only possible nonzero eigenvalues are $\pm |y| \|q_\perp\|/(2\|X\|)$ (the zero case is immediate). Thus $|\text{tr}(T_D H)| \leq B_0 \|H\|_{\text{op}}$, justifying in particular the dimension-free $2B_0$ bound used below.

5. From prediction to the channel random variable. Average (Tay) under the full posterior and subtract it from the true prediction. Then $(g_\Pi - \bar{g})/\lambda = -h_{\Pi-M_O}(D_\lambda) + \varepsilon$, $|\varepsilon| \leq 2K\lambda B_0$. By (P) it differs from $Q_\lambda = -h_{\Pi-M_Z}(D_\lambda)$ by at most $\lambda B_0(2K + 2\sqrt{r} bW_0)$. The L^2 norm of this bound is $O(\lambda)$, because its square is bounded by a fixed Gaussian polynomial times λ^2 . The joint convergence in (J), continuity of B and of T_D imply $Q_\lambda \Rightarrow -h_{\Pi-M_\rho(Y)}(D_0)$.

6. Actual expectation transfer. Both Π and every posterior mean have operator norm at most one. Therefore $|Q_\lambda| \leq 2B_0$ and

$$\sup_j \mathbb{E}|Q_\lambda|^4 \leq 16b^2(d+2)^2, \quad \mathbb{E}[Q_\lambda^2 \mathbf{1}_{|Q_\lambda|>M}] \leq 16b^2(d+2)^2/M^2. \quad (\text{UI})$$

Here $\mathbb{E}\|\eta\|^4 = \mathbb{E}\|\nu\|^4 = d(d+2)$ was used. Bounded continuous truncations and (UI) prove convergence of the square expectations. The L^2 replacement above changes second moments by at most $2\|Q_\lambda\|_2 \|E_\lambda\|_2 + \|E_\lambda\|_2^2 = O(\lambda)$, and hence does not change the limit. Finally D_0 is independent of (Π, Y) . The residual $H = \Pi - M_\rho(Y)$ has trace zero, since both traces equal r . Applying (M) conditional on (Π, Y) proves (L). A general symmetric residual would instead retain its trace correction; Frobenius MMSE was derived, not assumed.

7. Zero-information endpoint. For arbitrary $m \geq 1$ let $K_m = \alpha(S_Z - mI)$. Its conditional mean is $m\alpha\lambda\Pi$ and $\mathbb{E}\|K_m - \mathbb{E}[K_m | \Pi]\|_F^2 = \alpha^2 m\{(\text{tr } \Sigma)^2 + \text{tr } \Sigma^2\} \leq C_{d,\lambda_0} m\lambda^2$. If $m\lambda^2 \rightarrow 0$, both terms vanish, and continuity and boundedness of the compact posterior map give $M_Z \rightarrow rI/d$ in every finite L^p . For $m = 0$ this mean is exact. The error $|h_{M_Z - rI/d}(D_\lambda)| \leq B_0 \|M_Z - rI/d\|_F$ tends to zero in L^2 by Hölder, its L^4 matrix convergence and the L^4 envelope (E). The same Taylor, tilt and fourth-moment argument gives the endpoint $\|\Pi - rI/d\|_F^2/[d(d+2)] = r(d-r)/[d^2(d+2)]$. $\square$

For $d = 2, r = 1$, write $\Pi = (I + B(U))/2$ with U uniform on the unit circle. The two traceless coordinates of Y are $\sqrt{\rho}U/2 + G$, $G \sim N(0, I_2)$. Define $F_{\text{circle}}(a) = \frac{1}{4}\mathbb{E}\|U - \mathbb{E}[U | \sqrt{a}U + G]\|^2$. Thus $\mathcal{F}_{2,1}(\rho) = \frac{1}{16}\mathbb{E}\|U - \mathbb{E}[U | \sqrt{\rho/4}U + G]\|^2 = F_{\text{circle}}(\rho/4)/4$. Because $\lambda/t \rightarrow 2$, the original V/t^2 limit is exactly $F_{\text{circle}}(mt^2)$.

Posterior mean versus a subspace decision. For any rank- r projector decision $\hat{P}(Y)$,

$$\mathbb{E}\|\Pi - \hat{P}\|_F^2 | Y] = \mathbb{E}\|\Pi - M_\rho\|_F^2 | Y] + \|\hat{P} - M_\rho\|_F^2.$$

The optimal constrained decision selects the leading r eigenvectors of M_ρ , whereas the unconstrained squared-loss decision is M_ρ itself. For every finite Y and finite ρ , posterior density has full Haar support; every eigenvalue of M_ρ lies strictly between zero and one. Thus the two decisions cannot be interchanged. At $\rho = 0$, their mean losses are respectively $2r(d-r)/d$ and $r(d-r)/d$, an exact factor-of-two discrepancy.

Appendix E Optimal label timing and its complete-risk transfer

E.1 The optimized comparator

Proof Conditionally minimize the quadratic $\mathbb{E}[(\beta y - Y_*)^2 \mid W]$. Its normal equation is $\beta \mathbb{E}_\tau v = \mathbb{E}_\tau c$. The optimal intercept is zero, so allowing affine label rules does not improve this value. Conditional-mean orthogonality proves (2). The exact geometric attainment is proved in Appendix K; every $Q(W)$ rule is label-linear, and the constructed representative of $I - \alpha M_W$ is label-blind. Averaging the normalized precision actions under the full posterior gives the representative for $\bar{g}$. In general this average is not the normalized action of $\mathbb{E}(P \mid O)$. $\square$

Lemma 14 (Variance-level cancellation) *At a fixed W with $X \neq 0$, the conditional gap in (2) is zero if and only if*

$$\mathbb{E}_\tau[c_P \mid v_P] = \beta v_P \quad \tau\text{-almost surely.} \quad (\text{E1})$$

Proof The label density is strictly positive on $\mathbb{R}$. Equality to βy almost everywhere says that

$$\int (c_P/v_P - \beta) v_P^{-1/2} e^{-x/(2v_P)} \tau(dP) = 0$$

for $x = y^2$ almost everywhere. This is analytic in x since v_P lies in a compact positive interval, so it is zero for every $x \geq 0$. Push the finite signed measure inside this expression onto $t = 1/(2v)$. Differentiating at zero gives zero polynomial moments of every order. Polynomial approximation on its compact support implies that the signed measure is zero, which is equivalent to (E1). The converse is immediate. Thus distinct component slopes alone do not imply a gap: components at the same variance may cancel. $\square$

E.2 Finite-reference strictness

Proof Write $e = X/\sqrt{s}$, $z = e^T P e$, $b = q^T X/s$, and $k = q^T P X/s - bz$. Since $c_P/v_P = b - \alpha k/(1 - \alpha z)$, (E1) requires

$$\mathbb{E}_\tau[k \mid z] = \gamma(1 - \alpha z), \quad \gamma = (b - \beta)/\alpha.$$

But $|k| \leq (\|q_\perp\|/\sqrt{s})\sqrt{z(1-z)}$. The marginal of z under any positive Bingham tilt has full support $(0, 1)$, so approaching zero through the almost-everywhere equality forces $\gamma = 0$.

Consider $X = e_1$, $q = e_1 + e_2$, and $Z_i = 0$. Here $k = P_{12}$, and (1) is proportional to $\exp\{\alpha z + \alpha P_{12} + \alpha P_{22}/2\}$. At fixed z , the Haar-preserving reflection $\text{diag}(1, -I)$ negates P_{12} and preserves P_{22} . The conditional numerator for $\mathbb{E}_\tau[P_{12} \mid z]$ symmetrizes to

$$\mathbb{E}_H[P_{12} \sinh(\alpha P_{12}) e^{\alpha P_{22}/2} \mid z] > 0,$$

since $\mathbb{E}_H[P_{12}^2 \mid z] = z(1-z)/(d-1) > 0$. The latter follows from $\|(I - ee^T)Pe\|^2 = z(1-z)$ and conditional rotational symmetry. The zero criterion therefore fails at this configuration.

This center has probability zero; the argument uses a neighborhood, not the center as a sampled event. The conditional cost is continuous near it. Compact posterior integration is continuous, v is bounded away from zero locally, conditional means grow at most $C|y|$, and mixture densities have a common $C \exp(-y^2/(2v_{\max}))$ envelope. Dominated integration proves continuity of the cost. Every sufficiently small neighborhood then has positive cost and positive probability under the true Gaussian-input mixture, including when $m < d$. Integrating proves strictness. $\square$

E.3 Fourth-order risk transfer

Risk transfer Appendix F proves, uniformly in m ,

$$\|\bar{g} - g_{\text{corr}}\|_2 \leq C_{d,\Lambda} \alpha^3, \quad g_{\text{corr}} = \ell(W)y + \frac{\alpha^2 d}{2s} \text{Cov}_{\mu_Z}(k, z)y^3,$$

where ℓ is label-blind and μ_Z is the reference posterior. The class $\mathcal{L}$ is closed in actual L^2 : multiplication $\beta \mapsto y\beta(W)$ is an isometry from the complete weighted space $L^2(\mathbb{E}[y^2 | W]dP_W)$, since $\mathbb{E}[y^2 | W] = \mathbb{E}_\tau v > 0$ almost surely. Its projection of h is $y\mathbb{E}[yh | W]/\mathbb{E}_\tau v$. Let $P_{\mathcal{L}}$ be this orthogonal projection. Conditional Gaussian mixture moments give

$$(I - P_{\mathcal{L}})g_{\text{corr}} = \frac{\alpha^2 d}{2s} \text{Cov}_{\mu_Z}(k, z) \left\{ y^3 - 3 \frac{\mathbb{E}_\tau v^2}{\mathbb{E}_\tau v} y \right\}.$$

The projection uses the true mixture variance, not the zero-spike variance. Its squared norm divided by α^4 is

$$A_\alpha = \mathbb{E}_W \frac{s}{4d} \text{Cov}_{\mu_Z}(k, z)^2 B_\alpha, \quad (\text{E2})$$

$$B_\alpha = 15\mathbb{E}_\tau u^3 - 9 \frac{(\mathbb{E}_\tau u^2)^2}{\mathbb{E}_\tau u}, \quad u = 1 - \alpha z. \quad (\text{E3})$$

Cauchy-Schwarz gives $6(1 - \alpha)^3 \leq B_\alpha \leq 15$, and $B_\alpha \rightarrow 6$ uniformly in τ . Projection contraction implies

$$\left| \frac{\sqrt{N_m}}{\alpha^2} - \sqrt{A_\alpha} \right| \leq C_{d,\Lambda} \alpha. \quad (\text{E4})$$

Exactly, $s \text{Cov}_{\mu_Z}(k, z)^2 = (q^T b_{\mu_Z}(e))^2$.

Write $Z_i = \Sigma_P^{1/2} \zeta_i$ using standard independent Gaussians and set $U_m = \sum_i \zeta_i \zeta_i^T$. The centered reference posterior field is

$$H_\lambda = \frac{\alpha}{2}(S_Z - mI) = \frac{\alpha m \lambda}{2} P + \frac{\alpha \sqrt{m}}{2} \Sigma_P^{1/2} \frac{U_m - mI}{\sqrt{m}} \Sigma_P^{1/2}.$$

At $m = 0$ the centered term is defined to be zero. For any symmetric T , the summand $\zeta^T T \zeta - \text{tr } T$ has variance $2 \text{tr } T^2$ and finite third moment. Applying the scalar characteristic-function CLT to each linear combination establishes the joint matrix CLT, whose diagonal variances are two and off-diagonal variances one, with all upper-triangle cross covariances zero. Its Gaussian limit is independent of P and of the independent ξ, ν generating X, q . Thus H_λ converges jointly to $\rho P/2 + \sqrt{\rho} G/2$, while X, q converge jointly to independent standard ξ, ν . At $\rho = 0$, the mean and centered second moment of H_λ instead vanish directly; this also covers bounded m . The removed mI term is constant on fixed-rank projectors. The matrix-channel likelihood is proportional to $e^{-\|Y - \sqrt{\rho} P\|_F^2/4}$; its $\rho \text{tr}(P^2)/4$ term is constant, giving the stated posterior.

Compact-orbit integrals defining b are continuous in H, e , with positive normalizer. Outside $\xi = 0$, $e = X/\|X\| \rightarrow \xi/\|\xi\|$. Equation (E2) therefore converges in distribution to $3(\nu^T b_{\mu_Y}(e))^2/(2d)$. To transfer expectations, bound it by $15\|q\|^2/(64d)$. For bounded small spikes,

$$\mathbb{E} \|q\|^4 \leq (1 + \Lambda)^2 d(d + 2), \quad \mathbb{E} [\|q\|^2 \mathbf{1}_{\|q\| > L}] \leq \mathbb{E} \|q\|^4 / L^2.$$

Truncation and this explicit tail bound prove expectation convergence. Integrating the independent standard ν yields $\|b\|^2$. Finally (E4), boundedness of A_α , and $\alpha/\lambda \rightarrow 1$ prove (4). This argument transmits the full original risk after subtraction of its truly optimal label-linear component, not only the reference experiment. $\square$

Coefficient structure At $\rho = 0$, Haar reflection fixing e negates the transverse component and preserves $e^T P e$, so $b = 0$. Couple $Y_\rho = \sqrt{\rho}P + G$; continuity of compact posterior integrals and the bounded covariance prove continuity of C .

For strict positivity, set $e = e_1$ and consider the realizable posterior tilt $e^{aP_{11}+tP_{12}}$. At $t = 0$, its e_2 covariance component is zero by reflection. Its t derivative is

$$\text{Cov}_a(P_{12}^2, z) = \frac{1}{d-1} \text{Cov}_a(z(1-z), z), \quad z = P_{11}.$$

The function $F(a) = \mathbb{E}_a z(1-z)$ is positive at zero and tends to zero as $a \rightarrow \infty$, because exponential tilting of the Beta-supported z concentrates at one. Hence $F'(a)$ is nonzero at some finite a , and some small nonzero t makes the transverse covariance nonzero. Every $\rho > 0$ matrix channel has full support on symmetric matrices; continuity gives an open set of positive channel and input-direction probability. This proves $C > 0$ also for balanced ranks.

Let $M_Y = \mathbb{E}(P | Y)$. Vector Cauchy-Schwarz gives $\|b\|^2 \leq \mathbb{E}[\|P - M_Y\|_F^2 | Y]/4$. The unconstrained matrix estimator $Y/\sqrt{\rho}$ has mean square error $\mathbb{E}\|G\|_F^2/\rho = d(d+1)/\rho$, so Bayes optimality proves the $1/\rho$ bound and the zero endpoint at infinity. The universal bound uses $\|b\| \leq 1/4$. $\square$

For each fixed λ , $N_m \rightarrow 0$ as $m \rightarrow \infty$. To see this, let Q_m be reference PCA. The sample covariance converges to Σ_P almost surely; its eigenvalue gap is λ , so $Q_m \rightarrow P$. Whiten the inputs and put $B = \Sigma_P^{1/2}(I - \alpha Q_m)\Sigma_P^{1/2}$. Conditional on (P, Z) its oracle excess is exactly

$$\frac{1}{d} \mathbb{E}_e \left\| \frac{Be}{e^T Be} - e \right\|^2. \quad (\text{E5})$$

This follows by integrating the independent standard Gaussian coefficient and query after $X = \Sigma_P^{1/2} R e$. The eigenvalues of B are in $[(1+\lambda)^{-1}, 1+\lambda]$; therefore the integrand is bounded uniformly in m and converges to zero. Since $R(f_{\text{lin}}) \leq R(g_{Q_m})$ and $R(\bar{g}) \geq R_{\text{oracle}}$, (E5) bounds N_m and proves the claim.

E.4 Reference information and decision timing

Proof Fix $\rho_* > 0$ and take $m(\lambda) = \lfloor \rho_*/\lambda^2 \rfloor$. Theorem 10 yields a positive limit for $N_{m(\lambda)}/\lambda^4$, while the $m = 0$ path yields zero. The fixed- λ zero endpoint follows from (E5). In either decision class additional references can be ignored, proving monotonicity of each optimized total risk. Their difference need not be monotone. No unique peak or uniform growing- ρ regression expansion is asserted. $\square$

Appendix F Complete fixed-data correction and integrable remainder

This appendix proves the input used in the risk transfer, without relying on an unpublished earlier proof. Keep visible $D = (X, y, q)$ fixed throughout differentiation. Put

$$z = e^T P e, \quad k = q^T P X / s - (q^T X / s) z, \quad T_P = \text{tr}[P(XX^T + qq^T)], \quad t_y = dy^2/s,$$

$$S_P = \{T_P + (1 - t_y)z\}/2.$$

Hold $\mu = \mu_Z$ fixed and introduce an auxiliary parameter $a \in [0, \alpha]$. The remaining full-context log likelihood is

$$L_a(P) = \frac{a}{2} T_P - \frac{1}{2} \log(1 - az) - \frac{t_y}{2} \{(1 - az)^{-1} - 1\}.$$

Terms constant in P have been removed only from the posterior normalizer. In particular $L_0 = 0$ and $L'_0 = S_P$. At the true $a = \alpha$ its tilted posterior is exactly $P \mid O$. Also

$$g_{P,a} = yb - \frac{ayk}{1 - az}, \quad b = q^T X/s.$$

Let $B_0 = |y| \|q\| / \|X\|$ and $W_D = 1 + \|X\|^2 + \|q\|^2 + t_y$. For $\lambda \leq \Lambda$, $1 - az \geq (1 + \Lambda)^{-1}$, so direct differentiation gives

$$\begin{aligned} \|L'_a\|_\infty + \|L''_a\|_\infty &\leq C_\Lambda W_D, \\ \|L_a - aS\|_\infty &\leq C_\Lambda a^2 W_D, \\ \sup_P |g_{P,a} - yb + ayk + a^2 ykz| &\leq C_\Lambda a^3 B_0. \end{aligned}$$

All suprema are on the entire projector orbit. For example

$$L'_a = T_P/2 + \frac{z}{2(1 - az)} - \frac{t_y z}{2(1 - az)^2}, \quad L''_a = \frac{z^2}{2(1 - az)^2} - \frac{t_y z^2}{(1 - az)^3}.$$

Thus the complete label score is retained.

Let $\mu_a \propto e^{L_a} \mu$. For bounded f, h ,

$$\frac{d}{da} \mathbb{E}_{\mu_a} f = \text{Cov}_{\mu_a}(f, L'_a),$$

and differentiating a covariance bounds its change by a constant times the product of the three sup norms. Consequently

$$\mathbb{E}_{\mu_a} k = \mathbb{E}_\mu k + a \text{Cov}_\mu(k, S) + O_\Lambda(a^2 \|k\|_\infty (W_D + W_D^2)),$$

$$\mathbb{E}_{\mu_a} kz = \mathbb{E}_\mu kz + O_\Lambda(a \|k\|_\infty W_D).$$

These are Taylor integral remainders: differentiating the first covariance introduces L''_a and a third centered moment of k, L'_a, L''_a , bounded by $C \|k\|_\infty (W_D + W_D^2)$. No exponential likelihood-ratio moment is used.

With $M_Z = \mathbb{E}_{\mu_Z} P$, the soft metric predictor is $g_{M_Z} = yb - \alpha y \mathbb{E}_\mu k / (1 - \alpha \mathbb{E}_\mu z)$. Substituting the preceding expansions gives

$$\bar{g} - g_{M_Z} = -\alpha^2 y \text{Cov}_\mu(k, z + S) + O_\Lambda(\alpha^3 B_0 (1 + W_D + W_D^2)).$$

Define $g_{\text{corr}} = g_{M_Z} - \alpha^2 y \text{Cov}_\mu(k, z + S)$. Since $z + S = (3/2 - t_y/2)z + T_P/2$, it is exactly

$$g_{\text{corr}} = \ell(W)y + \frac{\alpha^2 d}{2s} \text{Cov}_{\mu_Z}(k, z)y^3.$$

Both the full posterior mean and its cubic correction may be nonlinear in y ; only $\ell(W)$ is label-blind.

To prove an actual L^2 bound, use the true generating coupling $X = L_P \xi$, $q = L_P \nu$, $y = \xi^T \eta / \sqrt{d}$. Put $W_0 = 1 + \|\xi\|^2 + \|\nu\|^2 + \|\eta\|^2$. Then

$$B_0 \leq \sqrt{(1 + \Lambda)/d} \|\eta\| \|\nu\|, \quad t_y \leq \|\eta\|^2, \quad W_D \leq C_\Lambda W_0.$$

The squared remainder is bounded by α^6 times a fixed Gaussian polynomial with finite expectation. This proves $\|\bar{g} - g_{\text{corr}}\|_2 \leq C_{d,\Lambda} \alpha^3$, uniformly over all reference counts and all reference posterior realizations. Radial denominators cancel before expectation; no inverse-radius assumption is hidden. Holding μ_Z fixed in the auxiliary derivative is a proof device, not a change to the true generating law or posterior at $a = \alpha$.

Appendix G Geometry-feedback and rank-balance refinements

G.1 The cost of cutting label geometry feedback

The leading risk limit alone does not distinguish complete posterior feedback from a rule that uses the same label to learn the regression task but omits its marginal geometry likelihood. Define

$$f_y = \int g_P(D) \mu(dP \mid Z, X, q), \quad f_{Xy} = \int g_P(D) \mu(dP \mid Z, q).$$

The latter removes both context-input and label geometry feedback; neither is an unlabelled predictor. They are legal cut rules, not a proposed new algorithm (Frazier and Nott 2025, 2026).

Theorem 15 (Critical geometry-feedback cost) *Fix d, r , let $\lambda \downarrow 0$ and $m\lambda^2 \rightarrow \rho \in (0, \infty)$, and use the channel posterior of Theorem 9. Put*

$$c_Y(e) = (I - ee^T) \text{Cov}_{P|Y}(Pe, e^T Pe), \quad \mathcal{K}_{d,r}(\rho) = \mathbb{E}_{Y,e} \|c_Y(e)\|^2,$$

where e is uniform on the unit sphere and independent of Y . Then

$$\begin{aligned} \frac{R(f_y) - R(\bar{g})}{\lambda^4} &\longrightarrow \frac{5}{2d} \mathcal{K}_{d,r}(\rho), \\ \frac{R(f_{Xy}) - R(\bar{g})}{\lambda^4} &\longrightarrow \frac{d^2 - 2d + 10}{4d} \mathcal{K}_{d,r}(\rho). \end{aligned}$$

The endpoint coefficient is zero at $\rho = 0$, including bounded-reference sequences. The continuous function $\mathcal{K}$ is positive for every finite $\rho > 0$ and tends to zero as $\rho \rightarrow \infty$.

Appendix H gives the full square-risk proof, with orbit-global likelihood interpolation and a polynomial L^2 remainder. Posterior linear response itself is classical (Giordano et al. 2018). Here $\mathcal{K} = \mathbb{E} \|\nabla_{\mathbb{S}} \text{Var}_{P|Y}(e^T P e)\|^2 / 16$ measures directional variation of remaining geometry uncertainty. Reference information can enable a nonzero label-feedback effect that cancels at the Haar endpoint. This is not the total value of obtaining a label, not a discontinuous phase transition, and not a uniform fourth-order theorem when dimension or channel strength grows.

G.2 Rank balance and the feedback onset

The coefficient in Theorem 15 distinguishes two statistical regimes even though both share the same leading geometry-risk limit. At unbalanced rank, $\mathcal{K}_{d,r}(\rho)$ starts at order ρ ; at exact balanced rank $r = d/2$, complement symmetry suppresses that term and it starts at order ρ^2 with a strictly positive coefficient. Theorem 16 gives both coefficients and polynomial remainder control, and Proposition 17 establishes the two-dimensional endpoint constants. This is not a new symmetry or cumulant principle (Guo et al. 2011), but an explicit directional prediction consequence. Because V_m is of order λ^2 , the feedback regret is relatively $O(\lambda^2)$ at fixed positive channel strength. The refinement does not overturn the leading information bottleneck.

Appendix H Geometry-posterior feedback and fourth-order prediction cost

This appendix keeps the model and the observation boundary unchanged. It studies two legal observable rules, not new coefficient priors or a new inference algorithm.

H.1 Rules, exact risk identity and fixed-data derivatives

Write $D = (X, y, q)$ and let μ_Z be the reference-only geometry posterior. Define

$$f_{Xy} = \int g_P(D) \mu(dP | Z, q), \quad f_y = \int g_P(D) \mu(dP | Z, X, q).$$

Both use the label inside g_P . Only its marginal geometry likelihood is removed in the second rule. For $S = Xy, y$ put $A_S = R(f_S) - R(\bar{g})$. Conditional-mean orthogonality gives

$$A_S = \mathbb{E}(f_S - \bar{g})^2.$$

This is classical Bayes excess risk; the cut rules are not conditional means under nested observation sigma fields.

Fix the visible data, put $e_X = X / \|X\|$, $t_P = e_X^T P e_X$, $Q_X = I - e_X e_X^T$, and $\alpha = \lambda / (1 + \lambda)$. Since $A_P = I - \alpha P$,

$$\begin{aligned} g_P &= g_0 - \lambda h_P + \epsilon_P, \quad g_0 = y q^T X / \|X\|^2, \\ h_P &= y q^T Q_X P X / \|X\|^2, \quad B_0 = |y| \|q\| / \|X\|, \\ \sup_P |\epsilon_P| &\leq C_b \lambda^2 B_0. \end{aligned} \tag{F1}$$

Indeed the exact correction is $-\alpha h_P/(1 - \alpha t_P)$, so its difference from $-\lambda h_P$ has the asserted bound when $0 \leq \lambda \leq \lambda_0$ and $b = 1 + \lambda_0$.

Ignoring terms constant in P , the omitted log likelihoods relative to $\lambda = 0$ are

$$\ell_y(P) = -\frac{1}{2} \log v_P - \frac{dy^2}{2\|X\|^2}(v_P^{-1} - 1), \quad v_P = 1 - \alpha t_P.$$

$$\ell_{Xy}(P) = \ell_y(P) + \frac{\alpha}{2} \|X\|^2 t_P.$$

Thus their fixed-observation derivatives are

$$s_y(P) = \frac{1}{2}(1 - dy^2/\|X\|^2)t_P, \quad s_{Xy}(P) = \frac{1}{2}(\|X\|^2 + 1 - dy^2/\|X\|^2)t_P. \quad (\text{F2})$$

These are derivatives of likelihoods at fixed observations, not derivatives along the data-generating trajectory.

H.2 A uniform-in-reference-count square-integrable remainder

Couple data as in Appendix D: $X = L_\lambda \xi$, $q = L_\lambda \nu$, $w = L_\lambda^{-1} \eta/\sqrt{d}$, where ξ, ν, η are independent standard Gaussians and $L_\lambda = (I + \lambda \Pi)^{1/2}$. Then $y = \xi^T \eta/\sqrt{d}$. Set $W_0 = 1 + \|\xi\|^2 + \|\nu\|^2 + \|\eta\|^2$. Uniformly on the projector orbit,

$$B_0 \leq \sqrt{b/d} \|\eta\| \|\nu\|, \quad dy^2/\|X\|^2 \leq \|\eta\|^2, \quad \|X\|^2 \leq b \|\xi\|^2.$$

In particular no inverse radial moment is needed. Since $v_P \geq 1/b$, direct differentiation of ℓ_y gives

$$\sup_P |\ell_S| \leq C_b \lambda W_0, \quad \sup_P |\ell_S - \lambda s_S| \leq C_b \lambda^2 W_0. \quad (\text{F3})$$

For detail, $|\alpha'| \leq 1$, $|\alpha''| \leq 2$ and $|v'| \leq 1$, $|v''| \leq 2$. The absolute second derivative of $-\log v/2$ is at most $b + b^2/2$; that of the remaining label term is at most $(dy^2/\|X\|^2)(b^2 + b^3)$. The X term adds at most $\|X\|^2$. Taylor's integral remainder proves (F3) over the entire orbit.

Let Q be either cut geometry posterior and $Q_a \propto e^{a\ell_S} Q$. For bounded functions u, v ,

$$\frac{d}{da} \mathbb{E}_{Q_a} u = \text{Cov}_{Q_a}(u, \ell_S), \quad \left| \frac{d}{da} \text{Cov}_{Q_a}(u, v) \right| \leq 6 \|u\|_\infty \|v\|_\infty \|\ell_S\|_\infty.$$

The latter follows by differentiating $\mathbb{E}uv - \mathbb{E}u \mathbb{E}v$ and bounding each of its three covariance terms by twice the product of sup norms. Therefore the exact interpolation identity is

$$f_S - \bar{g} = - \int_0^1 \text{Cov}_{Q_a}(g_P, \ell_S) da.$$

The constant g_0 drops out. Applying (F1)–(F3) inside the covariance and then the derivative bound yields

$$|f_S - \bar{g} - \lambda^2 \text{Cov}_Q(h_P, s_S)| \leq C_b \lambda^3 B_0 (W_0 + W_0^2). \quad (\text{F4})$$

For example, the covariance error from ϵ_P or $\ell_S - \lambda s_S$ is $O(\lambda^3 B_0 W_0)$; changing Q_a back to Q contributes $O(\lambda^3 B_0 W_0^2)$. This argument never bounds a likelihood ratio by e^{CW_0} .

The X and q marginal geometry tilts have orbit oscillation at most $C_b \lambda W_0$. Applying the same covariance derivative bound to h_P, s_S replaces Q in (F4) by μ_Z at an additional $C_b \lambda^3 B_0 W_0^2$ cost. The square expectation of (F4), divided by λ^4 , is $O_{d,b}(\lambda^2)$, because $B_0^2 (W_0 + W_0^2)^2$ is bounded by a fixed Gaussian polynomial with finite expectation. This is an explicit L^2 replacement, valid for any reference count.

H.3 Joint limit and exact coefficients

In the regime of Theorem 9, the reference sufficient statistic, true projector and coupled Gaussians converge jointly to (Y, Π, ξ, ν, η) , where $Y = \sqrt{\rho} \Pi + W$ and (ξ, ν, η) are independent of (Y, Π) . This is the joint reference CLT proved in Appendix D; its covariance is diagonal variance two and off-diagonal variance one. Compact-orbit integrals of bounded continuous functions of P converge under the posterior

$$\mu_\rho(dP \mid Y) \propto \exp\{\sqrt{\rho} \text{tr}(PY)/2\} d\mu(P).$$

Write $\xi = Re$, $u = e^T \eta / \sqrt{d}$, so $R^2 \sim \chi_d^2$, e is uniform on the sphere and $u \sim N(0, 1/d)$, all independent of each other and of ν, Y . Define

$$c_Y(e) = (I - ee^T) \text{Cov}_{\mu_\rho}(Pe, e^T Pe), \quad \mathcal{K}_{d,r}(\rho) = \mathbb{E}_{Y,e} \|c_Y(e)\|^2.$$

After the L^2 replacements above, the scaled prediction differences converge in distribution respectively to

$$\frac{u}{2} (R^2 + 1 - du^2) \nu^T c_Y(e), \quad \frac{u}{2} (1 - du^2) \nu^T c_Y(e). \quad (\text{F5})$$

The approximating variables in (F4) have uniformly bounded fourth moments: $|\text{Cov}_{\mu_Z}(h_P, s_S)| \leq C_b B_0 W_0$. Convergence of the squares' expectations follows by truncating at level M , applying bounded continuous convergence, and bounding the tail by the fourth moment divided by M^2 . The L^2 remainder then transfers the squared expectation by Cauchy–Schwarz. This supplies risk convergence, not merely formal linear response.

Since $\mathbb{E}_\nu(\nu^T c)^2 = \|c\|^2$, put $U = \sqrt{d}u$ and use $\mathbb{E}U^2 = 1, \mathbb{E}U^4 = 3, \mathbb{E}U^6 = 15, \mathbb{E}R^2 = d, \mathbb{E}R^4 = d(d+2)$. Then

$$\frac{\mathbb{E}[U^2(R^2 + 1 - U^2)^2]}{4d} = \frac{d(d+2) - 4d + 10}{4d}, \quad \frac{\mathbb{E}[U^2(1 - U^2)^2]}{4d} = \frac{10}{4d}.$$

Consequently

$$\frac{A_{Xy}}{\lambda^4} \longrightarrow \frac{d^2 - 2d + 10}{4d} \mathcal{K}_{d,r}(\rho), \quad \frac{A_y}{\lambda^4} \longrightarrow \frac{5}{2d} \mathcal{K}_{d,r}(\rho). \quad (\text{F6})$$

The first coefficient is not a pure label effect. The X -only response has coefficient $(d+2)/4$, the label-only response has $5/(2d)$ and their double cross term is -1 .

H.4 Cancellation, strict positivity and large-channel decay

For $F_Y(e) = \text{Var}_{\mu_\rho}(e^T P e)$, differentiation on the sphere gives $\nabla_{\mathbb{S}} F_Y(e) = 4c_Y(e)$. Hence

$$\mathcal{K}_{d,r}(\rho) = \frac{1}{16} \mathbb{E}_{Y,e} \|\nabla_{\mathbb{S}} F_Y(e)\|^2.$$

At $\rho = 0$ the posterior is Haar and $F_Y(e) = 2r(d-r)/[d^2(d+2)]$ is constant. Thus $\mathcal{K}(0) = 0$, and the zero-information endpoint of (F6) follows by the same compact posterior argument even when m remains bounded. This does not assert a uniform sixth-order rate on all paths with $m\lambda^2 \rightarrow 0$.

Here is a general-dimension proof that $\mathcal{K}(\rho) > 0$ for every finite $\rho > 0$. The channel mixture density is strictly positive everywhere. Consider realizable posterior tilts $\mu_\kappa(dP) \propto e^{\kappa P_{11}} d\mu(P)$. If $r \geq 2$, as $\kappa \rightarrow +\infty$ the posterior concentrates on $P_{11} = 1$: every neighborhood of this maximizer set has positive Haar measure and the exponential weight suppresses its complement. Compactness gives subsequential limits, and invariance under rotations fixing e_1 makes the limit the unique invariant distribution on projections containing e_1 . Thus it has a Haar rank- $(r-1)$ projection on $e_1^\perp$. Consequently

$$\text{Var}(P_{11}) \rightarrow 0, \quad \text{Var}(P_{22}) \rightarrow \frac{2(r-1)(d-r)}{(d-1)^2(d+1)} > 0.$$

If $r = 1, d \geq 3$, taking $\kappa \rightarrow -\infty$ instead gives a Haar rank-one projector inside $e_1^\perp$, and $\text{Var}(P_{22}) \rightarrow 2(d-2)/[(d-1)^2(d+1)] > 0$ while $\text{Var}(P_{11}) \rightarrow 0$. Some finite tilt therefore has a nonconstant spherical variance profile; its gradient is nonzero on an open sphere set. Continuity of the compact posterior integrals, and full support of Y , prove positive average gradient energy.

For the remaining case $d = 2, r = 1$, write $P = n_\theta n_\theta^T$ and take $Y = \text{diag}(y_1, y_2)$. The density is proportional to $e^{a \cos 2\theta}$, with $a = \sqrt{\rho}(y_1 - y_2)/4$. Putting $A = I_1(a)/I_0(a)$, $B = I_2(a)/I_0(a)$ gives

$$F_Y(e_\phi) = \frac{1}{8} [1 - A^2 + (B - A^2) \cos 4\phi], \quad c_Y(e_\phi) = \frac{1}{8} (A^2 - B) \sin 4\phi e_\phi^\perp.$$

For small nonzero a , $A^2 - B = a^2/8 + O(a^4)$, proving positivity.

Finally $\mathcal{K}$ is continuous on finite $\rho \geq 0$: couple $Y_\rho = \sqrt{\rho} \Pi + W$, use continuity of the posterior and boundedness of c . With $M = \mathbb{E}[P \mid Y]$, let $T = P - M$. The eigenvalues of T lie in $[-1, 1]$ and

$$\|c_Y(e)\|^2 \leq \mathbb{E}[\|T\|_F^2 \mid Y].$$

This follows from $c = (I - ee^T)\mathbb{E}[Te(e^T Te) \mid Y]$ and Jensen. In the channel, Ky-Fan's variational principle for the leading rank- r projector of Y gives $\sqrt{\rho} \left\| \hat{P} - P \right\|_F^2 / 2 \leq \text{tr}[W(\hat{P} - P)] \leq \|W\|_F \left\| \hat{P} - P \right\|_F$. Squaring the resulting inequality gives $\left\| \hat{P} - P \right\|_F^2 \leq 4 \|W\|_F^2 / \rho$. Bayes optimality then implies $\mathbb{E} \|P - M\|_F^2 \rightarrow 0$ and $\mathcal{K}(\rho) \rightarrow 0$. Thus $\mathcal{K}$ vanishes at both endpoints and is strictly positive between them. This last limit is a limit of the channel coefficient, not a uniform joint λ, ρ theorem.

H.5 Scope and classical coverage

Posterior tilt derivatives are classical linear-response identities (Giordano et al. 2018); likelihood feedback removal is classical modular Bayes (Frazier and Nott 2025, 2026). The present calculation identifies a model-specific critical coefficient and its information-dependent cancellation/nonzero behavior, while proving the full-regression fourth-order risk transfer. It neither proposes a new cut procedure nor establishes priority. It is distinct from acquiring a label: both rules already use y to update the regression task.

Appendix I Rank balance and endpoint behavior of feedback risk

Write $K_{d,r} = \mathcal{K}_{d,r}$. This appendix concerns the channel coefficient $K_{d,r}(\rho)$ in Appendix H, not a joint limit with growing dimension or channel strength. Write $p = r(d - r)$ and let $Q_e = I - ee^T$.

Theorem 16 (Rank-dependent onset) *At fixed d, r , as $\rho \downarrow 0$, if $r \neq d/2$ then*

$$K_{d,r}(\rho) = \frac{4p^2(d - 2r)^2}{d^4(d - 1)(d + 2)^2(d + 4)^2} \rho + O_{d,r}(\rho^2).$$

If d is even and $r = d/2$, then

$$K_{d,d/2}(\rho) = C_d \rho^2 + O_d(\rho^3), \quad C_d = \frac{d^2(d^3 + 10d^2 + 4d - 8)}{64(d - 1)^2(d + 1)(d + 2)^4(d + 6)^2} > 0.$$

In particular, $C_2 = 1/16384$. The exponent changes at exact rank balance, not merely the multiplicative constant.

Proof Put $S = P - (r/d)I$, $\delta = \sqrt{\rho}$, $\varepsilon = \delta/2$, and remove trace noise: $W_0 = W - (\text{tr } W/d)I$. The observed traceless matrix is $H_\rho = W_0 + \delta S_*$, where S_* is the true centered projection. The posterior is proportional to $\exp\{\varepsilon \text{tr}(SH)\}$ on the Haar orbit. For fixed H, e set $x = e^T Se$, $a = \text{tr}(SH)$ and $c_{\varepsilon,H}(e) = Q_e \text{Cov}_{\varepsilon,H}(Se, x)$. Differentiating only the direction argument e gives the exact identity

$$c_{\varepsilon,H}(e) = \frac{1}{4} \nabla_S \text{Var}_{\varepsilon,H}(x). \quad (\text{N1})$$

The posterior does not depend on this auxiliary independent e .

At the Haar prior, $e^\top Pe$ has beta parameters $r/2, (d-r)/2$. Its third central moment, obtained from $\mathbb{E}t^j = (r/2)_j/(d/2)_j$, is

$$\mu_3 = \frac{8p(d-2r)}{d^3(d+2)(d+4)}.$$

The stabilizer of e and $\text{tr } S = 0$ imply $\mathbb{E}(x^2 S) = d\mu_3(ee^\top - I/d)/(d-1)$: its value along e is μ_3 , and all perpendicular eigenvalues are equal and have sum $-\mu_3$. Consequently

$$\text{Var}_{\varepsilon, H}(x) = \text{Var}_0(x) + \varepsilon \frac{d\mu_3}{d-1} e^\top H e + O_{d,r}(\varepsilon^2 \|H\|_F^2).$$

$$c_{\varepsilon, H}(e) = \frac{\varepsilon d\mu_3}{2(d-1)} Q_e H e + O_{d,r}(\varepsilon^2 \|H\|_F^2).$$

Since $\mathbb{E}\|Q_e W e\|^2 = d-1$, this proves the displayed unbalanced leading coefficient, once the expectation argument below is applied.

For balanced rank, $S^2 = I/4$ and S has the same law as $-S$. All odd posterior-variance derivatives in ε vanish. Write

$$\sigma^2 = \mathbb{E}x^2 = \frac{1}{2(d+2)}, \quad \mu_4 = \mathbb{E}x^4 = \frac{3}{4(d+2)(d+6)}, \quad \gamma = \frac{d}{2(d-1)(d+2)}.$$

The covariance $\mathbb{E}(xa) = \gamma e^\top H e$ follows from the same stabilizer argument. The fourth cumulant

$$T_e(H) = \mathbb{E}(x^2 a^2) - \sigma^2 \mathbb{E}a^2 - 2(\mathbb{E}xa)^2$$

has, for traceless symmetric H and $d \geq 4$, the invariant form

$$T_e(H) = A_0 \text{tr } H^2 + A_1 e^\top H^2 e + A_2 (e^\top H e)^2. \quad (\text{N2})$$

To see that this is exhaustive, split H relative to e into its scalar, perpendicular vector and perpendicular symmetric block. Orthogonal invariance of the perpendicular block permits only its trace squared, trace of its square, the squared vector norm and scalar-trace product; $\text{tr } H = 0$ reduces these to the three terms in (N2). Here are three all-dimension moment identities determining the coefficients:

$$\begin{aligned} \frac{d-1}{d} A_0 + \frac{(d-1)^2}{d^2} (A_1 + A_2) &= \mu_4 - 3\sigma^4, \\ 2A_0 + A_1 &= \frac{4(\sigma^4 - \mu_4)}{d-1}, \\ A_0 + \frac{A_1}{d} + \frac{2A_2}{d(d+2)} &= -\frac{4\gamma^2}{d(d+2)}. \end{aligned} \quad (\text{N3})$$

The first uses $H = ee^\top - I/d$, hence $a = x$. For the second use $e = e_1$, $H_{12} = H_{21} = 1$ and all other entries zero. The equation $S^2 = I/4$ and stabilizer invariance give $\mathbb{E}(S_{12}^2 | x) = (1/4 - x^2)/(d-1)$; also $\mathbb{E}xa = 0$. For the third, average over e conditional on S : the spherical fourth-moment formula gives $\mathbb{E}_e x^2 = \sigma^2$ for every such S . Thus the first two cumulant terms cancel in this average and $\mathbb{E}_e (e^\top H e)^2 = 2 \text{tr } H^2 / [d(d+2)]$ gives the third equation. Solving (N3), an identity of rational functions of d , gives

$$\begin{aligned} A_0 &= \frac{d}{(d-1)(d+1)(d+2)^2(d+6)}, & A_1 &= -\frac{2d}{(d-1)(d+1)(d+2)(d+6)}, \\ A_2 &= -\frac{d(d^2 + 2d + 4)}{(d-1)^2(d+1)(d+2)^2(d+6)}. \end{aligned}$$

The balanced expansion is $\text{Var}_{\varepsilon, H}(x) = \sigma^2 + \varepsilon^2 T_e(H)/2 + O_d(\varepsilon^4 \|H\|_F^4)$, so (N1) yields

$$\frac{c_{\varepsilon, H_\rho}(e)}{\rho} \longrightarrow L_e(W_0) = \frac{A_1}{16} Q_e W_0^2 e + \frac{A_2}{8} (e^\top W_0 e) Q_e W_0 e \quad \text{in } L^2. \quad (\text{N4})$$

For clarity the Gaussian contraction is explicit. Fix $e = e_1$ and write $W_0 = \begin{pmatrix} h & b^\top \\ b & D \end{pmatrix}$, $n = d - 1$. Then $b \sim N(0, I_n)$ is independent of (h, D) , $\text{tr } D = -h$, $\mathbb{E} h^2 = 2n/d$, and $\mathbb{E} \text{tr } D^2 = n(d-2/d)$. Since $Q_e W_0^2 e = (D + hI)b$, this gives

$$\begin{aligned} \mathbb{E} \|Q_e W_0^2 e\|^2 &= n(d-2)(d+4)/d, \\ \mathbb{E} h \langle Q_e W_0^2 e, Q_e W_0 e \rangle &= 2n(d-2)/d, \\ \mathbb{E} h^2 \|Q_e W_0 e\|^2 &= 2n^2/d. \end{aligned}$$

Substitution into $\mathbb{E} \|L_e(W_0)\|^2$ gives exactly C_d above. Its positivity follows either from this squared Gaussian polynomial or from the displayed formula. For $d = 2$ the three invariant terms are dependent, so no uniqueness of (N3) is asserted. Directly, a traceless symmetric matrix has H^2 scalar and $L_e(H) = -(e^\top H e) Q_e H e / 128$. Its two independent GOE coordinates have variance one, giving $C_2 = 1/16384$.

We finish the remainder and expectation argument rather than relying on formal series. On the compact projection orbit, every derivative of a normalized tilted expectation is a finite joint cumulant of bounded observables and a . At fixed derivative order it is bounded by $C_{d,r,k} \|H\|_F^k$, uniformly over the tilt. This follows by repeatedly differentiating numerator and denominator, regrouping into cumulants under a probability distribution, and using $|a| \leq C_{d,r} \|H\|_F$; no reciprocal normalizer is left unbounded. The same bound holds after the auxiliary spherical gradient. Taylor's theorem therefore supplies the polynomial remainders used above globally in H . Since S_* is bounded and W_0 Gaussian, all their required moments are uniformly finite for $\delta \leq 1$. In the unbalanced case write $c = \delta L_1(W_0) + \delta^2 \{L_1(S_*) + L_2(W_0)\} + O_{L^2}(\delta^3)$, where L_j has degree j . The cubic cross term has zero mean: $L_1(W_0)$ is centered and independent of S_* , and Gaussian odd polynomial moments vanish. Thus $K = C_1 \delta^2 + O(\delta^4)$. In the balanced case even symmetry gives $c = \delta^2 L(W_0) + \delta^3 D L(W_0)[S_*] + O_{L^2}(\delta^4)$; its degree-three Gaussian cross term again vanishes. Hence $K = C_d \delta^4 + O(\delta^6)$. These are fixed-dimension remainders. $\square$

Proposition 17 (Two-dimensional endpoints) *For $d = 2, r = 1$,*

$$\lim_{\rho \downarrow 0} \frac{K_{2,1}(\rho)}{\rho^2} = \frac{1}{16384}, \quad \lim_{\rho \rightarrow \infty} \rho^2 K_{2,1}(\rho) = \frac{1}{8}.$$

Proof Rotate the true circle direction to $(1, 0)$, set $s = \sqrt{\rho}/2$, $G \sim N(0, I_2)$ and $R_s = \|G + s(1, 0)\|$. The angle posterior is von Mises with concentration $a = sR_s$. Direct angular averaging gives

$$K_{2,1}(\rho) = \mathbb{E} k(sR_s), \quad k(a) = \frac{1}{128} \left[\left(\frac{I_1(a)}{I_0(a)} \right)^2 - \frac{I_2(a)}{I_0(a)} \right]^2. \quad (\text{N5})$$

This also gives an independent one-dimensional integral against the Rice density $r \exp\{-(r^2 + s^2)/2\} I_0(sr)$. At zero, the absolutely convergent Bessel series give $\Delta(a) = (I_1/I_0)^2 - I_2/I_0 = a^2/8 + O(a^4)$. A uniform bounded-second-derivative tilt bound gives $|\Delta(a)| \leq C a^2$ for $a \leq 1$, while bounded cosine moments give this inequality also for $a \geq 1$. Thus $k(a) \leq C a^4$ globally.

For $s \leq 1$, $R_s^4 \leq C(1 + \|G\|^4)$ and $\mathbb{E}\|G\|^4 = 8$. Dominated convergence yields $K/s^4 \rightarrow 8/8192 = 1/1024$, proving the first endpoint.

For the other endpoint, write the von Mises density on $[-\pi, \pi]$ as proportional to $e^{a(\cos \theta - 1)}$. The inequality $\cos \theta - 1 \leq -2\theta^2/\pi^2$, the substitution $z = \sqrt{a}\theta$, and a lower normalization bound obtained on $|\theta| \leq 1/\sqrt{a}$ prove $a\mathbb{E}(1 - \cos \theta) \rightarrow 1/2$ and $\mathbb{E}(1 - \cos \theta)^2 = O(a^{-2})$. Consequently, if $A = \mathbb{E} \cos \theta$ then $\Delta = 1 - A^2 - 2\text{Var}(\cos \theta)$ satisfies $a\Delta(a) \rightarrow 1$ and $|\Delta(a)| \leq C/a$ for $a \geq 1$. On $R_s \geq s/2$, $s^4 k(sR_s)$ is bounded uniformly for large s and tends almost surely to $1/128$. On its complement, $\Pr(R_s < s/2) \leq \Pr(\|G\| > s/2) = e^{-s^2/8}$; bounded k makes its contribution at most $Cs^4 e^{-s^2/8}$. Therefore $s^4 K \rightarrow 1/128$, and $\rho^2 = 16s^4$ gives $1/8$. $\square$

Remark 18 (Statistical meaning and the order of limits) *The feedback regret is of order λ^4 , whereas geometry information cost is of order λ^2 at fixed positive ρ . In particular*

$$\frac{A_y}{V_m} = \lambda^2 \frac{5K_{d,r}(\rho)}{2dF_{d,r}(\rho)} + o(\lambda^2), \quad F_{d,r}(\rho) = \frac{\mathbb{E}\|P - \mathbb{E}[P | Y]\|_F^2}{d(d+2)} > 0.$$

The feedback coefficient is zero at zero information, positive at finite positive information, and tends to zero at infinity. Thus its reference-strength dependence is not monotone; this does not say that Bayes risk increases when more data are observed, or that a cut rule's absolute risk increases. Rank balance delays its onset by one power of channel strength. First take the fixed- ρ critical limit in λ , then an endpoint limit of the resulting coefficient. No interchange or uniform joint growth bound is asserted. The covariance response and third/fourth cumulant mechanism are classical; the complete regression-loss transfer and its quantitative model-specific implications are the scope of the present calculation.

Appendix J Reference-only decisions and posterior-moment approximation

J.1 Reference-only hardening versus adaptive prediction

Let $a_\rho = \mathbb{E}\|\Pi - M_\rho(Y)\|_F^2$, and let $Q_\rho(Y)$ be the top- r spectral projector of Y . Define $j_\rho = \mathbb{E}\|Q_\rho(Y) - M_\rho(Y)\|_F^2$.

Theorem 19 (Actual prediction cost of reference-only hardening) *Under the fixed-dimension weak-identification limit of Theorem 9, with $\hat{\Pi}_Z$ the reference PCA projector,*

$$\begin{aligned} \frac{R(g_{M_Z}) - R_{\text{oracle}}}{\lambda^2} &\rightarrow \frac{a_\rho}{d(d+2)}, \\ \frac{R(g_{\hat{\Pi}_Z}) - R_{\text{oracle}}}{\lambda^2} &\rightarrow \frac{a_\rho + j_\rho}{d(d+2)}. \end{aligned}$$

For every finite $\rho > 0$, $a_\rho, j_\rho > 0$ and $1 < (a_\rho + j_\rho)/a_\rho < 2$. At $\rho = 0$ the ratio of the leading oracle-excess coefficients is exactly two. It is not a ratio of total prediction risks. As $\rho \rightarrow \infty$ after the weak-spike limit, $a_\rho, j_\rho \rightarrow 0$; no uniform high-channel rate is asserted.

The parameter-space posterior Pythagorean identity and minimum-mean-square subspace decision are classical (Besson et al. 2011). The proof in Appendix K additionally averages the actual regression loss, uses its independent-reference metric

normalization, and transfers it through the joint experiment. The soft reference rule achieves the leading Bayes coefficient. The next construction incorporates complete-context feedback at finite spike strength. The hardening comparison restricts metric selection to references; context-adaptive representations are covered by Theorem 4.

Appendix K Prediction decisions, hardening costs, and a posterior-moment correction

K.1 A finite-spike image theorem for the one-label decision

Fix $X \neq 0$, put $e = X/\|X\|$ and $0 < \alpha < 1$. For a rank- r orthogonal projector P , let $z_P = e^T P e$ and

$$v_P = \frac{(I - ee^T)Pe}{1 - \alpha z_P}.$$

Since $\|(I - ee^T)Pe\|^2 = z_P(1 - z_P)$, its largest possible norm is

$$t_* = (2\sqrt{1 - \alpha})^{-1}, \quad \max_{0 \leq z \leq 1} \frac{z(1 - z)}{(1 - \alpha z)^2} = \frac{1}{4(1 - \alpha)}.$$

Differentiating the scalar quotient gives its maximum at $z = (2 - \alpha)^{-1}$. Conversely, for any $v \perp e$ with $0 < t = \|v\| \leq t_*$, set

$$z = \frac{2t^2}{1 + 2\alpha t^2 + \sqrt{1 - 4(1 - \alpha)t^2}}, \quad u = \sqrt{z}e + \sqrt{1 - z}v/t.$$

The stable root satisfies $z(1 - z) = t^2(1 - \alpha z)^2$ and $0 < z \leq (2 - \alpha)^{-1}$. Complete u by $r - 1$ orthonormal vectors from $\text{span}(e, v)^\perp$, possible because $r - 1 \leq d - 2$, and let Q project onto their span together with u . Then

$$\frac{(I - \alpha Q)e}{e^T(I - \alpha Q)e} = e - \alpha v.$$

For $v = 0$, any rank- r projector in $e^\perp$ works. A measurable completion can be chosen by lexicographic Gram–Schmidt on coordinate vectors, on the finite partition determined by the first nonzero pivots. Boundary equality in t_* does not invalidate the construction. Thus the image of the normalized precision action of all rank- r projectors is exactly a closed transverse Euclidean ball, not merely its boundary.

For any probability distribution μ on these projectors, $\mathbb{E}_\mu v_P$ lies in this same ball by convexity. Consequently its mean normalized precision action is achieved by a single rank- r Q depending on X and μ . In particular, taking μ to be the actual posterior given O gives

$$g_Q(D) = \mathbb{E}[g_P(D) \mid O] = \bar{g}(O).$$

This is a decision representation, not a recovery of true geometry or a shortcut for evaluating the posterior. Q may depend on q, y, Z through that posterior, so it says nothing about reference-only geometry estimators. It does not eliminate V_m .

For any $0 \preceq M \preceq I$, set $v = (I - ee^T)Me/(1 - \alpha e^T Me)$. The inequality $M^2 \preceq M$ gives $\|(I - ee^T)Me\|^2 \leq z(1 - z)$, so the same construction gives $g_Q = g_M$ pointwise. If $M = M_Z$, Q uses only (Z, X) , even though both predictions also use the label and query. This supplies a precise counterexample to any extension of a reference-only hard-rank prediction penalty to all data-adaptive hard metrics.

K.2 Actual prediction loss of reference-only hardening

For a reference-only matrix S with $0 \preceq S \preceq I$ and $\text{tr } S = r$, let $B = L(I - \alpha S)L$. Conditional on true P, S , direct averaging of the Gaussian task and query gives

$$R(g_S) - R(g_P) = \frac{1}{d} \mathbb{E}_{P, S, e} \left\| \frac{Be}{e^T Be} - e \right\|^2.$$

The whitened direction e is independent of P, S , which is essential. Uniformly over this compact matrix class, $B = I + \lambda(P - S) + O(\lambda^2)$; its denominator is bounded below by $1/(1 + \lambda)$, so squaring the expansion leaves an $O_{d, \lambda_0}(\lambda^3)$ remainder independent of the distribution of S . Sphere fourth moments yield

$$R(g_S) - R(g_P) = \frac{\lambda^2}{d(d+2)} \mathbb{E} \|P - S\|_F^2 + O_{d, \lambda_0}(\lambda^3).$$

Without the equal-trace condition the trace correction from (M) remains.

In the critical reference experiment $Y = \sqrt{\rho}P + W$, put $M = \mathbb{E}[P \mid Y]$ and $Q = \text{Top}_r(Y)$. Its posterior density is proportional to $\exp\{\sqrt{\rho} \text{tr}(PY)/2\}$ relative to Haar. Sign flips in eigen-coordinates show that M commutes with Y . For eigenvalues $\eta_i > \eta_j$, average the Haar integral with the coordinate swap $i \leftrightarrow j$. The numerator of $M_{ii} - M_{jj}$ becomes

$$\int (P_{ii} - P_{jj}) e^{tC(P)} \sinh\{t(\eta_i - \eta_j)(P_{ii} - P_{jj})/2\} d\mu(P) > 0, \quad t = \sqrt{\rho}/2,$$

where $C(P)$ is the swap-invariant part of $\text{tr}(PY)$. Strictness follows from full Haar support and $1 \leq r < d$. Thus the top eigenspaces of Y and M agree for $\rho > 0$; at $\rho = 0$ every rank- r subspace is optimal. Ky–Fan optimization gives the classical minimum-mean-square-distance constrained decision, not a new estimator.

The joint experiment convergence from Appendix D applies to M_Z by continuity and to the reference PCA projector by continuity off eigenvalue ties. The limiting matrix has a density and hence ties have probability zero. Bounded Frobenius losses give convergence of their expectations. Define

$$a_\rho = r - \mathbb{E} \text{tr } M^2, \quad j_\rho = \mathbb{E} \|Q - M\|_F^2, \quad b_\rho = 2r - 2\mathbb{E} \text{tr}(MQ) = a_\rho + j_\rho.$$

The above actual prediction identity proves

$$\frac{R(g_{M_Z}) - R_{\text{oracle}}}{\lambda^2} \rightarrow \frac{a_\rho}{d(d+2)}, \quad \frac{R(g_{\hat{P}_Z}) - R_{\text{oracle}}}{\lambda^2} \rightarrow \frac{b_\rho}{d(d+2)}.$$

Combined with the complete Bayes risk transfer, the soft plugin has vanishing leading Bayes excess, while hardening costs $\lambda^2 j_\rho / [d(d+2)] + o(\lambda^2)$. A measurable limiting channel hard decision minimizes this leading coefficient by choosing the top- r eigenspace; no all-observation projection optimality is asserted.

At every finite positive ρ , each eigenvalue μ_i of M lies strictly between zero and one and is strictly ordered. Therefore $j_\rho > 0$, $a_\rho > 0$, and $1 < b_\rho/a_\rho < 2$. Indeed $\sum_i \mu_i^2 < \sum_{i \leq r} \mu_i$ by maximizing the linear functional with coefficients μ_i over $[0, 1]^d$ with sum r . At zero information, $M = rI/d$, $a_0 = j_0 = r(d-r)/d$, so the leading oracle-excess risk of a hard decision is exactly twice the soft/Bayes coefficient. This is not a ratio of total prediction risks. The endpoint also follows for arbitrary sequences $m\lambda^2 \rightarrow 0$, since $M_Z \rightarrow rI/d$ and the relevant hard loss is then constant regardless of the projector choice.

As $\rho \rightarrow \infty$ after the critical experiment limit, $Y/\sqrt{\rho} = P + W/\sqrt{\rho}$ and the spectral projector converges almost surely to P . Bounded convergence and Bayes optimality imply $a_\rho, b_\rho, j_\rho \rightarrow 0$. No uniform growing-dimensional expansion or high-channel rate is claimed here.

K.3 An implementable second-order posterior-moment correction

Fix the actual reference posterior μ_Z , and hold it fixed while differentiating the auxiliary likelihood at fixed visible data. Put $s = \|X\|^2$, $b_D = q^T X/s$, $a = dy^2/s$ and

$$z_P = X^T P X / s, \quad k_P = q^T P X / s - b_D z_P, \quad S_P = \{\text{tr}[P(XX^T + qq^T)] + (1-a)z_P\}/2.$$

The exact target and auxiliary log likelihood at a separate interpolation parameter t are

$$g_{t,P} = g_I - \frac{tyk_P}{1 - tz_P},$$

$$h_t(P) = \frac{t}{2} \text{tr}[P(XX^T + qq^T)] - \frac{1}{2} \log(1 - tz_P) - \frac{a}{2} \{(1 - tz_P)^{-1} - 1\}.$$

Let $F(t)$ be the normalized tilted mean of $g_{t,P}$ under $e^{h_t} d\mu_Z$, and $M_Z = \mathbb{E}_{\mu_Z} P$. At $t = \alpha$ this mean is exactly $\bar{g}$. Differentiating at zero, with the reference measure fixed, gives

$$F'(0) = -y \mathbb{E} k_P, \quad F''(0)/2 = -y \{\mathbb{E}(k_P z_P) + \text{Cov}(k_P, S_P)\}.$$

The soft metric predictor's second coefficient is $-y(\mathbb{E} k_P)(\mathbb{E} z_P)$. Define

$$f_{\text{corr}} = g_{M_Z} - \alpha^2 y \text{Cov}_{\mu_Z}(k_P, z_P + S_P).$$

This uses first and second projection moments only and includes the label-variance and query feedback; it does not use true geometry, true coefficients or the query answer.

We give a finite-sample remainder uniform over reference count. Let $b = 1 + \Lambda$ and $B_0 = |y| \|q\| / \|X\|$. For $0 \leq t \leq \alpha$, set

$$S_1 = (\|X\|^2 + \|q\|^2)/2 + b/2 + ab^2/2, \quad S_2 = b^2/2 + ab^3, \quad S_3 = b^3 + 3ab^4.$$

Direct differentiation gives $|h'| \leq S_1$, $|h''| \leq S_2$, $|h'''| \leq S_3$ and $|g^{(j)}| \leq j!b^{j+1}B_0$ for $j = 0, 1, 2, 3$. The $j = 0$ bound follows directly from the metric quotient. Normalized-mean differentiation gives

$$\begin{aligned} F''' = & \mathbb{E}g''' + 3 \text{Cov}(g'', h') + 3 \text{Cov}(g', h'') + 3 \text{cum}(g', h', h') \\ & + \text{Cov}(g, h''') + 3 \text{cum}(g, h', h'') + \text{cum}(g, h', h'). \end{aligned}$$

For bounded variables the corresponding products of suprema bound covariance, third cumulant and fourth cumulant by factors 2, 8, 64, respectively. The last factor follows by bounding a centered fourth product and its three covariance pairings; it is deliberately nonoptimal. Adding the soft target third-derivative bound gives

$$B_0[12b^4 + 12b^3S_1 + 6b^2S_2 + 24b^2S_1^2 + 2bS_3 + 24bS_1S_2 + 64bS_1^3].$$

Only now substitute the true Gaussian coupling. If $W_0 = 1 + \|\xi\|^2 + \|\nu\|^2 + \|\eta\|^2$, then

$$B_0 \leq \sqrt{b/d} \|\eta\| \|\nu\|, \quad a \leq \|\eta\|^2, \quad S_1 \leq b^2W_0, \quad S_2 \leq \frac{3}{2}b^3W_0, \quad S_3 \leq 4b^4W_0.$$

The displayed derivative sum is at most $165b^7B_0W_0^3$. Taylor's formula, and the exact second-order correction, therefore yield

$$|f_{\text{corr}} - \bar{g}| \leq 28b^7\alpha^3B_0W_0^3.$$

As $\|\eta\|^2 \|\nu\|^2 \leq W_0^2/4$, the actual Bayes prediction excess satisfies for every $m \geq 0$

$$R(f_{\text{corr}}) - R(\bar{g}) \leq \frac{196b^{15}}{d} \left[\sum_{j=0}^8 \binom{8}{j} 2^j (3d/2)_j \right] \alpha^6.$$

The constant depends on fixed dimension and the bounded spike, not on m . This absolute bound is intentionally loose; it is not dimension-uniform relative efficiency, cannot remove V_m , and does not give a sharp finite-spike sample prescription. The same argument gives $R(g_{M_Z}) - R(\bar{g}) = O_{d,\Lambda}(\alpha^4)$. Posterior linear response and covariance differentiation are classical ingredients; the explicit complete-context correction and its true regression-risk control are their application here.

K.4 Closed observable reference moments in two dimensions

For $d = 2, r = 1$, write

$$P = \frac{1}{2} \begin{pmatrix} 1 + U_1 & U_2 \\ U_2 & 1 - U_1 \end{pmatrix}, \quad U = (\cos 2\theta, \sin 2\theta).$$

The exact reference posterior density is proportional to $e^{v^T U}$, with $v = \alpha(S_{Z,11} - S_{Z,22}, 2S_{Z,12})/4$. Let $k = \|v\|$ and $n = v/k$ when $k > 0$. If $A_j = I_j(k)/I_0(k)$, the Fourier integrals give

$$\mathbb{E}U = A_1 n, \quad \text{Cov}(U) = \frac{1-A_2}{2} I + (A_2 - A_1^2) nn^T.$$

At $k = 0$, these are 0 and $I/2$. Scaled Bessel functions evaluate the ratios without the exponential overflow of separate I_j values. For a symmetric matrix T , define $c(T) = ((T_{11} - T_{22})/2, T_{12})^T$, so $\text{tr}(PT) = \text{tr} T/2 + c(T)^T U$. For the correction, set

$$K = \frac{qX^T + Xq^T}{2s} - b_D \frac{XX^T}{s}, \quad C = \frac{XX^T}{s} + \frac{1}{2} \left(XX^T + qq^T + (1-a) \frac{XX^T}{s} \right).$$

Then $k_P = \text{tr}(PK)$, $z_P + S_P = \text{tr}(PC)$ and the correction coefficient is $y c(K)^T \text{Cov}(U) c(C)$. This computes the finite-reference rule exactly up to special-function rounding; full regression-posterior quadrature is used for verification only, not by the rule itself. For larger dimension the first two Bingham moments may be computed by differentiating its normalizer or by a separately diagnosed integration scheme. No high-dimensional posterior solver is certified by the two-dimensional implementation.

Appendix L Complete frozen method comparison

λ	ρ	m	A_{iso}	A_{PCA}	A_{soft}	A_{corr}
0.0625	0	0	$3.71e-07$	$2.15e-04$	$3.71e-07$	$2.28e-10$
0.0625	0.25	64	$6.26e-06$	$1.75e-04$	$2.81e-07$	$2.95e-10$
0.0625	1	256	$2.00e-05$	$1.31e-04$	$3.91e-07$	$5.44e-10$
0.0625	4	1024	$6.23e-05$	$3.40e-05$	$1.25e-07$	$2.18e-10$
0.125	0	0	$9.15e-06$	$1.08e-03$	$9.15e-06$	$2.07e-08$
0.125	0.25	16	$1.71e-05$	$6.04e-04$	$3.91e-06$	$1.56e-08$
0.125	1	64	$1.20e-04$	$5.48e-04$	$4.04e-06$	$1.35e-08$
0.125	4	256	$2.51e-04$	$2.05e-04$	$1.71e-06$	$8.47e-09$
0.25	0	0	$7.43e-05$	$2.73e-03$	$7.43e-05$	$3.14e-07$
0.25	0.25	4	$1.55e-04$	$2.97e-03$	$7.73e-05$	$5.82e-07$
0.25	1	16	$3.31e-04$	$2.01e-03$	$3.91e-05$	$3.72e-07$
0.25	4	64	$1.00e-03$	$1.02e-03$	$2.82e-05$	$5.03e-07$

All entries are empirical means of the conditionalized algorithmic regret on the complete frozen grid. Per-method SE, raw target losses, actual paired-loss differences,

negative taskwise differences and posterior refinement diagnostics are included in the machine-readable records. No additional grid cells have been selected or generated for this presentation.

References

- E. Akyol, K. B. Viswanatha and K. Rose. On conditions for linearity of optimal estimation. *IEEE Transactions on Information Theory* 58(6):3497–3508, 2012. <https://doi.org/10.1109/TIT.2012.2188850>.
- O. Besson, N. Dobigeon and J.-Y. Tournieret. Minimum mean square distance estimation of a subspace. *IEEE Transactions on Signal Processing* 59(12):5709–5720, 2011. <https://doi.org/10.1109/TSP.2011.2166548>.
- T. T. Cai, Z. Ma and Y. Wu. Sparse PCA: Optimal rates and adaptive estimation. *Annals of Statistics* 41(6):3074–3110, 2013. <https://doi.org/10.1214/13-AOS1178>. Original comparison version: arXiv:1211.1309v3.
- E. Chaumette, A. Renaux and M. N. El Korso. A class of Weiss–Weinstein bounds and its relationship with the Bobrovsky–Mayer–Wolf–Zakai bounds. *IEEE Transactions on Information Theory* 63(4):2226–2240, 2017. <https://doi.org/10.1109/TIT.2017.2671883>.
- D. A. Cohn, Z. Ghahramani, and M. I. Jordan. Active learning with statistical models. *Journal of Artificial Intelligence Research*, 4:129–145, 1996. Original author version: <https://arxiv.org/abs/cs/9603104>.
- M. De Lara and O. Gossner. Payoffs–beliefs duality and the value of information. *SIAM Journal on Optimization* 30(1):464–489, 2020. <https://doi.org/10.1137/18M1230049>.
- J. T. Flåm, S. Chatterjee, K. Kansanen and T. Ekman. On MMSE estimation: A linear model under Gaussian mixture statistics. *IEEE Transactions on Signal Processing* 60(7):3840–3845, 2012. <https://doi.org/10.1109/TSP.2012.2192112>.
- D. T. Frazier and D. J. Nott. Cutting feedback and modularized analyses in generalized Bayesian inference. *Bayesian Analysis* 20(4):1647–1675, 2025. <https://doi.org/10.1214/24-BA1448>.
- D. T. Frazier and D. J. Nott. Posterior risk of modular and semi-modular Bayesian inference. *Journal of the American Statistical Association* 121(554):1588–1600, 2026. <https://doi.org/10.1080/01621459.2025.2580690>.
- R. D. Gill and B. Y. Levit. Applications of the van Trees inequality: A Bayesian Cramér–Rao bound. *Bernoulli* 1(1–2):59–79, 1995. <https://doi.org/10.2307/3318681>.

- R. Giordano, T. Broderick and M. I. Jordan. Covariances, Robustness, and Variational Bayes. *JMLR* 19(51):1–49, 2018.
- D. Guo, Y. Wu, S. Shamai (Shitz) and S. Verdú. Estimation in Gaussian noise: properties of the minimum mean-square error. *IEEE Transactions on Information Theory* 57(4):2371–2385, 2011. <https://doi.org/10.1109/TIT.2011.2111010>. Comparison version: arXiv:1004.3332v1.
- P. D. Hoff. Simulation of the matrix Bingham–von Mises–Fisher distribution, with applications to multivariate and relational data. *Journal of Computational and Graphical Statistics* 18(2):438–456, 2009.
- J. Piironen, M. Paasiniemi and A. Vehtari. Projective inference in high-dimensional problems: Prediction and feature selection. *Electronic Journal of Statistics* 14(1):2155–2197, 2020. <https://doi.org/10.1214/20-EJS1711>.
- C. Radhakrishna Rao. Antieigenvalues and antisingularvalues of a matrix and applications to problems in statistics. *Mathematical Inequalities & Applications* 10(3):471–489, 2007.
- A. Renaux, P. Forster, P. Larzabal, C. D. Richmond and A. Nehorai. A fresh look at the Bayesian bounds of the Weiss–Weinstein family. *IEEE Transactions on Signal Processing* 56(11):5334–5352, 2008. <https://doi.org/10.1109/TSP.2008.927075>.
- H. G. Sung. *Gaussian Mixture Regression and Classification*. PhD dissertation, Rice University, 2004. Sections 4.1–4.2 and Theorem 4.5.1. <https://www.stat.rice.edu/~hgsung/thesis.pdf>.
- M. Tsang. Physics-inspired forms of the Bayesian Cramer–Rao bound. *Physical Review A* 102, 062217, 2020. arXiv:2007.04849v3.
- M. Wahl. Van Trees inequality, group equivariance, and estimation of principal subspaces. In *Foundations of Modern Statistics*, Springer Proceedings in Mathematics & Statistics 425, pp. 301–321, 2023. https://doi.org/10.1007/978-3-031-30114-8_8.
- A. Xu and M. Raginsky. Minimum excess risk in Bayesian learning. *IEEE Transactions on Information Theory* 68(12):7935–7955, 2022. <https://doi.org/10.1109/TIT.2022.3176056>.