SHARP LOW-DEGREE PREDICTION EXPONENTS FOR NOISY POLYNOMIAL SINGLE-INDEX MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Let X be standard Gaussian in $\mathbb{R}^D$ and let $Y = f(u \cdot X) + sG$, where u is an unknown unit vector, G is independent standard Gaussian noise, and f is a fixed centered polynomial with $\mathbb{E}f(Z)^2 = 1$ for standard Gaussian Z. Suppose its lowest nonzero Hermite order is $p \geq 3$. For $s = D^a$, $a > 0$, we determine the sample exponent $B_f(a)$ for polynomial prediction of $f(u \cdot x)$ from finitely many transformed-moment indices. Below $B_f(a)$ by any fixed positive margin, expected minimax risk tends to one for label degree a positive power of D, even with arbitrary measurable design and query coefficients. Above it by any fixed positive margin, an explicit predictor of total degree $O(\log D)$ in the design, labels and query has expected risk tending to zero, with supplied link coefficients and no noise parameters. The upper constructions use known transformed-moment recovery ingredients. For pure odd Hermite links, the upper scale $\min\{Ds^4, D^{p/2}s^2\}$ already follows for recovery, up to logarithms, from existing quantitative single-index bounds. The unrestricted fixed-accuracy sample requirement is $\Theta_P(Ds^2)$ for link degree at most P. Hermite rank alone does not determine the polynomial exponent: for fixed odd rank-three links of arbitrary degree, the possible curves are exactly $B_f(a) = \min\{1 + \kappa a, 3/2 + 2a\}$ for even $\kappa \geq 4$. Thus every fixed $a > 0$ admits a fixed link attaining $3/2 + 2a$, whose degree may depend on a. A degree-21 example gives $7/4$ at $a = 1/8$, versus the pure cubic's $3/2$.

1 INTRODUCTION

For every fixed centered normalized polynomial link of Hermite rank at least three, we determine the polynomial prediction exponent under Gaussian design and noise (Section 2.2). Rank alone does not determine this exponent: at noise standard deviation $s = D^{1/8}$, a fixed degree-21 rank-three link has exponent $7/4$, versus the pure cubic's $3/2$. The lower restricts label degree but allows arbitrary design/query processing. The matching upper has logarithmic joint degree and receives the coefficients. We first develop the pure Hermite case.

Let $h_p = \mathrm{He}_p/\sqrt{p!}$ be the normalized Hermite polynomial for fixed odd $p \geq 3$; in particular, $h_3(z) = (z^3 - 3z)/\sqrt{6}$. The observations and prediction target are

$$X_i \sim \mathcal{N}(0, I_D), \qquad Y_i = h_p(u \cdot X_i) + sG_i, \qquad f_u(x) = h_p(u \cdot x), \qquad \|u\| = 1, \tag{1}$$

where all design rows and noise variables are independent and $G_i \sim \mathcal{N}(0, 1)$. The entire design is observed. The fresh input $x \sim \mathcal{N}(0, I_D)$ is independent of training. The parameters p, D, s are known; no information correlated with u is supplied separately.

Affine, detection and scalar-signal comparisons concern the pure link in (1).

The cubic Hermite link is used as a benchmark for learning dynamics by Tsiolis et al. (2025, Figure 1).

We study the expected squared error in predicting $f_u(x)$. The zero predictor has risk one, since $\mathbb{E}[f_u(x)^2] = 1$. The labels have zero linear correlation with the inputs: $\mathbb{E}[YX] = 0$ for $p \geq 3$. The cubic transformation instead satisfies $\mathbb{E}[Y^3X] = A_p u$, where $A_p = \mathbb{E}[Z(h_p(Z) + sG)^3] > 0$ for independent standard Gaussians Z, G. This constant depends only on p, not on s. Proposition F.1 and Appendix G give the exact moment and vector identities. Such transformed signals underlie the

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

generative exponent of Damian et al. (2024). The signal remains nonzero as noise grows, but its sampling error increases. We determine the prediction sample exponent under growing noise and compare it with a sufficient sample count for detecting dependence.

We study low-degree prediction in the observed data and fresh input. Our lower holds for the larger class of polynomials in the training labels with arbitrary measurable coefficients in the entire design and query. The matching upper has logarithmic total degree in all these variables. For this pure link, both classes have the same sample exponent, corresponding to the scale

$$\min\{Ds^4, D^{p/2}s^2\}. \tag{2}$$

For $p = 3$, the crossover is at $s = D^{1/4}$ (Figure 1(a)). For general fixed odd p, it occurs at $D^{(p-2)/4}$. Below the crossover the quadratic-label matrix statistic gives the smaller sample count. Above it an order-p tensor statistic gives the smaller count.

For the cubic link at $s = D^{1/8}$, unrestricted prediction and the sufficient detection count have exponent $5/4$, while polynomial prediction has exponent $3/2$. At $N = \lceil D^{21/16} \rceil$, a label-degree-six score detects dependence and unrestricted prediction risk vanishes, but the stated growing-label-degree minimax risk tends to one. Corollary G.3 gives the explicit rates and the null preserving the joint label/input-norm law. The lower permits arbitrary design processing; it does not bound every polynomial-time learning algorithm.

1.1 Prior bounds and the comparison proved here

Damian et al. (2024) define the generative exponent using $\lambda_k^2 = \mathbb{E}[\mathbb{E}[h_k(Z) \mid Y]^2]$. Their Example 2.7 and Theorem 5.2 give generative exponent one for every non-even polynomial link with additive Gaussian noise. The index alone does not quantify the shrinking signals as s grows. Their Theorem 4.1 in the cited version gives recovery up to sign, at fixed directional accuracy, using $O_k((D^{k/2}+D)\log^{2k}(3/\lambda_k)/\lambda_k^2)$ samples. The moment bounds proved in Section 4 give $\lambda_2^2 \gtrsim_p s^{-4}$ and $\lambda_p^2 \gtrsim_p s^{-2}$ for $s \geq 1$. Thus both scales in (2), up to logarithms, already follow for that recovery task. Their label transformations and normalizations need not be polynomials in the labels.

We prove that the smaller of these two scales is necessary for prediction throughout the noise range, up to fixed positive margins in the sample exponent. The lower applies to every polynomial of the specified label degree, with arbitrary measurable coefficients in the full design and query. We construct predictors attaining the same exponent in this class, with full expected-risk control, using logarithmic joint degree. Thus the lower remains valid after granting arbitrary design processing, while the upper needs only polynomial processing.

Damian et al. (2025) extend the generative exponent to Gaussian multi-index models. Their Theorem 3 gives subspace recovery with a channel-dependent constant $C(\mathsf{P}, K)$, using a kernel over labels. Their Theorem 1 is a detection lower in input degree. Our channel changes with dimension, and its conditional signal strengths tend to zero at different rates. Section 4 evaluates these rates, including optimization over square-integrable transformations of one label. Recovery without the link being supplied is already treated by Damian et al. (2024, Corollary 4.4) and Damian et al. (2025, Lemma 3 and Algorithms 1–2). For the family in Corollary 2.4, our coefficient-independent construction attains the growing-noise prediction exponent inside the specified joint-polynomial class, with full expected-risk control and without the noise level being supplied.

The lower uses the conditional-cumulant method of Schramm and Wein (2022). Polynomial-degree estimation bounds and graph-excess arguments also appear in Sohn and Wein (2026). We apply this method to a Hermite prediction target while conditioning on the revealed Gaussian-vector design. The coordinate count in Section A supplies the two competing noise exponents in (2).

Joshi et al. (2025) study single-index models under spherically symmetric inputs. Their Theorem 5 gives a detection lower bound for polynomials in the input directions with unrestricted dependence on the labels and input norms, and permits dimension-dependent channels. Our prediction class instead allows arbitrary dependence on the design and restricts label degree. The distinction between the tasks is quantitative here: Section 3 exhibits a sample range where a polynomial score detects dependence but polynomial prediction fails. The null retains the label/norm information available in their model.

Defilippis et al. (2026) identify a noise-sensitivity exponent governing high-noise statistical-to-computational gaps. For even single-index links, their Theorems 3.2 and 3.3 give the optimal approximate-message-passing and statistical weak-recovery thresholds, respectively, in a proportional limit followed by high noise. The even-link high-noise gap is therefore known. Our finite-profile rule treats simultaneous growth of noise and dimension for every fixed polynomial link of rank at least three. It determines the sample exponent for full expected prediction risk in the stated polynomial class.

Luo and Zhang (2024) study a related rank-one tensor regression problem with independent Gaussian entries in each tensor covariate. Their Theorem 9 in the cited version gives a detection lower bound with a related high-noise scale, restricting joint degree in covariates and responses. Our design consists of Hermite features of a Gaussian vector and its entries are dependent. We restrict only the response degree and bound prediction risk directly. Equality of feature covariances does not identify the two design distributions.

2 Prediction classes and main result

Write γ_D for standard Gaussian measure. Let $\mathcal{Q}_J$ be the class of polynomials of total degree at most J in all scalar coordinates of the training design, training labels, and fresh query. After fixing an independent seed, their coefficients are real numbers that may depend on known parameters and measurably on the seed, but not on the unknown teacher. We first state the result for a larger class. We allow unrestricted processing of the observed Gaussian inputs and bound the degree with which the training labels enter the prediction. A predictor belongs to $\mathcal{P}_L$ if it has a representation

$$\widehat{f}(x) = \sum_{\substack{\alpha\in\mathbb{N}^N \\ |\alpha|\le L}} a_\alpha(X_1,\dots,X_N,x,R)Y^\alpha, \tag{3}$$

where $\mathbb{N}$ includes zero, $Y^\alpha = \prod_i Y_i^{\alpha_i}$, and each coefficient is jointly measurable. The variable R is independent learner randomness. Coefficients may depend on the known parameters and have unrestricted dependence on the entire design and fresh input. They cannot depend on the unknown u. The output need not be a single Hermite ridge.

To read (3), fix the design, the query and the seed. The prediction is then a polynomial of total degree at most L in the N label variables. At $L = 1$, it has the form $a_0 + \sum_i a_i Y_i$, with arbitrary measurable weights determined by the design, query and seed. Larger L permits products of labels; processing the design does not consume this degree allowance. Proposition H.1 gives another instance: unnormalized gradient descent on squared loss for a Hermite ridge, with label-independent initialization, step sizes and sample selection, produces a polynomial in the labels. The worst-teacher prediction lower applies at each fixed iteration count in the proposition's explicit logarithmic range.

For integers $D \ge 2$, $N, L \ge 0$, fixed odd $p \ge 3$ and $s \ge 1$, define

$$\mathcal{R}(D,N,L,p,s) = \inf_{\widehat{f}\in\mathcal{P}_L} \sup_{\|u\|=1} \mathbb{E}_u[(\widehat{f}(x) - f_u(x))^2]. \tag{4}$$

The expectation includes all training data, R, and x. Infinite risk is allowed. The zero predictor has risk one, because $\|f_u\|_{L^2(\gamma_D)} = 1$.

Theorem 2.1. *Fix an odd integer $p \ge 3$ and real numbers $a, \eta > 0$. Put*

$$b = \min\{1+4a, p/2+2a\}, \qquad \tau = \min\{2a, \eta/2\},$$
$$L_- = \left\lfloor D^{\tau/[4(3p+5)]} \right\rfloor, \qquad L_+ = 2p\left\lceil \log_2(100pD^{1+\eta/4}) \right\rceil + 1.$$

For all sufficiently large integers D, with $s = D^a$,

$$\mathcal{R}(D,N,L_-,p,D^a) \ge 1 - 2D^{-\tau/2} \qquad \text{if } 0 \le N \le D^{b-\eta}, \tag{5}$$
$$\mathcal{R}(D,N,L_+,p,D^a) \le D^{-\eta/4} \qquad \text{if } N \ge \lceil D^{b+\eta} \rceil. \tag{6}$$

The dimension threshold can depend on p, a, η. The upper predictor has polynomial real-arithmetic work and storage for fixed p.

The lower permits a positive power of D in label degree, whereas the upper uses $O_{p,\eta}(\log D)$. In particular $L_+ \le L_-$ eventually. Since the classes are nested, the lower also holds at L_+ and the upper also holds at L_- for sufficiently large D. These statements retain specified degree ranges; no dimensionwise unrestricted union of polynomial classes is used.

Appendix N deduces Theorem 2.1 from the complete finite-sample lower and upper bounds.

Let $\mathcal{R}_{\text{joint}}(D, N, J, p, s)$ denote (4) with $\mathcal{P}_L$ replaced by $\mathcal{Q}_J$. The following consequence gives the same exponent for joint degree. Appendix C gives the complete degree count.

Corollary 2.2. *Fix odd $p \ge 3$ and $a, \eta > 0$, and let b, τ, L_- be as in Theorem 2.1. Put*

$$J_+ = 2p(p+1)\left\lceil \log_2(100pD^{1+\eta/4})\right\rceil + 2p + 1.$$

For all sufficiently large integers D and every integer $N \ge 0$, with $s = D^a$,

$$\mathcal{R}_{\text{joint}}(D, N, L_-, p, s) \ge 1 - 2D^{-\tau/2} \qquad \text{if } N \le D^{b-\eta},$$
$$\mathcal{R}_{\text{joint}}(D, N, J_+, p, s) \le D^{-\eta/4} \qquad \text{if } N \ge \lceil D^{b+\eta}\rceil.$$

Eventually $J_+ \le L_-$, and both bounds hold at every integer degree between J_+ and L_-.

2.1 Proof structure

The complete lower proof is in Section A. The target f_u lies in the span of the degree-p multivariate Hermite basis functions. Projecting a finite-risk predictor onto this span preserves label degree and decreases risk. A conditional cumulant bound in this finite-dimensional function space bounds the improvement over risk one by a sum of squared cumulant norms. We apply the certificate with the entire design fixed and then average over that design. This permits arbitrary measurable design coefficients in the prediction class.

For label degree k and r distinct observed rows, the coordinate count gives a dimension loss of at least

$$\delta_p(k, r) = \max\{r, (p-1)r - (p-2)k/2 + p/2\}.$$

The first term follows from connectedness. The second follows from the cycle rank of a marked graph recording cross-replica Gaussian pairings and target contractions. Both are needed across the noise crossover. In the finite energy bound of Theorem A.1, the degree-k contribution involving r distinct observed rows has a factor $N^r s^{-2k} D^{-\delta_p(k,r)}$. For $s = D^a$ and $N \le D^{b-\eta}$, Corollary A.3 uses the two branches of δ_p to prove

$$N^r s^{-2k} D^{-\delta_p(k,r)} \le D^{(b-\eta)r - 2ak - \delta_p(k,r)} \le D^{-\tau k}, \qquad 1 \le r \le \min\{N, k\}.$$

Here b and τ are as in Theorem 2.1. This decay absorbs the combinatorial factor $(2pk)^{(3p+5)k}$ through the theorem's allowed label degree. Summing the resulting energies bounds the total possible improvement over risk one.

For the upper, the final output must remain polynomial in the labels while its full expected squared error is controlled. Section B constructs two unbiased matrix estimates of $uu^\top$. One uses $(Y^2 - 1 - s^2)(XX^\top - I)/(2p)$. The other uses the symmetrized product of two independent tensor estimates after flattening. Products of independent matrix blocks control the second moment along u and in its perpendicular space. The matrix $uu^\top$ does not retain the sign of u, which matters for odd p. A final independent order-p tensor supplies the signed target without eigenvector selection or label-dependent normalization. Applying p independent such products to this tensor estimate produces an unbiased prediction tensor. Independence keeps its expected squared error within the second-moment calculation while its total label degree is $O(\log D)$.

The extension below retains the same dimension loss $\delta_p(k, r)$ when different fixed Hermite orders of one parity occur at a row. Appendix I proves this count after separating the target chaoses. Appendix J checks the link-dependent covariance bounds and constructs all target orders using the same matrix products. To remove knowledge of the link coefficients, Appendix K replaces their scalar normalization by a polynomial in independent trace estimates. Its matrix-product bound keeps the additional parallel variance separate from the perpendicular error, so this calibration incurs no extra power of dimension in the sample count. A deterministic allocation between the two constructions also removes knowledge of the noise level. Appendix Q retains moment cancellations by grouping row powers before the Hermite expansion, then proves the general rule with matching predictors.

2.2 Fixed polynomial links

Let f be a known link of the form

$$f(z) = \sum_{q=p,p+2,\dots,P} c_q h_q(z), \qquad c_p \neq 0, \qquad \sum_q c_q^2 = 1, \tag{7}$$

where $3 \le p \le P$ are integers of the same parity and all coefficients are fixed independently of D and s. In this subsection the observations are $Y_i = f(u \cdot X_i) + sG_i$ and the target is $f(u \cdot x)$. Let $\mathcal{R}_f(D, N, L, s)$ and $\mathcal{R}_{f,\text{joint}}(D, N, J, s)$ be the minimax risks over $\mathcal{P}_L$ and $\mathcal{Q}_J$, respectively, for this target and model. The coefficients of the predictors may depend on the known link. In the theorem below, the smallest nonzero order p determines the sample exponent; P enters the constants, degree range, and computational cost. The tensor branch estimates $uu^\top$ from order p, while the final output applies matrix products to each target chaos.

Theorem 2.3. *Fix the link (7), suppose $\beta_f = \mathbb{E}[(Z^2-1)f(Z)^2] \neq 0$ for standard Gaussian Z, and fix $a, \eta > 0$. Put*

$$\varepsilon_{p,P}(D) = \mathbf{1}_{\{p \text{ even}\}}(P-1)!!D^{-p/2}.$$

$$b = \min\{1+4a, p/2+2a\}, \qquad \tau = \min\{2a, \eta/2\},$$
$$L_f = \left\lfloor D^{\tau/[4(3P+6)]} \right\rfloor, \qquad J_f = 2P(p+1)\left\lceil \log_2(100PD^{1+\eta/4}) \right\rceil + 2P + 1.$$

For all sufficiently large integers D, every integer $N \ge 0$, and $s = D^a$,

$$\mathcal{R}_f(D, N, L_f, s) \ge 1 - \varepsilon_{p,P}(D) - 2D^{-\tau/2} \qquad \text{if } N \le D^{b-\eta},$$
$$\mathcal{R}_{f,\text{joint}}(D, N, J_f, s) \le D^{-\eta/4} \qquad \text{if } N \ge \lceil D^{b+\eta} \rceil.$$

Eventually $J_f \le L_f$. Both polynomial classes have sample exponent b at every integer degree between J_f and L_f.

Appendix J.5 gives the proof.

The lower does not need $\beta_f \neq 0$; this condition supplies the quadratic-label branch of the matching upper. For the mixed link, the cross moments $\mathbb{E}[h_p(Z)Y] = c_p$ and $\mathbb{E}[h_2(Z)(Y^2-1-s^2)] = \beta_f/\sqrt{2}$ also give the two sufficient exponents in the cited recovery bound. The extension permits higher terms of the same parity without introducing a lower Hermite order. For odd links, the quadratic condition is automatic through degree 15: every normalized link supported on $\{3, 5, \dots, 15\}$ has $\beta_f \ge 1/8$ (Proposition J.5). At degree 17, β_f can vanish even with $c_3 \neq 0$. For odd rank-three links with $\mathbb{E}[Zf(Z)^3] \neq 0$, including this witness, the same upper exponent holds without supplied coefficients under unknown iid centered unit-variance noise with a fixed sixth-moment bound (Corollary P.2). The constants need not be uniform when the link varies with dimension.

The next corollary uses one predictor collection for all fixed links in this family. The learner receives the least order p and an upper bound P on the degree, but no link coefficients, noise level, or positive lower bounds on the nonzero signals.

Corollary 2.4. *Fix integers $3 \le p \le P$ of the same parity and $\eta > 0$. There is a collection $\widehat{f}_{p,P,D,N,\eta}$ of predictors, indexed by integers $D \ge 2$ and $N \ge 0$, whose coefficients depend only on these indices and whose joint degrees are $O_{p,P,\eta}((\log(eD))^3)$. For every fixed link satisfying (7) and $\beta_f \neq 0$, and every fixed $a > 0$, put $s = D^a$ and $b = \min\{1+4a, p/2+2a\}$. For all sufficiently large D and every integer $N \ge \lceil D^{b+\eta} \rceil$,*

$$\sup_{\|u\|=1} \mathbb{E}\big(\widehat{f}_{p,P,D,N,\eta}(x) - f(u \cdot x)\big)^2 \le D^{-\eta/8}.$$

These predictors have polynomial real-arithmetic work and storage for fixed p, P, η.

Proof. Theorem K.6 constructs the collection with exactly these inputs, degree bounds, and expected-risk guarantee. □

The threshold dimension may depend on the fixed link and a, η. For fixed positive lower bounds on $|c_p|$ and $|\beta_f|$, it is uniform over that coefficient family, with p, P, a, η fixed. No optimality of these degree bounds is asserted.

The upper in Corollary 2.4 and the weak-linear upper below hold uniformly when G_i are replaced by iid noise independent of the inputs, with mean zero, variance one, and fourth moment at most a fixed finite $M \geq 1$. The law, M, and s are not supplied; the dimension threshold may depend on M. Proposition K.7 and Theorem M.1 prove the respective extensions. The lower bounds retain Gaussian noise.

Now let $f = \sum_{q=p}^{P} c_q h_q$ be any fixed centered polynomial, with $c_p \neq 0$, $\sum_q c_q^2 = 1$ and $p \geq 3$, allowing both parities. For $Z \sim N(0,1)$ and $s = D^a$, $a > 0$, its prediction exponent is

$$m_j = \min\{m \in \mathbb{N}_{\geq 1} : \mathbb{E}[\mathrm{He}_j(Z) f(Z)^m] \neq 0\}, \quad 1 \leq j \leq p,$$

$$C_* = 2 \min\left\{ m_2, \min_{3 \leq j \leq p} \frac{(j-2)m_1 + m_j}{j-1} \right\},$$

$$B_f(a) = \min\left\{ 1 + C_* a, \min_{3 \leq j \leq p} (j/2 + 2m_j a) \right\}.$$

Empty minima are $+\infty$; C_* is finite (Lemma Q.1). The index m_2 is the noise-sensitivity exponent of Defilippis et al. (2026, arXiv version 1, Definition 2.1). After label transformation, quadratic matrices or tensor contractions with first-order pilots give the $1 + C_* a$ branch. Tensor Gram estimates give the other active branches. Corollary Q.4 gives minimax risk tending to one below $B_f(a)$ by any fixed positive margin under Gaussian noise, for label degree a positive power of D and arbitrary design/query coefficients. Proposition Q.5 gives expected risk $o(1)$ above it by the same margin, with joint degree $O(\log D)$, supplied coefficients and no noise parameters. The upper permits iid centered unit-variance noise independent of the inputs, with a fixed finite noise moment whose order depends on the link. Given p, P, a common collection of joint degree $O(\log^3 D)$ attains exponent $p/2 + 2a$ uniformly over centered unit-variance iid noise laws, without supplied f, s or noise law; its dimension threshold may depend on the fixed link (Proposition Q.6, Corollary Q.7). For fixed odd rank-three links under Gaussian noise, $\kappa = C_* = \min\{2m_2, m_1 + 1\}$ and $B_f(a) = \min\{1 + \kappa a, 3/2 + 2a\}$. Across fixed links of arbitrary degree, the possible κ are exactly $4, 6, 8, \ldots$ (Corollary Q.11). For each fixed $a > 0$, some fixed link has exponent $3/2 + 2a$; its degree may depend on a. The degree-21 witness (Proposition Q.8) has $\kappa = 6$, giving $7/4$ at $a = 1/8$ versus the pure cubic's $3/2$.

2.3 Fixed links of mixed parity

Mixing even and odd orders can change the noise exponent. Fix integers $3 \leq p \leq P$ and a link $f = \sum_{q=p}^{P} c_q h_q$, with coefficients fixed independently of D, s, $\sum_q c_q^2 = 1$, $c_p \neq 0$, and

$$\alpha_f = \mathbb{E}[Z f(Z)^2] = 2 \sum_{q=p}^{P-1} \sqrt{q+1}\, c_q c_{q+1} \neq 0.$$

For $s = D^a$, $a > 0$, Theorem O.1 proves the polynomial-prediction exponent

$$B_f(a) = \min\left\{ 1 + \left(4 - \frac{2}{p-1}\right) a, \frac{p}{2} + 2a \right\}.$$

Below this exponent by a fixed positive margin, minimax risk tends to one even for label degree a positive power of D, with arbitrary design and query coefficients. Above it, a joint polynomial of degree $O(\log D)$ has full expected risk at most $D^{-\eta/4}$ for margin $\eta > 0$. This upper receives the coefficients and permits the finite-fourth-moment noise laws above, without their scale, law or moment bound being supplied. The lower retains Gaussian noise. Given p, P, Theorem O.2 attains the same exponent without coefficients or noise parameters, with joint degree $O(\log^3 D)$ and expected risk at most $D^{-\eta/8}$. This collection assumes the same fixed nonzero c_p and α_f; its dimension threshold may depend on the link, noise exponent and finite fourth-moment bound.

For $f = (h_3 + t h_4)/\sqrt{1+t^2}$, fixed $t \neq 0$, we have $\alpha_f = 4t/(1+t^2)$ and $B_f(a) = \min\{1 + 3a, 3/2 + 2a\}$. The quadratic-label pilot and tensor contraction have recovery ancestry in Damian et al. (2024, arXiv version 2, Lemma F.4). Constants need not remain bounded as α_f or c_p tends to zero.

2.4 A VANISHING FIRST HERMITE COEFFICIENT

Fix a known odd link f_0 as in (7), with $\beta_{f_0} = \mathbb{E}[(Z^2-1)f_0(Z)^2] \neq 0$, and fix $a, b > 0$. Under the same Gaussian model, set

$$f_D(z) = \frac{f_0(z) + D^{-b}h_1(z)}{\sqrt{1+D^{-2b}}}, \qquad s = D^a,$$
$$B = \min\left\{1+4a, \frac{p}{2}+2a, 1+2a+\frac{2b(p-2)}{p-1}\right\}.$$

For each fixed $\eta > 0$, put $\tau = \min\{2a, \eta/2\}$ and $L = \lfloor D^{\tau/[4(3P+6)]} \rfloor$. For all sufficiently large D, every integer $0 \le N \le D^{B-\eta}$ has minimax risk at least $1 - D^{-2b} - 2D^{-\tau/2}$ in $\mathcal{P}_L$. For every integer $N \ge \lceil D^{B+\eta} \rceil$, a predictor in $\mathcal{Q}_{O_{p,P,\eta}(\log D)}$ has worst-teacher expected risk at most $D^{-\eta/4}$. The upper receives the link coefficients and permits the noise laws above. Both bounds concern the full target $f_D(u \cdot x)$. Theorems L.1 and M.1 give the complete lower and upper proofs, respectively.

The unperturbed exponent is unchanged exactly when

$$b \ge b_{\text{crit}} := (p-1)\min\left\{\frac{a}{p-2}, \frac{1}{4}\right\}.$$

For smaller b, the hybrid branch is strictly smaller. A fixed nonzero linear coefficient instead permits fixed-accuracy direction estimation at the Ds^2 scale.

The warm-start contraction bound of Damian et al. (2024, arXiv version 2, Lemma F.4) suggests the hybrid recovery scale for $b < (p-1)/4$. Montanari and Richard (2014) also study tensor recovery with linear side information. Here weighted coordinate counting gives the lower for the larger label-polynomial class, and the upper attains it with logarithmic joint degree and full expected-risk control. Corollary 2.5 gives statistical exponent $1+2a$, leaving excess $B-(1+2a) = \min\{2a, (p-2)/2, 2b(p-2)/(p-1)\}$. Its learner needs neither coefficients nor noise level; the polynomial upper still uses the coefficients.

2.5 STATISTICAL COMPARISON

For the pure link h_p, let $N_*(D,s,p)$ be the least integer sample count at which an arbitrary measurable learner has worst unit-teacher expected risk at most $1/10$. Theorem D.1 proves

$$\frac{Ds^2}{16p} < N_*(D,s,p) \le K_p Ds^2 \tag{8}$$

for every $D \ge 2$ and $s \ge 1$, with a finite constant K_p depending only on p. Finite-net median-of-means scores give the upper; hypercube testing gives the lower for improper, randomized learners. Adding a coefficient net extends the comparison to all normalized links supported on $\{1,\dots,P\}$, including both parities. Efficient search over the net is not claimed.

Corollary 2.5. *Fix an integer $P \ge 3$. For integers $D \ge 2$, $N \ge 0$ and a real $0 < \epsilon \le 1$, there is a learner depending only on P, D, N, ϵ whose expected risk is at most ϵ, uniformly over unit teachers and normalized links supported on $\{1,\dots,P\}$, whenever $s \ge 1$ and*

$$N \ge C_P^\star Ds^2\epsilon^{-2}\log(e/\epsilon).$$

Here $C_P^\star$ is finite and depends only on P. The bounds permit coefficients that vary with dimension. For every such link at each dimension, every unrestricted learner has worst-teacher expected risk at least $1 - 1/D - PN/((D-1)s^2)$, even when the link and noise are known.

Proof. Theorem D.6 constructs the common learner and Proposition D.7 proves the lower. □

For fixed P and $a > 0$, unrestricted minimax risk tends to one uniformly over these coefficients when $s = D^a$ and $N = o(D^{1+2a})$. With $\epsilon_D = 1/\log(eD)$, one collection depending only on P, D, N has vanishing expected risk above exponent $1+2a$ for every fixed $a > 0$ (Corollary D.8). The excess polynomial exponent for Theorem 2.3's fixed-link family is

$$b - (1+2a) = \min\{2a, p/2 - 1\}. \tag{9}$$

For the family in Theorem O.1, the excess is $\min\{2a(p-2)/(p-1), (p-2)/2\}$. Degree-dependent sample barriers for rotationally invariant kernels under uniform-sphere inputs were established by Ghorbani et al. (2021). Here the affine comparison allows arbitrary measurable Gaussian-design coefficients and keeps the dependence on growing noise explicit.

For the pure link h_p, the affine-label class has a larger sample requirement. Theorem E.1 proves, uniformly for $D \geq p^2$, $N \geq 1$ and $s \geq 1$,

$$\mathcal{R}(D, N, 1, p, s) = \Theta_p\big(\min\{1, s^2 D^p/N\}\big) .$$

Its upper is a deterministic shrinkage of the empirical Hermite-series predictor; its lower allows every measurable design-dependent affine rule. This class includes kernel ridge regression $k_x^\top (K + \lambda I)^{-1} Y$, where K is the kernel Gram matrix and k_x its query vector, provided the measurable, finite-valued positive-semidefinite kernel and $\lambda > 0$ are selected using only inputs and an independent seed. The theorem does not assert that every kernel attains the affine upper, and does not cover kernels and regularization chosen using labels. Table 1 summarizes the resulting prediction comparisons. Nonlinear label dependence reduces the affine exponent $p + 2a$ to $b \leq p/2 + 2a$.

3 Detection before polynomial prediction

We compare prediction with testing whether the labels retain an association with the input directions. Let P_u be the planted law in (1). Under the null Q, draw each pair (Y, ρ) with the same joint distribution as $(Y, \|X\|)$ under P_u, independently across observations. Draw an independent uniform direction ω on the unit sphere and set $X = \rho\omega$. Rotational invariance makes the distribution of $(Y, \|X\|)$ independent of u, so this specifies a known null. Its input marginal remains standard Gaussian, and no statistic of the labels and input norms alone distinguishes the hypotheses. Detection must therefore use their association with the input directions.

For $N \geq 2$, define the score

$$S_N = \frac{2}{N(N-1)} \sum_{i<j} Y_i^3 Y_j^3 \langle X_i, X_j \rangle. \tag{10}$$

Let A_p be the positive moment in (79). The test declares the planted model when $S_N > A_p^2/2$. Under P_u, the vectors $Y_i^3 X_i$ have mean $A_p u$; under Q, their mean is zero. Consequently S_N has mean A_p^2 under every planted direction and mean zero under the null. The score estimates the squared norm of the transformed mean without estimating its direction. Proposition G.1 proves a finite error bound uniformly over unit u. In particular, for fixed odd $p \geq 3$, the sum of testing errors tends to zero whenever $N/(\sqrt{D}\, s^6)$ tends to infinity, with $D \geq 2$ and $s \geq 1$. The score has label degree six and joint degree eight in the observed inputs and labels. Its thresholded decision is not itself a polynomial predictor. It can be evaluated in $O(ND)$ real-arithmetic operations by writing the numerator as $\|\sum_i Y_i^3 X_i\|^2 - \sum_i Y_i^6 \|X_i\|^2$.

For every fixed $0 < a < 1/4$, the interval

$$\max\{1 + 2a, 1/2 + 6a\} < r < 1 + 4a \tag{11}$$

is nonempty. Corollary G.2 shows that, for $s = D^a$ and $N = \lceil D^r \rceil$, the test succeeds, unrestricted prediction has expected risk tending to zero, and the minimax risk in a specified growing-label-degree class tends to one. This includes the cubic example in the introduction. A second-moment

Table 1: Sample exponents for the pure link h_p at worst-teacher expected prediction risk at most $1/10$, with fixed odd $p \geq 3$, $s = D^a$, and $a > 0$. The polynomial row uses the degree ranges in Theorem 2.1 and Corollary 2.2. Each row permits vanishing expected risk above its exponent by a fixed positive margin. The unrestricted fixed-accuracy law is sharp; no optimal dependence on the requested accuracy is claimed.

Prediction class	Sample exponent
Affine in labels, arbitrary design coefficients	$p + 2a$
Joint or label degree in the stated ranges	$\min\{1 + 4a, p/2 + 2a\}$
Unrestricted measurable predictors	$1 + 2a$

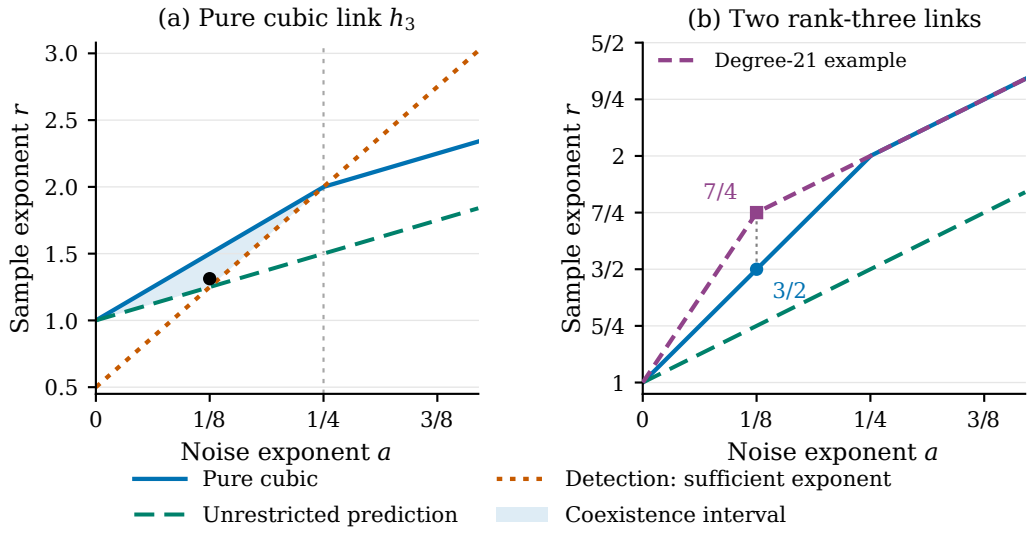

Figure 1: Analytic exponents for fixed $a > 0$, $s = D^a$ and $N = D^r$; $a = 0$ is a continuous endpoint. (a) Pure h_3 has prediction exponent $\min\{1 + 4a, 3/2 + 2a\}$ for both polynomial classes, allowing arbitrary design/query coefficients in the label-degree class. Unrestricted fixed-accuracy prediction has exponent $1 + 2a$. Dotted $1/2 + 6a$ is sufficient for detection, with no optimality claim. Shading $\max\{1 + 2a, 1/2 + 6a\} < r < 1 + 4a$ gives vanishing detection errors and unrestricted risk while polynomial minimax risk tends to one; the point is $(1/8, 21/16)$. (b) Under Gaussian noise, the specially constructed fixed degree-21 link (Proposition Q.8) has Hermite rank three, like the cubic, but exponent $\min\{1 + 6a, 3/2 + 2a\}$. Both have unrestricted exponent $1 + 2a$. At $a = 1/8$, cubic and example exponents are $3/2$ and $7/4$; the curves coincide for $a \geq 1/4$. Bounds use the stated degree ranges and fixed positive sample-exponent margins.

calculation bounds the error of this transformed-mean test. The prediction lower supplies the separation: successful detection does not suffice to predict in the stated class. Appendix G gives the null construction, complete variance calculation, and both comparisons.

4 Moment signals and the two scales

For the pure odd link h_p, define $\lambda_k^2 = \mathbb{E}[\mathbb{E}[h_k(Z) \mid Y]^2]$, where $Y = h_p(Z) + sG$ and Z, G are independent standard Gaussians. Proposition F.1 in Appendix F.1 proves $\lambda_1 > 0$ for every finite $s \geq 0$ and, for fixed p and $s \geq 1$, $\lambda_2^2 \gtrsim_p s^{-4}$ and $\lambda_p^2 \gtrsim_p s^{-2}$. These give the prior recovery scales; positivity alone gives no dimension-uniform lower bound. Appendix F proves

$$\lambda_1^2 \sim A_p^2/(6s^6), \qquad \lambda_2^2 \sim p^2/s^4, \qquad \lambda_p^2 \sim s^{-2}.$$

Each λ_k^2 is the supremum of $\mathbb{E}[h_k(Z)T(Y)]^2/\mathbb{E}[T(Y)^2]$ over $0 < \mathbb{E}[T(Y)^2] < \infty$. Among fixed harmonic orders in the cited recovery guarantees, Appendix F proves exponent b is minimal, attained at orders two and p; the order-one guarantee has sufficient scale Ds^6 up to logarithms. The separate cumulant argument gives the prediction lower.

The weak-linear upper fixes f_0, p, P, a, b and receives coefficients, but no noise scale. Coefficient-independent upper bounds apply to their stated fixed-link families; the statistical comparison is uniform over coefficients at each dimension. The exponent bounds use fixed positive margins; no optimal crossover constants, logarithmic factors or uniform resource bounds for growing link degree are claimed. Real-arithmetic cost is polynomial for each stated fixed family; thresholds and resource bounds may depend on the fixed link and its moment indices. Tensor storage and constants can be large; no practical runtime comparison or finite-bit implementation is claimed.

Reproducibility statement

All results have complete proofs in the appendices. Section 2.1 explains the main estimates, and Figure 1 plots the proved exponent formulas. Appendix Q.3 specifies the exact arithmetic certificate for the degree-21 example. The source includes its self-contained verifier `verify_double_cancellation.py`, which uses only the Python 3 standard library.

AI-use disclosure

The author designed and directed the research workflow, set research priorities, evaluated proposed directions, and checked the mathematical results. AI tools assisted with formulating mathematical claims, developing and checking proofs, literature searches, manuscript drafting and revision, and code development and verification. The author takes responsibility for the final content of this paper.

References

Alex Damian, Loucas Pillaud-Vivien, Jason D. Lee, and Joan Bruna. Computational-statistical gaps in Gaussian single-index models. arXiv:2403.05529v2, 2024. https://arxiv.org/abs/2403.05529v2.

Alex Damian, Jason D. Lee, and Joan Bruna. The generative leap: Tight sample complexity for efficiently learning Gaussian multi-index models. *Advances in Neural Information Processing Systems*, 38, 2025. https://proceedings.nips.cc/paper_files/paper/2025/hash/290056cc784d4825eb8cf5c72f808ebb-Abstract-Conference.html.

Leonardo Defilippis, Florent Krzakala, Bruno Loureiro, and Antoine Maillard. A noise sensitivity exponent controls large statistical-to-computational gaps in single- and multi-index models. arXiv:2603.17896v1, 2026. https://arxiv.org/abs/2603.17896v1.

Jingqiu Ding, Samuel B. Hopkins, and David Steurer. Estimating rank-one spikes from heavy-tailed noise via self-avoiding walks. *Advances in Neural Information Processing Systems*, 33, 2020. https://papers.nips.cc/paper_files/paper/2020/file/3c0de3fec9ab8a3df01109251f137119-Paper.pdf.

Jean Dolbeault, Maria J. Esteban, Michal Kowalczyk, and Michael Loss. Sharp interpolation inequalities on the sphere: New methods and consequences. *Chinese Annals of Mathematics, Series B*, 34(1):99–112, 2013. https://arxiv.org/abs/1210.1853v1.

Behrooz Ghorbani, Song Mei, Theodor Misiakiewicz, and Andrea Montanari. Linearized two-layers neural networks in high dimension. *Annals of Statistics*, 49(2):1029–1054, 2021. https://doi.org/10.1214/20-AOS1990.

Nirmit Joshi, Hugo Koubbi, Theodor Misiakiewicz, and Nati Srebro. Learning single index models via harmonic decomposition. *Advances in Neural Information Processing Systems*, 38, 2025. https://papers.neurips.cc/paper_files/paper/2025/hash/400693b9f90cd47ceae029071950179d-Abstract-Conference.html.

Dmitriy Kunisky. Hypothesis testing with low-degree polynomials in the Morris class of exponential families. *Proceedings of the Thirty Fourth Conference on Learning Theory*, PMLR 134:2822–2848, 2021. https://proceedings.mlr.press/v134/kunisky21a.html.

Yuetian Luo and Anru R. Zhang. Tensor-on-tensor regression: Riemannian optimization, over-parameterization, statistical-computational gap, and their interplay. arXiv:2206.08756v3, 2024. https://arxiv.org/abs/2206.08756v3.

Andrea Montanari and Emile Richard. A statistical model for tensor PCA. *Advances in Neural Information Processing Systems*, 27, 2014. https://proceedings.neurips.cc/paper_files/paper/2014/hash/cdb011b40de6e23c09618d0651fb6555-Abstract.html.

Tselil Schramm and Alexander S. Wein. Computational barriers to estimation from low-degree polynomials. *Annals of Statistics*, 50(3):1833–1858, 2022. https://doi.org/10.1214/22-AOS2179.

Youngtak Sohn and Alexander S. Wein. Sharp phase transitions in estimation with low-degree polynomials. arXiv:2502.14407v3, 2026. https://arxiv.org/abs/2502.14407v3.

Konstantinos Christopher Tsiolis, Alireza Mousavi-Hosseini, and Murat A. Erdogdu. From information to generative exponent: Learning rate induces phase transitions in SGD. *Advances in Neural Information Processing Systems*, 38, 2025. https://proceedings.neurips.cc/paper_files/paper/2025/hash/77b7d565f2370979b898d76d7ea27288-Abstract-Conference.html.

A Lower bound for polynomial prediction

We use a Gaussian estimation certificate based on cumulants (Schramm and Wein, 2022). Cumulant recursions and graph excess also enter the polynomial-degree recovery bounds of Sohn and Wein (2026). We prove the conditional prediction certificate and the required coordinate-counting bound below.

Fix an odd integer $p \geq 3$, integers $D \geq 1$ and $N, L \geq 0$, and $s > 0$. Write $h_j = \mathrm{He}_j / \sqrt{j!}$, where

$$\exp(tz - t^2/2) = \sum_{j \geq 0} \mathrm{He}_j(z) \frac{t^j}{j!}.$$

Let $\epsilon_1, \dots, \epsilon_D$ be independent uniform signs and put $U = \epsilon/\sqrt{D}$. Independently, let $A \in \mathbb{R}^{N \times D}$ have independent standard Gaussian entries and let $G \sim N(0, I_N)$. The entire design A and the labels

$$V_i = T_i(A, U) + G_i, \qquad T_i(A, U) = \frac{h_p(A_i^{\mathsf{T}} U)}{s} \tag{12}$$

are observed. For independent $x \sim \gamma_D = N(0, I_D)$, the target is $F_U(x) = h_p(U^{\mathsf{T}} x)$. An allowed predictor has the form

$$P(A, V, \xi; x) = \sum_{|\alpha| \leq L} a_\alpha(A, \xi, x) V^\alpha, \tag{13}$$

where ξ is independent learner randomness and the real coefficients are jointly measurable. Their dependence on A and x is unrestricted. We write

$$\mathrm{Risk}(P) = \mathbb{E}\big[(P(A, V, \xi; x) - F_U(x))^2\big],$$

allowing infinite risk. No teacher-dependent side information is supplied. Scaling the labels by the supplied positive number s preserves the class (13).

Let $\mathcal{H}_p$ be the homogeneous degree-p Gaussian Hermite chaos in $L^2(\gamma_D)$. For each fixed design define

$$\begin{aligned} m_\alpha &= \mathbb{E}_U[T^\alpha], & v_\alpha &= \mathbb{E}_U[F_U T^\alpha], \\ \kappa_\alpha &= v_\alpha - \sum_{\substack{\beta \leq \alpha \\ \beta \neq \alpha}} \binom{\alpha}{\beta} \kappa_\beta m_{\alpha - \beta}. \end{aligned} \tag{14}$$

Here multiindices have N coordinates, inequalities are coordinatewise, $|\alpha| = \sum_i \alpha_i$, $\alpha! = \prod_i \alpha_i!$, and $\binom{\alpha}{\beta} = \prod_i \binom{\alpha_i}{\beta_i}$. All expectations in (14) keep the design fixed. Put

$$E_k = \mathbb{E}_A \sum_{|\alpha| = k} \frac{\|\kappa_\alpha(A)\|_{\mathcal{H}_p}^2}{\alpha!}, \qquad c_p = 3p + 5. \tag{15}$$

Theorem A.1. *In the experiment* (12)*, every predictor* (13) *satisfies*

$$\mathrm{Risk}(P) \geq 1 - \sum_{k=0}^{L} E_k. \tag{16}$$

The energies are finite, $E_0 = 0$*, and* $E_k = 0$ *for even* k*. For every odd integer* $k \geq 1$*,*

$$\begin{aligned} E_k &\leq (2pk)^{c_p k} s^{-2k} \sum_{r=1}^{\min\{N,k\}} N^r D^{-\delta_p(k,r)}, \\ \delta_p(k, r) &= \max\left\{ r, (p-1)r - \frac{p-2}{2} k + \frac{p}{2} \right\}. \end{aligned} \tag{17}$$

An empty sum is zero. These bounds hold for all the stated integers p, D, N, L, k *and all* $s > 0$*.*

The bound (17) is finite-dimensional and uniform in its displayed parameters. No fixed-order limit enters its proof. At $p = 3$ its prefactor is $(6k)^{14k}$ and its loss exponent is $\max\{r, 2r - k/2 + 3/2\}$.

A.1 The conditional prediction certificate

For a standard Gaussian Z, multiplying two Hermite generating functions and taking expectations gives $\exp(tv)$. Comparing coefficients proves

$$\mathbb{E}[\mathrm{He}_j(Z)\,\mathrm{He}_\ell(Z)] = j!\,\mathbf{1}_{\{j=\ell\}}.$$

For a multiindex ν put $\psi_\nu(x) = \prod_j \mathrm{He}_{\nu_j}(x_j)/\sqrt{\nu!}$. The functions with $|\nu| = p$ form an orthonormal basis of $\mathcal{H}_p$. For every unit vector u, the coefficient of t^p in the generating function gives

$$\begin{aligned} F_u(x) &= \sum_{|\nu|=p} \sqrt{\frac{p!}{\nu!}}\, u^\nu \psi_\nu(x), \\ \|F_u\|_2^2 = 1, \qquad \langle F_u, F_w\rangle &= (u^\mathsf{T} w)^p. \end{aligned} \tag{18}$$

The last two identities follow from the multinomial formula. Oddness gives $F_{-u} = -F_u$, so $\mathbb{E}_U F_U = 0$. These identities include $D = 1$.

We first justify projecting the merely measurable coefficients in (13). Suppose the risk is finite. Then $\mathbb{E}\|P\|_2^2 < \infty$, since

$$\|P\|_2^2 \le 2\|P - F_U\|_2^2 + 2.$$

Conditional on any fixed design a, the label law is a finite mixture of Gaussian laws with identity covariance. It has a strictly positive density and all polynomial moments. Its finite monomial Gram matrix

$$J_a(\alpha,\beta) = \mathbb{E}[V^{\alpha+\beta} \mid A = a], \qquad |\alpha|, |\beta| \le L,$$

is positive definite: a nonzero polynomial cannot vanish Lebesgue almost everywhere. Let $\lambda_a > 0$ be its smallest eigenvalue. For almost every (a, ξ), Tonelli's theorem gives

$$\mathbb{E}[\|P(a, V, \xi; \cdot)\|_2^2 \mid a, \xi] \ge \lambda_a \sum_\alpha \|a_\alpha(a, \xi, \cdot)\|_2^2.$$

Thus every coefficient belongs to $L^2(\gamma_D)$ there. Project each one orthogonally onto $\mathcal{H}_p$. This preserves label degree and cannot increase risk, because every F_U lies in $\mathcal{H}_p$. The projection coefficients are measurable integrals against the finite basis ψ_ν, and they are finite by Cauchy–Schwarz. Assign zero on the exceptional design and seed null set. This gives a measurable $\mathcal{H}_p$-valued polynomial. No uniform lower bound on λ_a or measurable optimization is required.

We now prove (16) for that polynomial. Fix a and ξ, and abbreviate $T = T(a, U)$ and $F = F_U$. The target may depend on U beyond what the finite vector T determines. Write

$$q_\alpha(y) = \prod_i \frac{\mathrm{He}_{\alpha_i}(y_i)}{\sqrt{\alpha_i!}}, \qquad P(y) = \sum_{|\alpha|\le L} b_\alpha q_\alpha(y), \qquad b_\alpha \in \mathcal{H}_p.$$

The generating function implies the finite shift identity

$$q_\alpha(g + t) = \sum_{\beta \le \alpha} \sqrt{\frac{\beta!}{\alpha!}} \binom{\alpha}{\beta} t^{\alpha-\beta} q_\beta(g). \tag{19}$$

In particular,

$$\mathbb{E}\langle F, P(T + G)\rangle = \sum_\alpha \langle z_\alpha, b_\alpha\rangle, \qquad z_\alpha = \frac{v_\alpha}{\sqrt{\alpha!}}.$$

Jensen's inequality over U at fixed G, followed by Gaussian orthogonality in (19), gives

$$\begin{aligned} \mathbb{E}\|P(T + G)\|^2 &\ge \sum_\beta \|d_\beta\|^2, \\ d_\beta &= \sum_\alpha M_{\beta,\alpha} b_\alpha, \\ M_{\beta,\alpha} &= \mathbf{1}_{\{\beta\le\alpha\}} m_{\alpha-\beta} \sqrt{\frac{\beta!}{\alpha!}} \binom{\alpha}{\beta}. \end{aligned}$$

With $w_\beta = \kappa_\beta/\sqrt{\beta!}$, the recursion (14) says exactly $z = M^\mathsf{T} w$. Indeed, multiply its α coordinate by $\sqrt{\alpha!}$. The cross term is therefore $\sum_\beta \langle w_\beta, d_\beta \rangle$. Using $\mathbb{E}\|F\|^2 = 1$ and completing squares proves

$$\mathbb{E}\|P(T+G) - F\|^2 \geq 1 + \sum_\beta \|d_\beta - w_\beta\|^2 - \sum_\beta \|w_\beta\|^2$$
$$\geq 1 - \sum_{|\alpha| \leq L} \frac{\|\kappa_\alpha\|^2}{\alpha!}.$$

All sums and matrices are finite. Conditional signal moments exist because the sign prior is finite. By induction in (14), each basis coordinate of $\kappa_\alpha(a)$ is a polynomial in a, with coefficients depending on D, p, s and finitely many sign moments. Its squared norm has finite Gaussian-design expectation. Averaging over A, ξ proves (16). This finiteness also makes the inequality automatic for an infinite-risk predictor.

If $N = 0$, only the empty multiindex occurs. If $L = 0$, the retained coefficient is again $\kappa_0 = \mathbb{E}_U F_U = 0$. In both cases risk is at least one, and the zero predictor attains one.

A.2 Sign cumulants

For $S \subseteq [D]$ write $\chi_S(\epsilon) = \prod_{j \in S} \epsilon_j$. An ordered list of p coordinates reduces, after canceling even repetitions, to a nonempty character of odd support size at most p. For supports $S_0, \ldots, S_k$ of a target and k label factors, let $K(S_0, \ldots, S_k)$ be their joint cumulant. Its partition formula is

$$K(S_0, \ldots, S_k) = \sum_\pi (-1)^{|\pi|-1} (|\pi| - 1)! \prod_{B \in \pi} \mathbf{1}_{\{\triangle_{j \in B} S_j = \varnothing\}}, \tag{20}$$

where π ranges over partitions of $\{0, \ldots, k\}$ and $\triangle$ denotes symmetric difference. To derive this identity, expand

$$\log \mathbb{E} \exp\left(\sum_{j=0}^k t_j \chi_{S_j} \right)$$

and extract the coefficient using every t_j once. The logarithm groups positions into set partitions with the displayed coefficient. Differentiating first in t_0 instead gives the ratio of the target-weighted moment generating function to the unweighted one. Multiplying coefficient series yields (14). Repeated labels remain distinct positions before collecting multiindices, which gives its binomial factors. Boundedness of the characters justifies these operations near zero.

We will use three consequences. First, $K = 0$ unless $\triangle_{j=0}^k S_j = \varnothing$, since every nonzero term in (20) has empty symmetric difference within each block. Second, $K = 0$ if the intersection graph of $S_0, \ldots, S_k$ is disconnected. In that case its components depend on disjoint independent signs, so the generating function factors and its logarithm has no coefficient involving all the components. Third,

$$|K| \leq \sum_{j=1}^{k+1} j^k \leq (k+1)^{k+1}. \tag{21}$$

Indeed, $j!$ times the number of partitions into j blocks counts onto maps to j labels and is at most j^{k+1}. Dividing by j bounds the sum of absolute partition coefficients. Each S_j has odd size, so empty total symmetric difference requires $k+1$ to be even. Thus every even-order coefficient vanishes, including order zero as also proved by target symmetry.

A.3 Normal-ordered tensors

For an ordered tuple $j = (j_1, \ldots, j_p)$ define

$$: z_{j_1} \cdots z_{j_p} := \partial_{t_{j_1}} \cdots \partial_{t_{j_p}} \exp(t^\mathsf{T} z - \|t\|^2/2)\Big|_{t=0}.$$

Repeated derivatives are allowed. Differentiating along a unit vector u gives

$$h_p(u^\mathsf{T} z) = \frac{1}{\sqrt{p!}} \sum_{j \in [D]^p} u_{j_1} \cdots u_{j_p} : z_{j_1} \cdots z_{j_p} : . \tag{22}$$

For a symmetric order-p tensor B, define $J_p(B)$ by replacing the coordinate product in this sum by B_j. This map is a Frobenius isometry onto $\mathcal{H}_p$. The Gaussian average of two normal-ordered p-tuples has $p!$ matchings from one tuple to the other. Symmetry makes all tensor contractions equal, and the factor $1/p!$ cancels their number. The range contains every product-Hermite basis function: symmetrize a tensor supported on the corresponding repeated coordinate tuple and multiply by its normalization constant. In particular $F_U = J_p(U^{\otimes p})$.

For a product of more factors, the Gaussian average of their generating functions is

$$\exp\left(\sum_{v<w} t_v^{\mathsf{T}} t_w\right).$$

Coefficient comparison gives pairings of all legs with no pair inside a single factor. Distinct design rows are independent, so every Gaussian pair stays within its row. These finite polynomial identities include repeated coordinates and repeated factors.

Fix a multiindex α of odd order $k \geq 1$ with support size r. Its positive row multiplicities $m_1, \ldots, m_r$ sum to k. Expand its tensor cumulant using (22) and multilinearity, and square the Frobenius norm. There are two replicas, each with k label factors of p legs and one target tensor of p legs. The Frobenius product identifies corresponding target coordinates in order; after the isometry there is no additional target matching to count. Each label factor contributes $(p!)^{-1/2} D^{-p/2} s^{-1}$ and each target tensor contributes $D^{-p/2}$. Squaring therefore gives the exact scalar prefactor

$$(p!)^{-k} D^{-p(k+1)} s^{-2k}. \tag{23}$$

The $2pk$ label legs pair within rows. Assign one initial coordinate variable to every Gaussian pair and one to each of the p target bridges. There are exactly $e = p(k+1)$ initial variables. Their actual values in $[D]$ induce a partition into q equality classes. Each replica contributes the cumulant of its reduced characters. Terms with a zero cumulant in either replica can be removed.

A.4 The coordinate-identification graph

For one remaining term form a bipartite multigraph B. Its left vertices are the r row groups and one target group, and its right vertices are the q coordinate classes. Each initial variable gives one edge from its row or target group to its coordinate class. Parallel edges remain distinct. Thus B has e edges and $r + 1 + q$ vertices. Write $\ell = e - q$ for the coordinate-identification loss.

Lemma A.2. *Every term with nonzero cumulants in both replicas satisfies*

$$\ell \geq \max\left\{r, (p-1)r - \frac{p-2}{2}k + \frac{p}{2}\right\}.$$

Proof. The graph B is connected. The nonzero cumulant in one replica has a connected character intersection graph. Every shared coordinate gives a path through its coordinate vertex in B. Identifying occurrences with the same row preserves connections among all row groups and the target. A coordinate class absent from the reduced characters, because of an even repetition, still has an edge to one of these groups and therefore joins the same component. The cycle rank is consequently

$$\beta(B) = e - (r + 1 + q) + 1 = \ell - r \geq 0. \tag{24}$$

Mark the Gaussian pairs crossing replicas and all p target bridges. Let M be the marked subgraph with isolated vertices removed. Every coordinate vertex of M has even positive degree. Its coordinate count in the left replica is congruent modulo two to its marked degree: a left-internal pair contributes two, a right-internal pair zero, and a cross pair or target bridge one. Empty total symmetric difference in that replica forces evenness. Such vertices therefore have degree at least two.

The target has marked degree p. Let r_1 count singleton rows, r_o count other odd-multiplicity rows, and r_e count even rows. A singleton row has marked degree p, because its only factor on each side cannot pair inside itself. In a row of multiplicity m, if c pairs cross replicas, then the remaining $pm - c$ legs on each side pair internally. Since p is odd, c has the parity of m. Thus every other odd row has marked degree at least one, and an even row is isolated or has degree at least two.

The target ensures that the number c_M of components of M is at least one. Its degree-sum identity gives

$$\beta(M) = c_M + \frac{1}{2}\sum_{v\in M}(\deg_M(v) - 2)$$
$$\geq 1 + \frac{(p-2)(r_1+1) - r_o}{2} = \frac{(p-2)r_1 - r_o + p}{2}. \tag{25}$$

Coordinate vertices and nonisolated even-row vertices contribute nonnegatively. Each other odd row contributes at least -1, while the target and each singleton row contribute $p-2$. The inequality also holds if its right side is negative.

Cycle rank is monotone under taking subgraphs, including multigraphs: cycle vectors in a subgraph remain independent in the full cycle space. Equivalently, extend a spanning forest of the subgraph. Hence $\beta(B) \geq \beta(M)$. Since

$$k \geq r_1 + 3r_o + 2r_e = 2r - r_1 + r_o,$$
$$(p-2)r_1 - r_o = (p-2)(r_1 - r_o) + (p-3)r_o \geq (p-2)(2r-k),$$

where the last inequality uses $p \geq 3$, equations (24) and (25) prove both required lower bounds on ℓ. □

The argument uses necessary cumulant conditions. It does not require an arbitrary connected Wick diagram to have a nonzero cumulant. The nonnegative term $(p-3)r_o$ above explains the general-degree substitution when $p > 3$.

A.5 Counting the terms

Completion of the proof of Theorem A.1. The certificate and its finiteness were proved in Section A.1, and parity was proved in Section A.2. It remains to bound an odd order $k \geq 1$.

There are at most $(2pk)^{pk}$ pairings of the $2pk$ label legs even before imposing the row and normal-order restrictions. The number of partitions of the $e = p(k+1) \leq 2pk$ initial variables is at most $e^e \leq (2pk)^{2pk}$. Indeed, mapping every variable to its block's least-indexed member injects the partitions into e-valued maps. For a partition with q blocks, the number of distinct-coordinate assignments is $(D)_q \leq D^q$, and is zero when $q > D$. By (21), the product of its two sign cumulants has absolute value at most

$$(k+1)^{2(k+1)} \leq (2k)^{4k} \leq (2pk)^{4k}.$$

Taking absolute values termwise, using (23) and Lemma A.2, and dropping $(p!)^{-k} \leq 1$, gives for a fixed α of support r

$$\mathbb{E}_A\|\kappa_\alpha(A)\|^2 \leq (2pk)^{(3p+4)k} s^{-2k} D^{-\delta_p(k,r)}. \tag{26}$$

Remaining sign cancellations can only decrease this upper bound.

There are

$$\binom{N}{r}\binom{k-1}{r-1} \leq N^r 2^{k-1}$$

multiindices of total order k and support r. Drop $1/\alpha! \leq 1$ and absorb $2^{k-1} \leq (2pk)^k$ into (26). Summing over $1 \leq r \leq \min\{N, k\}$ proves (17), with $c_p = 3p+5$.

All expansions and sums are finite. After design averaging only finite Gaussian polynomial moments occur. This proves the bound for repeated rows and coordinates, small D, and $k > D$. When $N = 0$ the positive-order sums are empty, as required. □

A.6 The first-order coefficient

The first-order energy provides an exact check on the normalization. For independent sign-prior teachers U, W, equations (18) and the same covariance identity for one design row give

$$E_1 = \frac{N}{s^2}\mathbb{E}(U^\mathsf{T}W)^{2p} = \frac{N}{s^2 D^{2p}} \sum_{n_1+\cdots+n_D=p} \frac{(2p)!}{\prod_{j=1}^D (2n_j)!}. \tag{27}$$

The last sum ranges over nonnegative integers. It expands a power of a sum of D independent fair signs; only even multiplicities survive. Every coordinate tuple with even counts admits a pairing of equal coordinates. Summing over the $(2p-1)!!$ pairings and D^p coordinate assignments therefore gives

$$E_1 \le \frac{(2p-1)!!N}{s^2 D^p}.$$

For fixed p, tuples with p distinct paired coordinates contribute

$$\binom{D}{p}\frac{(2p)!}{2^p} = (2p-1)!!D^p + O_p(D^{p-1}),$$

and tuples using fewer coordinates number $O_p(D^{p-1})$. Thus the coefficient multiplying N/s^2 in (27) is $(2p-1)!!D^{-p} + O_p(D^{-p-1})$. At $D=1$ it is exactly one. The exponent $\delta_p(1,1) = p$ agrees with these identities.

A.7 Polynomially growing label degree

Corollary A.3. *Fix an odd integer $p \ge 3$ and real numbers $a, \eta > 0$. Put*

$$b = \min\{1+4a, p/2+2a\}, \qquad \tau = \min\{2a, \eta/2\}, \qquad c_p = 3p+5,$$
$$D_0 = \left\lceil \max\left\{2, (4p)^{4c_p/\tau}, 2^{2/\tau}\right\}\right\rceil.$$

For every integer $D \ge D_0$, let $s = D^a$. If the integer sample count satisfies $0 \le N \le D^{b-\eta}$, every predictor (13) of label degree at most

$$L = \left\lfloor D^{\tau/(4c_p)} \right\rfloor$$

has expected prediction risk at least $1 - 2D^{-\tau/2}$.

Proof. First suppose $N \ge 1$. For $1 \le r \le k$, set $q = p-2 \ge 1$ and $c = b-\eta-1$. The exponent of D in each summand of (17) is at most

$$e = (b-\eta)r - 2ak - \delta_p(k,r).$$

If $c < 0$, the first branch of δ_p gives $e \le cr - 2ak \le -\tau k$.

Suppose $0 \le c \le q$. A maximum dominates the convex combination with weight c/q on its second branch. Hence

$$\begin{aligned}\delta_p(k,r) &\ge (1-c/q)r + (c/q)\left((q+1)r - qk/2 + p/2\right)\\ &= (1+c)r - ck/2 + cp/(2q),\\ e &\le -(2a - c/2)k - cp/(2q) \le -\eta k/2 \le -\tau k.\end{aligned}$$

The penultimate inequality uses $c \le 4a - \eta$, which follows from $b \le 1+4a$, and $c \ge 0$. It includes both endpoints $c = 0$ and $c = q$.

Finally suppose $c > q$. Then $a > q/4$, since otherwise $b \le 1+4a \le 1+q$ would give $c < q$. Thus $b = p/2 + 2a$. Set $t = 2a - q/2$. The inequality $c = q + t - \eta > q$ gives $t > \eta$. Using the second branch of δ_p yields

$$e \le (t-\eta)r - tk - p/2 = -t(k-r) - \eta r - p/2 \le -\eta k - p/2 \le -\tau k,$$

where $k - r \ge 0$. These three cases exhaust the possibilities. Every summand is therefore bounded by $D^{-\tau k}$.

There are at most k possible support sizes, so

$$E_k \le k(2pk)^{c_p k} D^{-\tau k} \le (4pk)^{c_p k} D^{-\tau k}.$$

For $k \le L \le D^{\tau/(4c_p)}$, the definition of D_0 gives

$$(4pk)^{c_p} \le (4p)^{c_p} D^{\tau/4} \le D^{\tau/2}.$$

Thus $E_k \le D^{-\tau k/2}$ for $1 \le k \le L$; the vanishing even orders satisfy the same bound. Since $z = D^{-\tau/2} \le 1/2$,

$$\sum_{k=0}^{L} E_k \le \sum_{k\ge 1} z^k = \frac{z}{1-z} \le 2D^{-\tau/2}.$$

Apply (16).

If $N = 0$, risk is at least one and the result follows directly. This includes the case $\eta > b$, when the sample restriction forces $N = 0$. At $\eta = b$ it permits $N = 1$ as well, already covered. At $a = (p-2)/4$ both expressions for b agree and $c = q - \eta < q$, so one of the first two cases applies. All floors only reduce the allowed degree or sample count. The hypotheses require $a, \eta > 0$; setting $\tau = 0$ is not part of the conclusion. □

For fixed p, the sample-exponent envelope in Corollary A.3 is $\min\{Ds^4, D^{p/2}s^2\}$, with crossover $a = (p-2)/4$. The conclusion is below this envelope by a fixed polynomial margin. Although Theorem A.1 is finite and uniform in p, the corollary fixes p and does not assert a uniform growing-degree learning law.

The finite teacher prior implies that, for each learner and each allowed parameter tuple, a fixed unit teacher has at least the displayed expected risk. It is chosen before the observations and learner randomness. The conclusion concerns the label-polynomial class with arbitrary design and fresh-input coefficients. It supplies neither a common hard teacher for all learners nor a fixed-confidence or unrestricted computational lower bound.

B Polynomial prediction algorithms

We construct two predictors with polynomial dependence on the training labels. Throughout this section, $p \ge 3$ is a fixed odd integer and $h_j = \mathrm{He}_j / \sqrt{j!}$, where

$$e^{tz - t^2/2} = \sum_{j\ge 0} \mathrm{He}_j(z) \frac{t^j}{j!}.$$

For $Z \sim N(0,1)$, define

$$C_p = \max_{0\le j\le p} \mathbb{E}[h_p(Z)^2 h_j(Z)^2], \tag{28}$$

$$G_p = 1 + \frac{1}{4p^2} \max_{w\in\{1, z^2, (z^2-1)^2\}} \mathbb{E}[(h_p(Z)^4 + 2h_p(Z)^2 + 3)w(Z)]. \tag{29}$$

The maximum in (29) is over the three displayed polynomials. Both constants depend only on p, and $C_p, G_p \ge 1$. All their entries are finite Gaussian polynomial moments, computable from $\mathbb{E}Z^{2j} = (2j-1)\mathbb{E}Z^{2j-2}$ and $\mathbb{E}Z^{2j+1} = 0$. The even-moment recurrence follows by Gaussian integration by parts.

Theorem B.1. *Fix an integer $D \ge 2$, $s \ge 1$, and $0 < \epsilon \le 1$. The parameters p, D, s, ϵ are supplied to the learner. For an unknown unit vector $u \in \mathbb{R}^D$, the learner observes independent pairs (X_i, Y_i) generated by*

$$X_i \sim N(0, I_D), \qquad G_i \sim N(0,1), \qquad Y_i = h_p(u^\top X_i) + sG_i,$$

with all X_i, G_i independent. Put

$$r = \epsilon/p, \qquad t = \lceil \log_2(100D/r) \rceil, \qquad \Delta = C_p + s^2, \tag{30}$$

$$m = \lceil 2^{p+1}\Delta/\epsilon \rceil, \qquad n_Q = \lceil 100 G_p D s^4 t/r \rceil, \tag{31}$$

$$n_T = \left\lceil 400\Delta D t/r + 20\Delta D^{p/2}\sqrt{t/r} \right\rceil, \tag{32}$$

$$N_Q = ptn_Q + m, \qquad N_T = 2ptn_T + m. \tag{33}$$

There are predictors using N_Q and N_T samples, respectively, each of total degree at most $2pt+1$ in its training labels, such that

$$\sup_{\|u\|=1} \mathbb{E}_u[(\widehat{f}(x) - h_p(u^\top x))^2] \le \epsilon, \qquad x \sim N(0, I_D)$$

for an independent fresh x. Their coefficients are polynomials in the observed design and x. Choosing the smaller deterministic sample count gives

$$N = O_p\Big(\min\Big\{Ds^4\epsilon^{-1}\log^2(D/\epsilon),\quad s^2\left[D\epsilon^{-1}\log^2(D/\epsilon) + D^{p/2}\epsilon^{-1/2}\log^{3/2}(D/\epsilon)\right]\Big\}\Big). \quad (34)$$

For fixed p, both constructions have polynomial real-arithmetic work and storage. Extra observations may be ignored.

The proof uses independent matrix products and a linear tensor output. It bounds the full expected loss without label-dependent normalization or spectral selection.

B.1 Hermite tensors and covariance

Let $R_p(X)$ be the ordered p-tensor whose entry at $(i_1, \dots, i_p)$ is the corresponding mixed derivative at zero of $\exp(v^\top X - \|v\|^2/2)$. Tensor norms and inner products are Frobenius ones, with every ordered index tuple included. Let Π be the orthogonal projection onto symmetric tensors, given by averaging the $p!$ permutations of the tensor factors. Define

$$J_p(F)(X) = \frac{\langle F, R_p(X)\rangle}{\sqrt{p!}}. \quad (35)$$

Lemma B.2. *For deterministic tensors F, H and any unit vector u,*

$$\mathbb{E}[J_p(F)(X)J_p(H)(X)] = \langle \Pi F, \Pi H\rangle, \quad (36)$$

$$J_p(u^{\otimes p})(X) = h_p(u^\top X), \qquad \frac{\mathbb{E}[Y R_p(X)]}{\sqrt{p!}} = u^{\otimes p}. \quad (37)$$

Every function f in homogeneous Gaussian chaos of order p satisfies

$$\mathbb{E}[h_p(u^\top X)^2 f(X)^2] \le C_p \mathbb{E}[f(X)^2]. \quad (38)$$

Consequently, for an integer $n \ge 1$, the estimator

$$\widehat{C}_n = \frac{1}{\sqrt{p!}\,n}\sum_{i=1}^n Y_i R_p(X_i) \quad (39)$$

satisfies

$$\mathbb{E}\widehat{C}_n = u^{\otimes p}, \qquad \mathrm{Cov}(\widehat{C}_n) \preceq \frac{C_p + s^2}{n}\Pi \preceq \frac{C_p + s^2}{n} I. \quad (40)$$

Proof. The Gaussian exponential identity

$$\mathbb{E}e^{v^\top X - \|v\|^2/2} e^{w^\top X - \|w\|^2/2} = e^{v^\top w}$$

shows, by taking p derivatives in each argument, that the expectation of two entries of R_p is the sum of their $p!$ cross matchings. Repeated indices retain their multiplicities. Dividing by $p!$ and summing against F, H proves (36). Differentiating along the unit vector u gives the generating function $\exp(tu^\top X - t^2/2)$, and proves the first identity in (37). The same cross matchings, together with the independent zero-mean noise, prove its second identity.

Rotate u to the first coordinate and write $X = (Z, W)$. Products of the one-dimensional normalized Hermite polynomials form an orthonormal basis of Gaussian chaos. Collecting the terms by their degree in Z gives the decomposition

$$f(Z, W) = \sum_{j=0}^p h_j(Z) f_{p-j}(W),$$

where f_{p-j} belongs to chaos order $p - j$ in W. Distinct orders are orthogonal. Integrating W therefore removes the cross terms both in f^2 and in $h_p(Z)^2 f^2$, so

$$\begin{aligned}\mathbb{E}[h_p(Z)^2 f(Z, W)^2] &= \sum_{j=0}^p \mathbb{E}[h_p(Z)^2 h_j(Z)^2]\mathbb{E}[f_{p-j}(W)^2] \\ &\le C_p \mathbb{E}[f(Z, W)^2].\end{aligned}$$

This proves (38). For $D = 1$, the positive-order chaos spaces in W are zero and only the $j = p$ term remains.

Testing the covariance of (39) on any deterministic tensor F gives

$$\mathrm{Var}(\langle \widehat{C}_n, F\rangle) = \frac{1}{n}\,\mathrm{Var}(Y J_p(F)(X)) \le \frac{C_p + s^2}{n}\|\Pi F\|^2.$$

Here (38) applies to $J_p(F)$, and the independent noise adds $s^2\mathbb{E}J_p(F)^2$. Averaging independent samples and subtracting the squared mean can only reduce the variance. This proves the operator inequality on the whole ordered tensor space. All moments used are finite Gaussian polynomial moments. □

B.2 Independent matrix products

The next lemma controls matrix products using three scalar second moments.

Lemma B.3. *Let $D \ge 2$ be an integer, fix a unit vector u, and put $P = uu^\top$, $Q = I - P$, and $d = D - 1$. Suppose that a symmetric random matrix M has mean P, finite second moments, and a law invariant under conjugation by every orthogonal matrix fixing u. Define*

$$A = \mathbb{E}(u^\top M u)^2, \qquad B = \mathbb{E}\|QMu\|^2, \qquad C = \frac{\mathbb{E}\|QMQ\|_F^2}{d}.$$

Suppose $\gamma > 0$, $0 < r \le 1$, and an integer $t \ge \lceil \log_2(100D/r)\rceil$ satisfy

$$A \le 1 + \gamma, \qquad B \le d\gamma, \qquad C \le D\gamma, \qquad h := D\gamma \le \frac{r}{100t}. \tag{41}$$

For independent copies $M_1, \dots, M_t$, put $\mathcal{B}_t = M_t \cdots M_1$. Then

$$\mathbb{E}\mathcal{B}_t = P, \qquad K := \mathbb{E}\mathcal{B}_t^\top \mathcal{B}_t = a_t P + b_t Q, \qquad a_t \le e^{r/100}, \qquad \mathrm{tr}\, K < 2. \tag{42}$$

Proof. Independence gives $\mathbb{E}\mathcal{B}_t = P^t = P$. Write $S_j = \mathbb{E}\mathcal{B}_j\mathcal{B}_j^\top$. The invariance assumption gives $S_j = a_j P + b_j Q$, starting with $a_0 = b_0 = 1$. Indeed reflections in perpendicular coordinates remove mixed entries, and perpendicular-coordinate permutations make their diagonal entries equal. Independence of the newest factor, followed by the u quadratic form and the perpendicular trace, gives exactly

$$a_j = Aa_{j-1} + Bb_{j-1}, \qquad b_j = (B/d)a_{j-1} + Cb_{j-1}. \tag{43}$$

The first row sum is at most $1 + \gamma + d\gamma = 1 + h$. The second is at most $\gamma + h \le 1 + h$, since $\gamma \le h \le 1/100$. Hence

$$\max\{a_j, b_j\} \le (1 + h)^j \le e^{r/100}.$$

Keeping the two terms in the second recursion gives

$$b_t \le h^t + \frac{\gamma e^{r/100}}{1 - h}.$$

Reversing the order of the independent identically distributed symmetric factors shows that $\mathcal{B}_t$ and $\mathcal{B}_t^\top$ have the same law. Thus $K = S_t$, and

$$\mathrm{tr}\, K = a_t + db_t \le \frac{e^{r/100}}{1 - h} + Dh^t.$$

We used $d\gamma \le D\gamma = h$ in the last inequality. Since $h \le 1/2$ and $D2^{-t} \le r/100$, we have $Dh^t \le 1/100$. Also $e^z \le (1 - z)^{-1}$ for $0 \le z < 1$, by comparison of the two power series. Therefore

$$\mathrm{tr}\, K \le (100/99)^2 + 1/100 < 2.$$

This proves (42) using only second moments. □

B.3 Quadratic-label blocks

For the first construction, use n_Q fresh samples to form

$$M_Q = \frac{1}{2pn_Q}\sum_{i=1}^{n_Q}(Y_i^2 - (1+s^2))(X_iX_i^\top - I_D). \tag{44}$$

Lemma B.4. *The matrix M_Q has mean P and satisfies the assumptions of Lemma B.3 with $\gamma = G_p s^4/n_Q$ and the parameters in (30)–(31).*

Proof. The normalized Hermite recurrence is

$$zh_p(z) = \sqrt{p+1}\,h_{p+1}(z) + \sqrt{p}\,h_{p-1}(z).$$

It follows by differentiating the defining exponential series and comparing coefficients. Orthogonality gives $\mathbb{E}Z^2h_p(Z)^2 = 2p+1$, and hence

$$\mathbb{E}[h_p(Z)^2(Z^2-1)] = 2p. \tag{45}$$

Use coordinates $X = (Z, V)$ with first direction u, and set $H = h_p(Z)$ and $W_0 = Y^2 - (1+s^2)$. The vector V is independent of Z, G, and $\mathbb{E}[W_0 \mid Z] = H^2 - 1$. Equation (45) gives the u entry of $\mathbb{E}M_Q$. All off-diagonal means vanish by $\mathbb{E}V = 0$, and the perpendicular mean vanishes by $\mathbb{E}(VV^\top - I_d) = 0$. Thus $\mathbb{E}M_Q = P$. Symmetry and invariance under orthogonal transformations fixing u follow directly from (44).

The conditional second moment is

$$\mathbb{E}_G W_0^2 = (H^2-1)^2 + 4s^2H^2 + 2s^4 \le s^4(H^4 + 2H^2 + 3). \tag{46}$$

For the inequality, multiply the nonnegative first term by s^4 and use $s^2 \le s^4$, since $s \ge 1$. The three scalar moments of the sample average are

$$A = 1 + \frac{\mathbb{E}[W_0^2(Z^2-1)^2] - (2p)^2}{4p^2n_Q}, \tag{47}$$

$$B = \frac{d\mathbb{E}[W_0^2Z^2]}{4p^2n_Q}, \qquad C = \frac{D\mathbb{E}W_0^2}{4p^2n_Q}. \tag{48}$$

The first formula subtracts the squared mean from the variance of the u entry. For the second, cross-sample terms vanish and $\mathbb{E}\|V\|^2 = d$. For the third, the lower block has mean zero and

$$\mathbb{E}\|VV^\top - I_d\|_F^2 = \mathbb{E}\|V\|^4 - 2\mathbb{E}\|V\|^2 + d = d(d+1) = dD.$$

Here $\mathbb{E}\|V\|^4 = d(d+2)$ follows by independence and the one-dimensional fourth moment. Independence of V from W_0 justifies each factorization. Equations (29) and (46)–(48) give $A \le 1+\gamma$, $B \le d\gamma$, and $C \le D\gamma$. The choice of n_Q ensures $D\gamma \le r/(100t)$. □

B.4 Independent tensor Gram blocks

For the second construction, take independent copies C_1, C_2 of $\widehat{C}_{n_T}$, using $2n_T$ fresh samples. Flatten each ordered tensor to a D by J matrix, where $J = D^{p-1}$, with its first index as the row. Its mean is $uv^\top$ for $v = u^{\otimes(p-1)}$, and $\|v\| = 1$. Write

$$C_j = uv^\top + E_j, \qquad \kappa = \Delta/n_T, \qquad M_{\text{raw}} = C_1C_2^\top, \qquad M_T = \tfrac{1}{2}(M_{\text{raw}} + M_{\text{raw}}^\top). \tag{49}$$

Lemma B.5. *The matrix M_T has mean P and satisfies the assumptions of Lemma B.3 with $\gamma = 2\kappa + \kappa^2 J$ and the parameters in (30) and (32).*

Proof. Flattening preserves the Frobenius inner product. The errors E_1, E_2 are centered and independent, with covariance at most κI by (40). Consequently $\mathbb{E}M_T = P$. We prove the moment estimates for any column dimension J. For every deterministic $z \in \mathbb{R}^J$, summing the scalar covariance bounds over an orthonormal basis of $u^\perp$ gives

$$\mathbb{E}\|QE_jz\|^2 \le \kappa d\|z\|^2, \qquad \mathbb{E}\|E_j^\top Q\|_F^2 \le \kappa dJ, \qquad \mathrm{Cov}(E_j^\top u) \preceq \kappa I_J. \tag{50}$$

The second inequality also sums over the J column vectors; the third tests $\mathbb{E}(u^\top E_j z)^2 \le \kappa\|z\|^2$. Expanding gives

$$M_{\text{raw}} = P + u(E_2 v)^\top + (E_1 v)u^\top + E_1 E_2^\top.$$

In $u^\top M_{\text{raw}} u$, the two centered linear terms have variance at most κ each. The bilinear term is $(E_1^\top u)^\top (E_2^\top u)$; its second moment is the trace of the product of their covariance matrices, at most $\kappa^2 J$. All cross terms have mean zero by conditioning on one centered independent error. Scalar quadratic forms are unchanged by symmetrization, so

$$A - 1 \le 2\kappa + \kappa^2 J. \tag{51}$$

Next, $QM_{\text{raw}}u = QE_1 v + QE_1(E_2^\top u)$. The cross term has mean zero by centering of E_2. Conditioning on E_2 and using (50) gives

$$\mathbb{E}\|QM_{\text{raw}}u\|^2 \le \kappa d + \kappa d\mathbb{E}\|E_2^\top u\|^2 \le d(\kappa + \kappa^2 J).$$

Exchanging E_1, E_2 gives the same bound for the transpose. Convexity of squared norm therefore gives

$$B \le d(\kappa + \kappa^2 J). \tag{52}$$

Finally, $QM_{\text{raw}}Q = QE_1E_2^\top Q$. Apply the first inequality of (50) to every column of $E_2^\top Q$, then average, to obtain

$$\mathbb{E}\|QM_{\text{raw}}Q\|_F^2 \le \kappa d\mathbb{E}\|E_2^\top Q\|_F^2 \le \kappa^2 d^2 J.$$

The transpose has the same bound, and convexity yields $C \le \kappa^2 dJ$. Together with (51) and (52), this proves the three moment assumptions for $\gamma = 2\kappa + \kappa^2 J$. Moreover,

$$D\gamma = \frac{2\Delta D}{n_T} + \frac{\Delta^2 D^p}{n_T^2} \le \frac{r}{200t} + \frac{r}{400t} \le \frac{r}{100t}. \tag{53}$$

Each of the two positive summands in (32) gives the corresponding denominator bound, and rounding only strengthens it.

It remains to check invariance of the law. If O is orthogonal with $Ou = u$, replacing every design row by OX preserves the observation law. Substitution in the generating function gives $R_p(OX) = O^{\otimes p} R_p(X)$. After flattening, C_j transforms to $OC_j(O^{\otimes(p-1)})^\top$, and M_T transforms to $OM_TO^\top$. This proves the required invariance. All second moments are finite by the preceding estimates. □

B.5 Tensor output and expected risk

For either choice of block, construct p independent copies $\mathcal{B}^{(1)}, \dots, \mathcal{B}^{(p)}$ of the product in Lemma B.3. All their blocks and an additional final m-sample tensor $\widehat{C}_m$ use disjoint observations. Their common second-moment matrix is $K = aP + bQ$ from (42), and each product has mean P. Define

$$\widehat{T} = (\mathcal{B}^{(1)} \otimes \cdots \otimes \mathcal{B}^{(p)})\widehat{C}_m, \qquad \widehat{f}(x) = J_p(\Pi\widehat{T})(x). \tag{54}$$

Lemma B.6. *The predictor in (54) satisfies*

$$\mathbb{E}[(\widehat{f}(x) - h_p(u^\top x))^2] \le a^p - 1 + \frac{\Delta(\operatorname{tr} K)^p}{m}. \tag{55}$$

Only second moments of the independent matrix products are needed.

Proof. Independence gives $\mathbb{E}\widehat{T} = u^{\otimes p}$. The projection Π fixes this target. By (36) and contraction of orthogonal projection,

$$\mathbb{E}[(\widehat{f}(x) - h_p(u^\top x))^2] \le \mathbb{E}\|\widehat{T} - u^{\otimes p}\|^2 = \mathbb{E}\|\widehat{T}\|^2 - 1.$$

For a deterministic tensor c, the p independent second moments give

$$\mathbb{E}\|(\mathcal{B}^{(1)} \otimes \cdots \otimes \mathcal{B}^{(p)})c\|^2 = \langle c, K^{\otimes p} c\rangle.$$

Expanding the m independent summands of $\widehat{C}_m$ therefore gives

$$\mathbb{E}\|\widehat{T}\|^2 = (1-1/m)a^p + \frac{1}{m}\mathbb{E}[Y^2 V_K(X)], \tag{56}$$

$$V_K(X) = \left\| (K^{1/2})^{\otimes p} \frac{R_p(X)}{\sqrt{p!}} \right\|^2. \tag{57}$$

The off-diagonal sample terms use the mean $u^{\otimes p}$ and $Ku = au$. Diagonalize K. Equation (38) bounds each nonnegative weighted coordinate square in V_K, and independent noise adds s^2 times its expectation. Hence

$$\mathbb{E}[Y^2 V_K] \le \Delta \mathbb{E} V_K.$$

As an operator on ordered tensors, (36) says $\mathbb{E}[R_p(X)R_p(X)^\top]/p! = \Pi$. It follows that

$$\mathbb{E}V_K = \operatorname{tr}(K^{\otimes p}\Pi) \le \operatorname{tr}(K^{\otimes p}) = (\operatorname{tr} K)^p.$$

For the inequality, diagonalize the orthogonal projection Π. The omitted diagonal entries of the positive semidefinite operator $K^{\otimes p}$ are nonnegative. This argument also proves finiteness of every norm and cross term used above. Combining it with (56) and discarding $-a^p/m \le 0$ proves (55). No higher moment of a matrix product has been used. □

Proof of Theorem B.1. Lemmas B.4 and B.5 each give the assumptions of Lemma B.3. Form (54) using the selected block type. Since $r = \epsilon/p$, the product bound gives $a^p \le e^{\epsilon/100}$ and $\operatorname{tr} K < 2$. The choice of m then gives

$$\frac{\Delta(\operatorname{tr} K)^p}{m} \le \epsilon/2.$$

The power-series inequality used earlier gives $e^{\epsilon/100} - 1 \le \epsilon/99$ for $0 < \epsilon \le 1$. Lemma B.6 therefore bounds the expected risk by $(1/99 + 1/2)\epsilon < \epsilon$, uniformly over fixed unit teachers.

The first construction uses pt independent blocks of n_Q samples; the second uses pt pairs of blocks of n_T samples. Both use a final block of m samples, giving exactly (33). Each matrix block has label degree at most two. Each product thus has degree at most $2t$, and the output uses its p independent products once and the final tensor of degree one. Its total label degree is at most $2pt+1$. Both Π and J_p are fixed linear maps. The branch choice depends only on supplied parameters, so it preserves polynomial dependence on the labels and design.

For fixed p, $\Delta \le (C_p+1)s^2$ since $s \ge 1$, $r^{-1} = p/\epsilon$, and $t = O_p(\log(D/\epsilon))$. Substitution in (33), using $\lceil z \rceil \le z+1$, proves (34). The additive block counts and final m are absorbed by either displayed branch for $D \ge 2$. For fixed accuracy, the term $D\log^2 D$ in the tensor branch is at most a constant times $D^{p/2}\log^{3/2} D$: it suffices that $\log D \le D^{p-2}$, which holds up to an absolute constant for $D \ge 2$ and $p \ge 3$. Thus the fixed-accuracy sample envelope is $\min\{Ds^4, D^{p/2}s^2\}$ up to logarithmic factors.

There are D^p ordered Hermite tensor entries. For fixed p, each is a degree-p polynomial computable with a constant depending on p, and tensor sums can be accumulated while streaming observations. Forming a quadratic matrix costs $O(D^2)$ per observation; forming a tensor sum costs $O_p(D^p)$ per observation. A Gram product of two D by D^{p-1} tensors costs $O(D^{p+1})$ operations. Ordinary matrix products cost $O(D^3)$, and applying a matrix along one index of an order-p tensor costs $O(D^{p+1})$. Multiplying these costs by the stated block counts gives polynomial real-arithmetic work for fixed p. Streaming block sums and products uses $O_p(D^p + D^2)$ storage. Evaluation on a fresh input costs $O_p(D^p)$. Constants such as $\sqrt{p!}$ are real-arithmetic primitives; these bounds do not assert finite-bit complexity. □

B.6 Dimension one and scope

Proposition B.7. *For $D = 1$, fixed odd $p \ge 3$, $s \ge 1$ and $0 < \epsilon \le 1$, there is a predictor of label degree one and expected risk at most ϵ using*

$$m_1 = \left\lceil \frac{\mathbb{E}h_p(Z)^4 - 1 + s^2}{\epsilon} \right\rceil$$

samples, uniformly over the two unit teachers.

Proof. The teacher u is a sign and $h_p(uX) = uh_p(X)$ because p is odd. Set

$$\widehat{c} = \frac{1}{m_1}\sum_{i=1}^{m_1} Y_i h_p(X_i), \qquad \widehat{f}(x) = \widehat{c}\, h_p(x).$$

The coefficient has mean u and variance $(\mathbb{E}h_p(Z)^4 - 1 + s^2)/m_1$. The latter is the prediction risk, since $\mathbb{E}h_p(x)^2 = 1$. It is at most ϵ. The numerator is positive because $s \geq 1$ and $\mathbb{E}h_p^4 \geq (\mathbb{E}h_p^2)^2 = 1$, so $m_1 \geq 1$. □

The constructions are uniform over fixed unit teachers and therefore apply to any teacher prior. Their outputs are generally improper order-p chaos functions and need not be Hermite ridges in unit directions. The stated costs fix p; the constants may grow with p, and no efficiency assertion for a growing Hermite degree follows.

C Total degree in the design, labels, and query

We count the joint degree of the upper construction; its risk bound is unchanged.

Proof of Corollary 2.2. Grouping a joint polynomial by its label monomials expresses it in the form (3), with polynomial design and query coefficients. Thus $\mathcal{Q}_J \subseteq \mathcal{P}_J$ and $\mathcal{R}_{\text{joint}}(D, N, J, p, s) \geq \mathcal{R}(D, N, J, p, s)$. Theorem 2.1 therefore gives the asserted lower.

For the upper, apply the construction of Theorem B.1 with $\epsilon = D^{-\eta/4}$. Its block count is $t = \lceil \log_2(100pD^{1+\eta/4}) \rceil$. We count total degree with every scalar design coordinate, label, and query coordinate assigned degree one. In the defining expansion

$$\exp(v^\top X - \|v\|^2/2) = \exp(v^\top X)\exp(-\|v\|^2/2),$$

a coefficient of total v-degree p contains at most p factors of X. Every entry of $R_p(X)$ has degree at most p. Consequently $\widehat{C}_n$ in (39) has joint training degree at most $p + 1$.

Each entry of M_Q in (44) has degree at most four. Each entry of M_T in (49) has degree at most $2(p + 1)$, since it is a sum of products of two entries of independent copies of $\widehat{C}_{n_T}$. Flattening and transposition only reorder coordinates. Matrix multiplication sums products, so degrees add. Each product $\mathcal{B}^{(j)}$ contains t matrix blocks, and the output tensor in (54) uses p such products and one final empirical tensor. Its training degree is therefore at most $4pt + p + 1$ for quadratic blocks and $2p(p + 1)t + p + 1$ for tensor blocks.

The map Π averages fixed coordinate permutations and does not increase degree. The final evaluation $J_p(F)(x) = \langle F, R_p(x)\rangle/\sqrt{p!}$ adds at most p query degrees. The two predictors have full joint degree at most

$$4pt + 2p + 1 \qquad \text{and} \qquad 2p(p + 1)t + 2p + 1,$$

respectively. Since $p \geq 3$, both are at most J_+. Every division in the construction is by a nonzero constant depending only on supplied parameters and deterministic sample counts. The branch choice and sample blocks also depend only on these parameters. There is no data-dependent normalization or spectral selection. Hence both predictors belong to $\mathcal{Q}_{J_+}$.

The sample-count calculation in the proof of Theorem 2.1 gives expected risk at most $D^{-\eta/4}$ with at most $\lceil D^{b+\eta} \rceil$ samples for all sufficiently large D. Ignoring extra observations preserves degree and risk. Finally $J_+ = O_{p,\eta}(\log D)$ and L_- grows as a positive power of D. Thus $J_+ \leq L_-$ eventually. Nestedness transfers the lower to every smaller degree and the upper to every larger degree, proving the last assertion. □

D Statistical comparison

Let $p \geq 3$ be odd, $D \geq 2$, and $s \geq 1$. Write $h_j = \mathrm{He}_j/\sqrt{j!}$, let γ_D be standard Gaussian measure on $\mathbb{R}^D$, and put $f_u(x) = h_p(u \cdot x)$ for unit u. The observations are independent pairs

$$X_i \sim \gamma_D, \qquad Y_i = f_u(X_i) + sG_i, \qquad G_i \sim \mathcal{N}(0, 1),$$

with the Gaussian noises independent of the design. The learner knows p, D, s, receives no teacher-dependent side information, and may use arbitrary computation and independent randomization. For a jointly measurable output $\widehat{f}$, define

$$R_u(\widehat{f}) = \int (\widehat{f}(x) - f_u(x))^2 \, d\gamma_D(x), \qquad \mathrm{Risk}_u(\widehat{f}) = \mathbb{E}R_u(\widehat{f}),$$

allowing infinite risk. The expectation includes training data and learner randomness; the fresh input in the integral is independent. Let $N_*(D, s, p)$ be the least nonnegative integer sample count for which a learner has $\sup_{\|u\|=1} \mathrm{Risk}_u \le 1/10$.

Theorem D.1. *For every fixed odd $p \ge 3$, there is a finite constant K_p such that, for every $D \ge 2$ and $s \ge 1$,*

$$\frac{Ds^2}{16p} < N_*(D, s, p) \le K_p D s^2.$$

Thus the sample requirement at expected risk $1/10$ is $\Theta_p(Ds^2)$. The upper uses proper predictors. The lower applies to every jointly measurable predictor, including improper and randomized outputs, without any restriction on label transformations or computation.

D.1 Hermite covariance and score variance

For a standard Gaussian Z, define the finite constant

$$C_p = \max_{0 \le j \le p} \mathbb{E}[h_p(Z)^2 h_j(Z)^2] \ge 1, \qquad \Delta = C_p + s^2.$$

It depends only on p and is an explicit polynomial Gaussian moment.

Lemma D.2. *For unit u, v,*

$$\|f_u\|_2^2 = 1, \qquad \langle f_u, f_v \rangle = (u \cdot v)^p, \tag{58}$$

$$\|f_u - f_v\|_2^2 = 2[1 - (u \cdot v)^p] \le 4. \tag{59}$$

If f belongs to degree-p Gaussian Hermite chaos, then

$$\mathbb{E}[h_p(u \cdot X)^2 f(X)^2] \le C_p \|f\|_2^2. \tag{60}$$

Consequently the score $Z_v = Y f_v(X)$ satisfies

$$\mathbb{E}Z_v = (u \cdot v)^p =: \mu_v, \qquad \mathrm{Var}(Z_v) \le \Delta. \tag{61}$$

Proof. The defining generating function is

$$\exp(tz - t^2/2) = \sum_{j \ge 0} \mathrm{He}_j(z) \frac{t^j}{j!}.$$

For jointly standard Gaussian Z, W with correlation ρ,

$$\mathbb{E} \exp(tZ - t^2/2) \exp(vW - v^2/2) = \exp(\rho t v).$$

Comparing coefficients gives $\mathbb{E}[h_p(Z)h_p(W)] = \rho^p$, including $\rho = \pm 1$. This proves (58); (59) follows because $|u \cdot v| \le 1$. The coefficient comparisons use finite derivatives of Gaussian generating functions with finite exponential moments.

Products of the coordinate generating functions give

$$f_v(X) = \sum_{|\nu| = p} \sqrt{\frac{p!}{\nu!}} \, v^\nu \prod_j h_{\nu_j}(X_j).$$

These product Hermites form the orthonormal basis of degree-p Gaussian chaos. Rotate u to the first coordinate and write $X = (Z, W)$. Collecting basis terms by their degree in Z gives, for any f in this chaos,

$$f(Z, W) = \sum_{j=0}^{p} h_j(Z) f_{p-j}(W),$$

where f_{p-j} belongs to degree-$(p-j)$ chaos in the other coordinates. Distinct degrees in W are orthogonal. Averaging W therefore removes cross terms in both f^2 and $h_p(Z)^2 f^2$, and yields

$$\mathbb{E}[h_p(Z)^2 f(Z,W)^2] = \sum_{j=0}^{p} \mathbb{E}[h_p(Z)^2 h_j(Z)^2] \, \|f_{p-j}\|_2^2$$
$$\leq C_p \sum_{j=0}^{p} \|f_{p-j}\|_2^2 = C_p \|f\|_2^2.$$

This proves (60). Applying it to the unit-norm function f_v, and averaging the independent centered noise, gives

$$\mathrm{Var}(Z_v) \leq \mathbb{E}Z_v^2 = \mathbb{E}[h_p(u \cdot X)^2 f_v(X)^2] + s^2 \mathbb{E} f_v(X)^2 \leq C_p + s^2.$$

The score mean follows from (58). Finally, C_p is finite because its integrands are polynomials. Gaussian integration by parts gives $\mathbb{E}Z^{2j} = (2j-1)\mathbb{E}Z^{2j-2}$ from $\mathbb{E}1 = 1$, and odd moments vanish. The Gaussian density makes the boundary terms zero, so this recurrence computes every entry of C_p. Its $j = 0$ entry equals one. □

D.2 A PROPER FINITE-NET UPPER

Proposition D.3. *For $0 < \epsilon \leq 1$, put $r = \sqrt{\epsilon/(4p)}$ and fix an r-net $\mathcal{V} = \{v_1, \ldots, v_M\}$ of the unit sphere with $M \leq (1 + 2/r)^D$. Define*

$$b = \left\lceil \frac{4096\Delta}{\epsilon^2} \right\rceil, \qquad q = 2 \left\lceil \frac{1}{2} \log_2 \frac{8M}{\epsilon} \right\rceil + 1, \qquad N_{\mathrm{up}} = bq.$$

There is a deterministic proper learner using N_{up} observations whose expected risk is at most ϵ for every unit teacher. Additional observations may be ignored. At $\epsilon = 1/10$, $N_{\mathrm{up}} \leq K_p D s^2$ for a finite constant depending only on p.

Proof. First construct a maximal finite subset of the sphere with pairwise distances strictly greater than r. Starting from one point, add a point more than r from all previous ones whenever one exists. The open Euclidean balls of radius $r/2$ centered at the selected points are disjoint and lie in the ball of radius $1 + r/2$. Comparing their D-dimensional volumes bounds the number of points by $(1 + 2/r)^D$. The procedure must therefore stop after finitely many additions, and at termination it is an r-net. Fix such a net and an ordering before observing any data. This is an existence construction of a finite set; no efficiency or finite-bit claim is needed.

For each unit u, choose $v_0 \in \mathcal{V}$ with $\|u - v_0\| \leq r$. Then $\rho_0 = u \cdot v_0 = 1 - \|u - v_0\|^2/2 \geq 0$, since $r^2 \leq 1$. For $0 \leq \rho \leq 1$,

$$1 - \rho \leq 1 - \rho^p = (1 - \rho) \sum_{j=0}^{p-1} \rho^j \leq p(1 - \rho). \tag{62}$$

Together with (59), this gives

$$\|f_u - f_{v_0}\|_2^2 \leq p\|u - v_0\|^2 \leq \epsilon/4, \qquad \mu_{v_0} \geq 1 - \epsilon/8. \tag{63}$$

Partition the data into q consecutive blocks of size b. For each $v \in \mathcal{V}$, compute the mean of $Y f_v(X)$ in each block and let $\widehat{\mu}_v$ be the median of these q means. The integer q is odd. Select a maximizing index for $\widehat{\mu}_v$, breaking ties by the fixed ordering, and output the associated function $f_{\widehat{v}}$. This is a jointly measurable predictor: all scores, finite order statistics, comparisons and tie choices are measurable.

Let $t = \epsilon/16$. By (61), independence within a block and Chebyshev's inequality, each block mean B_v satisfies

$$\mathbb{P}\{|B_v - \mu_v| > t\} \leq \frac{\Delta}{bt^2} \leq \frac{1}{16}.$$

For a fixed v, the block events are independent. If its median deviates by more than t, at least $(q+1)/2$ blocks deviate by more than t. A union bound over subsets of blocks therefore gives

$$\mathbb{P}\{|\widehat{\mu}_v - \mu_v| > t\} \leq 2^q (1/16)^{(q+1)/2} \leq 2^{-q}.$$

Independence between different candidates is unnecessary. A union bound over $\mathcal{V}$, with $q \geq \log_2(8M/\epsilon)$, gives

$$\mathbb{P}\left\{\max_{v\in\mathcal{V}} |\widehat{\mu}_v - \mu_v| > t\right\} \leq \epsilon/8. \tag{64}$$

On the complementary event, maximality and (63) imply

$$\mu_{\widehat{v}} \geq \widehat{\mu}_{\widehat{v}} - t \geq \widehat{\mu}_{v_0} - t \geq \mu_{v_0} - 2t \geq 1 - \epsilon/4.$$

Its fresh prediction loss is at most $\epsilon/2$ by (59). On every outcome, including failure of (64), the output is still a unit-vector function in the net and its loss is at most four. Thus the full expected risk is at most

$$\epsilon/2 + 4(\epsilon/8) = \epsilon.$$

Every bound is uniform in the unknown unit teacher.

To make the sample constant explicit, set

$$B_p(\epsilon) = (4096/\epsilon^2 + 1)(C_p + 1),$$
$$Q_p(\epsilon) = \log_2(1 + 4\sqrt{p/\epsilon}) + \log_2(8/\epsilon) + 3.$$

Since $s \geq 1$, $\Delta \leq (C_p + 1)s^2$ and $b \leq B_p(\epsilon)s^2$. The inequality $2\lceil z/2\rceil + 1 \leq z + 3$ and the net cardinality give

$$\begin{aligned} q &\leq D\log_2(1 + 4\sqrt{p/\epsilon}) + \log_2(8/\epsilon) + 3 \\ &\leq DQ_p(\epsilon). \end{aligned}$$

Hence $N_{\text{up}} \leq B_p(\epsilon)Q_p(\epsilon)Ds^2$; take $K_p = B_p(1/10)Q_p(1/10)$. This absorbs all ceilings. The net is generally exponential in D, so the construction provides a statistical upper without a polynomial-time guarantee. □

D.3 A lower for improper predictors

Proposition D.4. *If the nonnegative integer N satisfies*

$$N \leq \frac{(D-1)s^2}{8p},$$

then every allowed learner has a fixed unit teacher u for which $\text{Risk}_u(\widehat{f}) \geq 3/16$.

Proof. Put $d = D - 1 \geq 1$. For $\theta \in \{-1, 1\}^d$, define

$$u_\theta = \frac{e_1}{\sqrt{2}} + \frac{1}{\sqrt{2d}}\sum_{j=1}^{d} \theta_j e_{j+1}, \qquad f_\theta = f_{u_\theta}.$$

These are unit vectors. If $H(\theta, \omega)$ denotes Hamming distance, then $u_\theta \cdot u_\omega = 1 - H(\theta, \omega)/d$ belongs to $[0, 1]$. Equations (59) and (62) imply

$$\frac{2H(\theta,\omega)}{d} \leq \|f_\theta - f_\omega\|_2^2 \leq \frac{2pH(\theta,\omega)}{d}. \tag{65}$$

Let P_θ be the law of one labeled observation. These laws have the same input density. Conditional on input x, the two label laws are Gaussian with means $f_\theta(x), f_\omega(x)$ and common positive variance s^2. Their log density ratio is

$$\frac{(Y - f_\omega(x))^2 - (Y - f_\theta(x))^2}{2s^2}.$$

Its expectation under the θ model is $(f_\theta(x) - f_\omega(x))^2/(2s^2)$. Integrating over input and adding the N independent log ratios gives

$$\text{KL}(P_\theta^N \| P_\omega^N) = \frac{N\|f_\theta - f_\omega\|_2^2}{2s^2}. \tag{66}$$

All integrals are finite Gaussian polynomial moments. At $N = 0$, both distributions are the same empty experiment, with zero divergence. For neighbors differing in one bit, the sample assumption gives

$$\mathrm{KL}(P_\theta^N \| P_\omega^N) \le \frac{Np}{ds^2} \le \frac{1}{8}. \tag{67}$$

We include the testing inequalities. For probability densities a, b of measures P, Q against a common dominating measure, let $A = \{a \ge b\}$. Then $\mathrm{TV}(P, Q) = P(A) - Q(A)$. Jensen's inequality for the density ratio on A and its complement gives, with $x = P(A)$ and $y = Q(A)$,

$$\mathrm{KL}(P\|Q) \ge x \log \frac{x}{y} + (1-x) \log \frac{1-x}{1-y}.$$

For $0 < y < 1$, the right side as a function of x has value and derivative zero at $x = y$, and second derivative $1/[x(1-x)] \ge 4$. Integrating twice proves it is at least $2(x-y)^2$. Endpoint cases follow by continuity or have infinite divergence unless the corresponding mass agrees. Thus

$$\mathrm{TV}(P, Q) \le \sqrt{\mathrm{KL}(P\|Q)/2}.$$

By (67), neighboring observation laws have total variation at most $1/4$. For any test selecting P on a measurable set B, its sum of error probabilities is

$$P(B^c) + Q(B) = 1 - [P(B) - Q(B)] \ge 1 - \mathrm{TV}(P, Q).$$

Randomized tests satisfy the same bound after averaging their common independent seed. Appending this seed leaves the divergence unchanged, so all learner processing may be included in the test.

Decode an arbitrary learner's output by minimizing $\|\widehat{f} - f_\theta\|_2$ over the finite cube, with a fixed ordering to break ties. Each nonnegative risk integral is measurable, so this selector is measurable. If $\widehat{f} \notin L^2(\gamma_D)$, its distance from every cube function is infinite; otherwise adding that L^2 function would put $\widehat{f}$ in L^2. Choose the first index on such outcomes. Their true risk is infinite, so the lower bound below holds there without applying a triangle inequality to infinite norms.

For an L^2 outcome, nearest-point minimality gives

$$\|f_\theta - f_{\widehat{\theta}}\|_2 \le \|f_\theta - \widehat{f}\|_2 + \|\widehat{f} - f_{\widehat{\theta}}\|_2 \le 2\|\widehat{f} - f_\theta\|_2.$$

Together with (65), this proves on every outcome

$$R_{u_\theta}(\widehat{f}) \ge \frac{H(\theta, \widehat{\theta})}{2d}. \tag{68}$$

Give θ the uniform distribution on the cube for analysis. For each coordinate j, pair vertices differing only at j. In each pair, the decoded bit is a test whose summed errors are at least $1 - \mathrm{TV} \ge 3/4$. Averaging the 2^{d-1} pairs over 2^d vertices yields $\mathbb{P}\{\widehat{\theta}_j \ne \theta_j\} \ge 3/8$. Summing coordinates gives $\mathbb{E}H \ge 3d/8$. Taking expectations in (68) gives average risk at least $3/16$. Since the prior is finite, at least one fixed unit teacher has that expected risk. Its selection precedes the observations and learner randomness. The argument permits improper and randomized outputs. □

Proof of Theorem D.1. Proposition D.3 makes the set defining N_* nonempty and gives $N_* \le K_p D s^2$. Proposition D.4 shows that every integer $N \le (D-1)s^2/(8p)$ has minimax expected risk at least $3/16 > 1/10$. Therefore

$$N_* > \frac{(D-1)s^2}{8p} \ge \frac{Ds^2}{16p},$$

where the last inequality uses $D \ge 2$. This treats nonintegral thresholds and the case in which the lower sample cap permits only $N = 0$. The constants depend on fixed p. The result concerns fixed expected accuracy, without asserting an optimal varying-accuracy law, a confidence-level lower with these constants, or a common hard teacher for all learners and dimensions. □

D.4 Vanishing risk above the statistical exponent

Corollary D.5. *Fix odd $p \geq 3$, $a > 0$, and $r > 1 + 2a$. Put $\gamma = (r - 1 - 2a)/4$, $s = D^a$ and $N_D = \lceil D^r \rceil$. For all sufficiently large integers D, there is a deterministic proper learner using N_D observations with worst-teacher expected prediction risk at most $D^{-\gamma}$.*

Proof. Use Proposition D.3 with $\epsilon_D = D^{-\gamma} \in (0, 1]$. Its proof gives $N_{\text{up}} \leq B_p(\epsilon_D)Q_p(\epsilon_D)Ds^2$, including all integer ceilings. For $D \geq 2$,

$$B_p(\epsilon_D) \leq 4097(C_p + 1)D^{2\gamma}.$$

Since $D^{\gamma/2} \geq 1$ and $\log_2 D \geq 1$,

$$\begin{aligned} Q_p(\epsilon_D) &\leq \log_2(1 + 4\sqrt{p}) + 6 + (3\gamma/2)\log_2 D \\ &\leq [\log_2(1 + 4\sqrt{p}) + 6 + 3\gamma/2]\log_2 D. \end{aligned}$$

Consequently

$$N_{\text{up}} \leq C_{p,\gamma} D^{1+2a+2\gamma} \log_2 D$$

for a finite constant depending only on p, γ. Because $r = 1 + 2a + 4\gamma$, the ratio of the right side to D^r tends to zero. Thus $N_{\text{up}} \leq D^r \leq N_D$ for all sufficiently large D. Ignoring extra observations preserves the proper predictor and its uniform expected-risk bound ϵ_D. □

D.5 Statistical prediction without the link or noise level

Fix an integer $P \geq 3$, put $I = \{1, 2, \ldots, P\}$ and $K = |I|$, and let

$$f_c(z) = \sum_{q \in I} c_q h_q(z), \qquad \|c\|_2 = 1, \qquad F_{c,u}(x) = f_c(u^\top x), \quad \|u\| = 1.$$

The observations are independent pairs $X_i \sim \gamma_D, Y_i = F_{c,u}(X_i) + sG_i$, with independent standard Gaussian noises. The query is independent with law γ_D. The learner below does not receive c, its lowest nonzero order, or s. No quadratic-signal condition is required. All bounds are uniform over coefficients at each dimension, so the coefficient vector may vary with D.

Theorem D.6. *There is a finite constant $C_P^\star$ and a collection of deterministic measurable learners indexed only by P, D, N, ϵ such that, for all integers $D \geq 2$, $N \geq 0$, all real $s \geq 1$ and $0 < \epsilon \leq 1$, the expected prediction risk is at most ϵ, uniformly over normalized c and unit u, whenever*

$$N \geq C_P^\star D s^2 \epsilon^{-2} \log(e/\epsilon).$$

The output belongs to the normalized polynomial ridge family with orders in I. No computational efficiency is asserted.

Proof. The Gaussian generating-function calculation in Lemma D.2 gives, for jointly standard Gaussian Z, W with correlation ρ,

$$\mathbb{E}[h_j(Z)h_q(W)] = \mathbf{1}_{\{j=q\}}\rho^q.$$

Thus every $F_{c,u}$ has squared norm one, and

$$\langle F_{c,u}, F_{c,v} \rangle = \sum_{q \in I} c_q^2 (u^\top v)^q. \tag{69}$$

For $0 \leq \rho \leq 1$, summing the geometric identity for $1 - \rho^q$ against c_q^2 gives

$$1 - \rho \leq 1 - \sum_{q \in I} c_q^2 \rho^q \leq P(1 - \rho). \tag{70}$$

Consequently $\|F_{c,u} - F_{c,v}\|_2^2 \leq P\|u - v\|_2^2$ when $u^\top v \geq 0$. At a common direction, $\|F_{c,v} - F_{a,v}\|_2 = \|c - a\|_2$ by orthonormality.

Put

$$H_P(z) = \sum_{q \in I} h_q(z)^2, \qquad M_P = \mathbb{E}[H_P(Z)^2], \qquad \Delta = M_P + s^2.$$

The finite Gaussian polynomial moment M_P depends only on P, and $M_P \geq K^2 \geq 1$. Cauchy–Schwarz gives $f_c^2 \leq H_P$ pointwise, so $\mathbb{E}F_{c,u}(X)^4 \leq M_P$. For every normalized candidate $g = F_{a,v}$ the same bound holds. Hence

$$\mathbb{E}[F_{c,u}(X)^2 g(X)^2] \leq M_P, \qquad \mu_g := \mathbb{E}[Yg(X)] = \langle F_{c,u}, g\rangle, \qquad \mathrm{Var}(Yg(X)) \leq \Delta. \quad (*)$$

The last bound averages the independent centered noise and uses $\|g\|_2 = 1$.

Fix nets of the unit spheres in $\mathbb{R}^D$ and $\mathbb{R}^K$ with radii $r_u = \sqrt{\epsilon/(16P)}$ and $r_c = \sqrt{\epsilon}/4$. The volume argument in Proposition D.3, applied in the respective ambient dimensions, gives cardinalities at most $(1 + 2/r_u)^D$ and $(1 + 2/r_c)^K$. This also covers $K = 1$. Fix their orderings before observing data and list all M pairs as candidate functions. For every true c, u, there is a candidate $g_0 = F_{a_0,v_0}$ with

$$\|F_{c,u} - g_0\|_2^2 \leq 2Pr_u^2 + 2r_c^2 = \epsilon/4.$$

Indeed $u^\top v_0 = 1 - \|u - v_0\|^2/2 \geq 0$, so the preceding direction and coefficient inequalities apply. It follows that $\mu_{g_0} \geq 1 - \epsilon/8$.

Set

$$q = 2\left\lceil \frac{1}{2}\log_2 \frac{8M}{\epsilon} \right\rceil + 1, \qquad b = \lfloor N/q \rfloor. \quad (71)$$

If $b = 0$, output the first candidate. Otherwise use q disjoint blocks of size b, ignoring any remaining observations. For each candidate let $\widehat{\mu}_g$ be the median of the block averages of $Yg(X)$. Output a maximizing candidate with a fixed tie ordering. The scores, finite order statistics and selected polynomial are jointly measurable. The definition uses no coefficient or noise parameter.

For the analysis only, put

$$b_{\mathrm{req}} = \left\lceil 4096\Delta/\epsilon^2 \right\rceil, \qquad N_{\mathrm{req}} = qb_{\mathrm{req}}.$$

If $N \geq N_{\mathrm{req}}$, then $b \geq b_{\mathrm{req}}$. By (*) and Chebyshev, a block mean deviates from μ_g by more than $t = \epsilon/16$ with probability at most $\Delta/(bt^2) \leq 1/16$. For a fixed candidate these events are independent across blocks. If its median deviates by more than t, at least $(q+1)/2$ blocks deviate. A union bound over subsets gives probability at most $2^q(1/16)^{(q+1)/2} \leq 2^{-q}$. A union bound over candidates therefore gives

$$\mathbb{P}\{\max_g |\widehat{\mu}_g - \mu_g| > t\} \leq \epsilon/8.$$

No independence between different candidates is used. On the complementary event, the selected candidate satisfies

$$\mu_{\widehat{g}} \geq \widehat{\mu}_{\widehat{g}} - t \geq \widehat{\mu}_{g_0} - t \geq \mu_{g_0} - 2t \geq 1 - \epsilon/4.$$

Its squared prediction error is at most $\epsilon/2$. On every outcome the target and selected candidate have norm one, so the squared error is at most four. The full expected risk is therefore at most $\epsilon/2 + 4(\epsilon/8) = \epsilon$.

To bound the sample requirement, define

$$Q_P^\star = \log_2(1 + 8\sqrt{P}) + K\log_2 9 + 6 + \frac{K+3}{2\log 2},$$
$$C_P^\star = 4097(M_P + 1)Q_P^\star.$$

The net cardinalities and (71) give

$$\begin{aligned} q &\leq D\log_2(1 + 8\sqrt{P/\epsilon}) + K\log_2(1 + 8/\sqrt{\epsilon}) + \log_2(8/\epsilon) + 3 \\ &\leq D\left[\log_2(1 + 8\sqrt{P}) + K\log_2 9 + 6 + \frac{K+3}{2}\log_2(1/\epsilon)\right] \\ &\leq DQ_P^\star \log(e/\epsilon). \end{aligned}$$

The middle inequality uses $\epsilon \leq 1$ and $D \geq 1$. Also $s \geq 1$ gives $b_{\mathrm{req}} \leq 4097(M_P + 1)s^2\epsilon^{-2}$. Thus $N_{\mathrm{req}} \leq C_P^\star Ds^2\epsilon^{-2}\log(e/\epsilon)$, including the integer ceilings. This proves the theorem for every allowed sample count. The generally exponential nets are the reason no efficient search claim is made. □

Proposition D.7. *Let $D \geq 2$ and $N \geq 0$ be integers, let $s \geq 1$, and fix any normalized link f_c with orders in I. Even if c and s are supplied, every jointly measurable learner, including improper learners and learners with an independent random seed, has a fixed unit teacher with expected risk at least*

$$1 - \frac{1}{D} - \frac{PN}{(D-1)s^2}.$$

Proof. Let ν be normalized uniform measure on the unit sphere in $\mathbb{R}^D$, and take $U \sim \nu$ independently of the design and noise. Work in the finite-dimensional Hilbert space $\mathcal{H} = \bigoplus_{q=1}^{P} \mathcal{H}_q \subset L^2(\gamma_D)$ of Gaussian chaoses, and abbreviate $F_u = F_{c,u}$. Equation (69) gives $\|F_u\|^2 = 1$. For every $-1 \leq \rho \leq 1$, $1 - \rho^q = (1-\rho)\sum_{j=0}^{q-1} \rho^j \leq q(1-\rho)$. Consequently

$$\|F_u - F_v\|^2 = 2\sum_q c_q^2 (1 - (u^\top v)^q) \leq P\|u - v\|^2.$$

For each unit $g \in \mathcal{H}$, the function $h_g(u) = \langle F_u, g\rangle$ is smooth and bounded on the sphere. Applying the last inequality along a unit-speed spherical curve gives $\|\nabla h_g\| \leq \sqrt{P}$.

Corollary 2 of Dolbeault et al. (2013), arXiv version 1, states the logarithmic Sobolev inequality

$$\mathrm{Ent}_\nu(\varphi^2) \leq \frac{2}{D-1}\mathbb{E}_\nu\|\nabla\varphi\|^2, \qquad \mathrm{Ent}_\nu(V) = \mathbb{E}_\nu[V\log(V/\mathbb{E}_\nu V)],$$

for smooth φ; its sphere dimension is $D - 1 \geq 1$. Put $H_g = h_g - \mathbb{E}_\nu h_g$ and $\psi(\lambda) = \log \mathbb{E}_\nu e^{\lambda H_g}$. Substituting $\varphi = e^{\lambda H_g/2}$ and dividing by $\mathbb{E}_\nu e^{\lambda H_g}$ gives, for $\lambda > 0$,

$$\lambda\psi'(\lambda) - \psi(\lambda) \leq \frac{P\lambda^2}{2(D-1)}.$$

Since $\psi(0) = \psi'(0) = 0$, integration of $(\psi(\lambda)/\lambda)'$ from zero to λ yields $\psi(\lambda) \leq P\lambda^2/(2(D-1))$. All differentiations are valid because H_g is bounded.

Fix a probability Q on the sphere with finite $\mathrm{KL}(Q\|\nu)$. Nonnegativity of KL relative to the tilted probability $e^{\lambda H_g}\nu/\mathbb{E}_\nu e^{\lambda H_g}$ implies

$$\langle \mathbb{E}_Q F - \mathbb{E}_\nu F, g\rangle \leq \frac{\mathrm{KL}(Q\|\nu)}{\lambda} + \frac{P\lambda}{2(D-1)}.$$

If the mean difference is nonzero, choose g in its direction and minimize the right side over $\lambda > 0$. A zero mean difference is immediate, and zero KL is covered by letting $\lambda \downarrow 0$. Thus for every such Q,

$$\|\mathbb{E}_Q F - \mathbb{E}_\nu F\|^2 \leq \frac{2P}{D-1}\,\mathrm{KL}(Q\|\nu). \tag{72}$$

This deterministic inequality holds for every posterior law; it does not interchange expectation and a direction supremum.

Write $W = (X_i, Y_i)_{i=1}^N$ and let Q_0 be the data law with the same independent Gaussian inputs and independent $N(0, s^2)$ labels. The Gaussian mean-shift calculation gives

$$\mathbb{E}_U\,\mathrm{KL}(\mathcal{L}(W \mid U)\|Q_0) = \frac{1}{2s^2}\sum_{i=1}^{N}\mathbb{E}[F_U(X_i)^2] = \frac{N}{2s^2}.$$

The relative-entropy chain rule writes its left side as $I(U;W) + \mathrm{KL}(\mathcal{L}(W)\|Q_0)$. Therefore

$$\mathbb{E}_W\,\mathrm{KL}(\mathcal{L}(U \mid W)\|\nu) = I(U;W) \leq \frac{N}{2s^2}.$$

Regular posteriors exist on these standard Borel spaces and have finite KL almost surely. Their feature means are measurable finite-dimensional vectors. Set $m_W = \mathbb{E}[F_U \mid W]$ and $m_0 = \mathbb{E}_\nu F_U$. Applying (72) to each posterior and then averaging gives

$$\mathbb{E}\|m_W - m_0\|^2 \leq \frac{PN}{(D-1)s^2}.$$

The same statements cover $N = 0$, with the empty experiment.

If a learner has infinite prior-averaged risk, the proposition already follows. Otherwise its output belongs to $L^2(\gamma_D)$ for almost every training sample and seed, since the target has norm one. Project onto $\mathcal{H}$ to decrease risk. The coefficients of this projection are integrals against a finite Hermite basis and hence measurable; assign zero on the exceptional null set. Independence of the seed leaves the posterior unchanged. The conditional squared-loss identity consequently bounds its prior-averaged risk below by

$$\mathbb{E}\|F_U - m_W\|^2 = 1 - \|m_0\|^2 - \mathbb{E}\|m_W - m_0\|^2.$$

Here $\mathbb{E}m_W = m_0$ and conditional expectation is the orthogonal projection minimizing squared loss.

For independent $U, V \sim \nu$, sign symmetry and (69) give

$$M_D(c) := \|m_0\|^2 = \sum_{\substack{1 \le q \le P \\ q \text{ even}}} c_q^2 \mathbb{E}(U^\top V)^q \le \mathbb{E}(U^\top V)^2 = \frac{1}{D}.$$

Indeed $|U^\top V|^q \le |U^\top V|^2$ for every positive even q, and rotational symmetry gives $\mathbb{E}VV^\top = I_D/D$. Combining the last three bounds proves the claimed average-risk lower. At least one fixed unit teacher has risk at least that average bound. Its choice precedes the observations and learner randomness, as required. □

At expected accuracy $1/10$, Theorem D.6 gives a common unknown-parameter upper of order Ds^2, with a constant depending only on P. At $N \le (D-1)s^2/(8P)$, Proposition D.7 gives risk at least $3/8$ even when the link and noise are known. This cutoff is at least $Ds^2/(16P)$ for $D \ge 2$. The sample order is thus $\Theta_P(Ds^2)$ at each actual noise level. There is no finite-sample uniform guarantee over unbounded noise levels, nor a claim of optimal dependence on varying accuracy.

Corollary D.8. *For a fixed integer $P \ge 3$, there is a predictor collection depending only on P, D, N such that, for every fixed $a > 0$ and $r > 1 + 2a$, all sufficiently large D, $s = D^a$, and all integers $N \ge \lceil D^r \rceil$, its expected risk is at most $1/\log(eD)$ uniformly in all normalized c and unit u.*

Proof. Use (71) with $\epsilon_D = 1/\log(eD)$. The required count is at most

$$C_P^\star D s^2 \log^2(eD) \log(e \log(eD)).$$

Its ratio to D^r tends to zero, since $r > 1 + 2a$. Thus every stated sample count eventually meets Theorem D.6's condition. The accuracy choice uses only D, proving the asserted parameter independence. □

E The affine-label baseline

We determine the risk order when labels enter affinely, while allowing arbitrary measurable processing of the design and query. This comparison uses the same loss as the main theorem.

Theorem E.1. *Fix an odd integer $p \ge 3$, an integer $D \ge p$, and $s \ge 1$. Put $a_{p,D} = (D)_p/D^p$, where $(D)_p = D(D-1)\cdots(D-p+1)$. For every integer $N \ge 0$,*

$$\mathcal{R}(D, N, 1, p, s) \ge \frac{a_{p,D}}{1 + p!N/(s^2 D^p)}. \tag{73}$$

For every integer $N \ge 0$, with C_p as in (28) and $V = (C_p + s^2)\binom{D+p-1}{p}$,

$$\mathcal{R}(D, N, 1, p, s) \le \frac{V}{N + V}. \tag{74}$$

In particular, uniformly for $D \ge p^2$, $s \ge 1$ and $N \ge 1$,

$$\mathcal{R}(D, N, 1, p, s) = \Theta_p\big(\min\{1, s^2 D^p/N\}\big). \tag{75}$$

At $N = 0$ the minimax risk is one. For fixed p, as $D \to \infty$ and $s \to \infty$,

$$\mathcal{R}(D, N, 1, p, s) \sim \frac{1}{1 + p!N/(s^2 D^p)} \tag{76}$$

uniformly over integers $N \ge 0$.

Proof. For the lower, let $\varepsilon_1, \dots, \varepsilon_D$ be independent uniform signs and put $U = D^{-1/2}\varepsilon$. Write $\psi_\alpha(x) = \prod_j h_{\alpha_j}(x_j)$. The Gaussian Hermite expansion in the proof of Lemma D.2 gives

$$f_U(x) = D^{-p/2} \sum_{|\alpha|=p} \sqrt{\frac{p!}{\alpha!}} \prod_j \varepsilon_j^{\alpha_j} \psi_\alpha(x).$$

Let $S(\alpha)$ be the set of coordinates where α_j is odd. Since p is odd, its cardinality is odd and lies between one and p. For every such subset $S \subseteq [D]$, define

$$g_S = D^{-p/2} \sum_{\substack{|\alpha|=p \\ S(\alpha)=S}} \sqrt{\frac{p!}{\alpha!}}\, \psi_\alpha, \qquad \lambda_S = \|g_S\|_2^2, \qquad e_S = g_S/\sqrt{\lambda_S}.$$

Each λ_S is positive: start with ones on S and add the even remainder $p - |S|$ to one coordinate of S. The functions g_S have disjoint supports in the orthonormal Hermite basis, so the e_S form an orthonormal family. Writing $\chi_S = \prod_{j\in S} \varepsilon_j$ gives

$$f_U = \sum_S \theta_S e_S, \qquad \theta_S = \sqrt{\lambda_S}\chi_S.$$

The sign characters are centered and orthogonal. Thus $\mathbb{E}\theta = 0$ and $\mathrm{Cov}(\theta) = C$, where $C = \mathrm{diag}(\lambda_S)$ is positive definite. No independence or Gaussian law for the coordinates of θ is asserted. Since $\|f_U\|_2^2 = 1$, $\mathrm{tr}\, C = 1$. For $|S| = p$, the only multiindex is $\alpha = \mathbf{1}_S$, so

$$\lambda_S = p!/D^p, \qquad \#\{S : |S| = p\} = \binom{D}{p}. \tag{77}$$

Condition on the training design and the independent learner seed. Put $\Phi_{i,S} = e_S(X_i)$. The conditional observation model is

$$Y = \Phi\theta + sG, \qquad \mathbb{E}Y = 0, \qquad M := \mathrm{Cov}(Y) = \Phi C \Phi^\top + s^2 I.$$

The prior, design, seed and noise are independent before forming Y, so the conditional covariance of θ remains C.

We justify projecting an arbitrary affine output onto the span of the e_S. If its prior-average risk is infinite, the lower is immediate. Otherwise, at almost every fixed design and seed, its conditional expected squared norm is finite, since

$$\mathbb{E}\|\widehat{f}\|_2^2 \le 2\mathbb{E}\|\widehat{f} - f_U\|_2^2 + 2 < \infty.$$

Write the output as $a_0(x) + a(x)^\top Y$. At these fixed designs and seeds,

$$\mathbb{E}[\widehat{f}(x)^2] = a_0(x)^2 + a(x)^\top M a(x) \ge a_0(x)^2 + s^2 \sum_i a_i(x)^2.$$

Integrating in x and applying Tonelli puts every coefficient in $L^2(\gamma_D)$. Their orthogonal projections are defined by integrals against the finite family e_S and are jointly measurable. Set them to zero on the null exceptional set. Projection preserves affine dependence and decreases risk, because the target belongs to the projected subspace. The projected coefficient vector has the form $b + AY$.

Its conditional risk is

$$\|b\|^2 + \mathrm{tr}\, C - 2\,\mathrm{tr}(A\Phi C) + \mathrm{tr}(AMA^\top).$$

The matrix M is positive definite. Completing the square in A shows that the minimum is attained at $b = 0$ and $A = C\Phi^\top M^{-1}$, with value

$$\mathrm{tr}(C - C\Phi^\top M^{-1}\Phi C) = \mathrm{tr}\big((C^{-1} + s^{-2}\Phi^\top\Phi)^{-1}\big). \tag{78}$$

For the matrix identity, multiply the matrix inside the first trace by $C^{-1} + s^{-2}\Phi^\top\Phi$ and use $(I + s^{-2}\Phi C\Phi^\top)M^{-1} = s^{-2}I$; the product is the identity.

For a positive definite matrix B and a coordinate vector v, Cauchy–Schwarz gives

$$1 = \langle B^{1/2}v, B^{-1/2}v\rangle^2 \le (v^\top B v)(v^\top B^{-1} v).$$

Therefore the S diagonal entry of the inverse in (78) is at least

$$\frac{1}{\lambda_S^{-1} + s^{-2}\sum_i e_S(X_i)^2}.$$

The function $z \mapsto (\lambda_S^{-1} + z)^{-1}$ is convex on $[0, \infty)$. Averaging over the design, Jensen's inequality and $\mathbb{E}e_S(X_i)^2 = 1$ give prior-average risk at least

$$\sum_S \frac{\lambda_S}{1 + N\lambda_S/s^2}.$$

Keeping the terms in (77) proves (73). Every learner's worst-teacher risk is at least its prior average; the bound therefore survives the infimum over affine learners and independent randomization. When $N = 0$, the conditional risk calculation has value $\operatorname{tr} C = 1$, attained by the zero predictor.

For the upper with $N \geq 1$, first use the empirical Hermite tensor without matrix blocks:

$$\widehat{f}_N(x) = J_p(\widehat{C}_N)(x) = \frac{1}{p!N}\sum_{i=1}^{N} Y_i\langle R_p(X_i), R_p(x)\rangle.$$

It is linear in the labels. Both $\widehat{C}_N$ and $u^{\otimes p}$ are symmetric tensors. By the isometry (36) and unbiasedness in (40),

$$\mathbb{E}_u[(\widehat{f}_N(x) - f_u(x))^2] = \mathbb{E}_u\|\widehat{C}_N - u^{\otimes p}\|_F^2 = \operatorname{tr}\operatorname{Cov}_u(\widehat{C}_N) \leq \frac{C_p + s^2}{N}\operatorname{tr}\Pi.$$

All moments are finite by Lemma B.2. The symmetric tensor space has one basis element for each multiplicity vector of a p-tuple from D coordinates, so $\operatorname{tr}\Pi = \binom{D+p-1}{p}$. Set $c = N/(N+V)$. Unbiasedness and the finite second moment give

$$\mathbb{E}_u\|c\widehat{f}_N - f_u\|_2^2 = (1-c)^2 + c^2\mathbb{E}_u\|\widehat{f}_N - f_u\|_2^2 \leq \frac{V^2 + NV}{(N+V)^2} = \frac{V}{N+V}.$$

The cross term vanishes because $\mathbb{E}_u[\widehat{f}_N - f_u] = 0$ in the finite-dimensional Hermite space. Equivalently, the predictor is

$$\frac{1}{p!(N+V)}\sum_{i=1}^{N} Y_i\langle R_p(X_i), R_p(x)\rangle.$$

This formula is zero when $N = 0$ and then has risk one. Its coefficients depend only on known parameters and the design and query, so it belongs to $\mathcal{P}_1$. This proves (74) for every $N \geq 0$.

For $D \geq p^2$, induction using $(1-x)(1-y) \geq 1 - x - y$ for $x, y \in [0, 1]$ gives

$$a_{p,D} = \prod_{j=0}^{p-1}(1 - j/D) \geq 1 - \frac{p(p-1)}{2D} \geq \frac{1}{2}.$$

For $z = N/(s^2 D^p) > 0$, the finite lower is at least

$$\frac{1}{2(1 + p!z)} \geq \frac{1}{2(1+p!)}\min\{1, z^{-1}\}.$$

Also $\binom{D+p-1}{p} \leq D^p$, since each multiset has an ordered representative, and $C_p + s^2 \leq (C_p + 1)s^2$. The upper is consequently at most $(C_p + 1)\min\{1, s^2 D^p/N\}$. These bounds prove (75).

For (76), put $v_0 = s^2 D^p/p!$. For fixed p,

$$\frac{V}{v_0} = \left(1 + \frac{C_p}{s^2}\right)\prod_{j=0}^{p-1}\left(1 + \frac{j}{D}\right) \longrightarrow 1, \qquad a_{p,D} \longrightarrow 1.$$

Since $V \geq v_0$, the finite lower and upper give, for every integer $N \geq 0$,

$$a_{p,D} \leq \frac{\mathcal{R}(D, N, 1, p, s)}{(1 + N/v_0)^{-1}} \leq \frac{V}{v_0}\frac{N + v_0}{N + V} \leq \frac{V}{v_0}.$$

Both outer terms converge to one independently of N, which proves the stated uniform relative asymptotic. □

For all parameters in (4), the first-energy certificate also gives the direct finite lower

$$\sup_{\|u\|=1} \mathbb{E}_u[(\widehat{f}(x) - f_u(x))^2] \geq 1 - \frac{(2p-1)!!N}{s^2 D^p}, \qquad \widehat{f} \in \mathcal{P}_1.$$

Indeed, (16) gives sign-prior average risk at least $1 - E_0 - E_1$. Here $E_0 = 0$, and Section A.6 gives $E_1 \leq (2p-1)!!N/(s^2 D^p)$. Rescaling labels by the known positive s preserves affine degree, and worst-teacher risk is at least the prior average.

F Moment identities and optimal scalar signals

F.1 Proof of the moment identities

Proposition F.1. *Let $p = 2m + 1 \geq 3$, let Z, G be independent standard Gaussians, let $s \geq 0$ be finite, and let $Y = h_p(Z) + sG$. Then*

$$\mathbb{E}[ZY^3] = \frac{3(p!)^{3/2}}{(m+1)!(m!)^2} > 0. \tag{79}$$

Writing $M_{4,p} = \mathbb{E}[h_p(Z)^4]$ and $\lambda_k^2 = \mathbb{E}[\mathbb{E}[h_k(Z) \mid Y]^2]$, one has

$$\lambda_1^2 \geq \frac{\mathbb{E}[ZY^3]^2}{\mathbb{E}[Y^6]} > 0, \tag{80}$$

$$\lambda_2^2 \geq \frac{2p^2}{M_{4,p} - 1 + 4s^2 + 2s^4}, \tag{81}$$

$$\lambda_p^2 \geq \frac{1}{1+s^2}. \tag{82}$$

Proof. The Hermite generating function gives $h_p' = \sqrt{p}\, h_{p-1}$. Gaussian integration by parts yields

$$\mathbb{E}[Z h_p(Z)^3] = 3\sqrt{p}\, \mathbb{E}[h_p(Z)^2 h_{p-1}(Z)].$$

The expectation of three generating functions is $\exp(t_1t_2 + t_1t_3 + t_2t_3)$. In its coefficient of $t_1^p t_2^p t_3^{p-1}$, the pair multiplicities are forced to be $(m+1, m, m)$. Thus

$$\mathbb{E}[h_p^2 h_{p-1}] = \frac{p!\sqrt{(p-1)!}}{(m+1)!(m!)^2}.$$

This proves (79) at $s = 0$. Expanding Y^3 adds only $3s^2\mathbb{E}[Zh_p(Z)]$, which vanishes by orthogonality of Hermite degrees one and p. All moments used here are finite Gaussian polynomial moments.

For any square-integrable function $T(Y)$, conditional expectation and Cauchy–Schwarz give

$$\mathbb{E}[h_k(Z)T(Y)]^2 \leq \lambda_k^2 \mathbb{E}[T(Y)^2]. \tag{83}$$

Taking $k = 1$ and $T(Y) = Y^3$ proves (80). For $T(Y) = Y^2 - 1 - s^2$,

$$\mathbb{E}[T(Y)^2] = M_{4,p} - 1 + 4s^2 + 2s^4.$$

Indeed $\mathbb{E}Y^2 = 1 + s^2$ and $\mathbb{E}Y^4 = M_{4,p} + 6s^2 + 3s^4$ by independence and Gaussian noise moments. The recurrence $Zh_p = \sqrt{p+1}h_{p+1} + \sqrt{p}h_{p-1}$ gives $\mathbb{E}[Z^2 h_p^2] = 2p + 1$. Since $h_2(Z) = (Z^2 - 1)/\sqrt{2}$,

$$\mathbb{E}[h_2(Z)T(Y)] = \mathbb{E}[h_2(Z)h_p(Z)^2] = \sqrt{2}\, p.$$

The recurrence follows by differentiating the generating function in its scalar parameter and comparing coefficients. Equation (83) now proves (81). Finally $\mathbb{E}[h_p(Z)Y] = 1$ and $\mathbb{E}Y^2 = 1 + s^2$, so the same inequality with $k = p$ and $T(Y) = Y$ proves (82). □

F.2 Conditional signals at high noise

Conditional Hermite coefficients optimize over square-integrable transformations of one label, as in Damian et al. (2024). We evaluate three such coefficients at high noise and compare the fixed-order recovery guarantees. The expansion of the second conditional moment is also the mechanism used by Defilippis et al. (2026). These scalar calculations do not replace the prediction lower in Appendix A.

Lemma F.2. *Let Z, G be independent standard Gaussians and let f be a fixed polynomial with $\mathbb{E}[f(Z)] = 0$ and $\mathbb{E}[f(Z)^2] = 1$. Put $H = f(Z)$ and $S_t = G + tH$ for $t > 0$. Let $V = v(Z)$ for a fixed polynomial v. Fix an integer $r \geq 1$ and suppose $\mathbb{E}[VH^j] = 0$ for every integer $0 \leq j < r$. Then*

$$\lim_{t \downarrow 0} t^{-2r} \mathbb{E}[\mathbb{E}[V \mid S_t]^2] = \frac{\mathbb{E}[VH^r]^2}{r!}.$$

Proof. Write ϕ for the standard Gaussian density and q_t for the density of S_t. Since $\mathbb{E}H = 0$ and $\mathbb{E}H^2 = 1$, Jensen's inequality gives, for every real y,

$$\frac{q_t(y)}{\phi(y)} = \mathbb{E} \exp(tyH - t^2H^2/2) \geq e^{-t^2/2}. \tag{84}$$

The expectation is finite because its integrand is at most $e^{y^2/2}$. Put $K = 16(r+1)$ and, for sufficiently small t,

$$A_t = \{|Z| \leq \sqrt{K \log(1/t)}\}.$$

For a fixed polynomial P and $R \geq 1$, integration by parts in the Gaussian tail bounds $\mathbb{E}[|P(Z)|\mathbf{1}_{\{|Z|>R\}}]$ by a constant times $e^{-R^2/2}$ times a fixed power of R. Indeed, after bounding $|P(z)|$ by a finite sum of powers of $|z|$, the integrals of $z^j e^{-z^2/2}$ on $[R, \infty)$ satisfy

$$\int_R^\infty z^j e^{-z^2/2}\, dz = R^{j-1} e^{-R^2/2} + (j-1) \int_R^\infty z^{j-2} e^{-z^2/2}\, dz$$

for $j \geq 1$, with the $j = 0$ integral at most $R^{-1}e^{-R^2/2}$. Applying this bound to V^2 and $|VH^j|$ shows, for each fixed j,

$$\begin{aligned} \mathbb{E}[V^2 \mathbf{1}_{A_t^c}] &= o(t^{2r}), \\ \mathbb{E}[|VH^j| \mathbf{1}_{A_t^c}] &= O\big(t^{K/2}(1 + \log(1/t))^{c_j}\big) \end{aligned} \tag{85}$$

for finite constants c_j. Conditional expectation is an L^2 contraction, so the L^2 norm of $\mathbb{E}[V\mathbf{1}_{A_t^c} \mid S_t]$ is $o(t^r)$.

Define

$$n_t(y) = \mathbb{E}[V\mathbf{1}_{A_t}\phi(y - tH)], \qquad c_j(t) = \mathbb{E}[V\mathbf{1}_{A_t}H^j], \qquad M_t = \max\{1, \max_{A_t} |H|\}.$$

Here $M_t = O_{f,r}((1 + \log(1/t))^{\deg(f)/2})$. For an independent copy (V', H', A_t'), completing the square in the Gaussian density gives

$$\begin{aligned} \int_{\mathbb{R}} \frac{n_t(y)^2}{\phi(y)}\, dy &= \mathbb{E}[VV'\mathbf{1}_{A_t}\mathbf{1}_{A_t'} e^{t^2 HH'}] \\ &= \sum_{j \geq 0} \frac{t^{2j} c_j(t)^2}{j!}. \end{aligned} \tag{86}$$

Both interchanges are absolutely justified: V, H, V', H' are bounded on the indicated events, and the integrated absolute Gaussian product is bounded by a constant times $e^{t^2 M_t^2}$. The exponential series is absolutely bounded by the same quantity.

For $j < r$, the full moment is zero, so (85) implies $t^{2j} c_j(t)^2 = o(t^{2r})$. Also $c_r(t) \longrightarrow \mathbb{E}[VH^r]$ by integrability. For $j \geq r + 1$, put $C = \mathbb{E}[|V||H|^{r+1}] < \infty$. Then $|c_j(t)| \leq C M_t^{j-r-1}$, and hence

$$\sum_{j \geq r+1} \frac{t^{2j} c_j(t)^2}{j!} \leq C^2 t^{2r+2} e^{t^2 M_t^2} = o(t^{2r}).$$

Since $\mathbb{E}[V\mathbf{1}_{A_t} \mid S_t = y] = n_t(y)/q_t(y)$, (84) and (86) bound its squared L^2 norm by

$$e^{t^2/2}\left(\frac{t^{2r}\mathbb{E}[VH^r]^2}{r!} + o(t^{2r})\right).$$

The L^2 triangle inequality and the tail estimate give the corresponding upper limit for the full conditional expectation. This also proves a zero limit when $\mathbb{E}[VH^r] = 0$.

For the reverse inequality, use the statistic $T_t = \mathrm{He}_r(S_t)$. The Hermite generating function and independence give

$$\mathbb{E}[T_t \mid Z] = t^r H^r, \qquad \mathbb{E}[VT_t] = t^r \mathbb{E}[VH^r].$$

Moreover $\mathbb{E}[T_t^2] \to r!$: this moment is a finite polynomial in t, and its value at zero is $\mathbb{E}[\mathrm{He}_r(G)^2] = r!$. Conditional expectation and Cauchy–Schwarz yield

$$\mathbb{E}[\mathbb{E}[V \mid S_t]^2] \geq \frac{t^{2r}\mathbb{E}[VH^r]^2}{\mathbb{E}[T_t^2]}.$$

Taking the lower limit completes the proof. □

Proposition F.3. *Fix an odd integer $p \geq 3$, let $Y = h_p(Z) + sG$, and define $\lambda_k^2(s) = \mathbb{E}[\mathbb{E}[h_k(Z) \mid Y]^2]$. With A_p as in (79), as $s \to \infty$,*

$$\lambda_1^2(s) \sim \frac{A_p^2}{6s^6}, \qquad \lambda_2^2(s) \sim \frac{p^2}{s^4}, \qquad \lambda_p^2(s) \sim \frac{1}{s^2}. \tag{87}$$

For every fixed integer $k \geq 1$ with $k \neq p$, $\lambda_k^2(s) = O_{p,k}(s^{-4})$. Each coefficient has the variational representation

$$\lambda_k^2(s) = \sup_{\substack{T:\mathbb{R}\to\mathbb{R} \text{ measurable} \\ 0<\mathbb{E}[T(Y)^2]<\infty}} \frac{\mathbb{E}[h_k(Z)T(Y)]^2}{\mathbb{E}[T(Y)^2]}. \tag{88}$$

Proof. Conditioning on Y and on $Y/s = G + s^{-1}h_p(Z)$ is equivalent for $s > 0$. Apply Lemma F.2 with $t = s^{-1}$. For $V = Z$, the moments of orders zero, one and two vanish by centering, Hermite orthogonality, and parity, respectively; the order-three moment is A_p by Proposition F.1. For $V = h_2(Z)$, the moments of orders zero and one vanish, and $\mathbb{E}[h_2(Z)h_p(Z)^2] = \sqrt{2}\,p$, also computed there. For $V = h_p(Z)$, the mean is zero and $\mathbb{E}[VH] = 1$. These applications with $r = 3, 2, 1$ prove (87).

For every fixed $k \neq p$, orthogonality gives $\mathbb{E}h_k(Z) = \mathbb{E}[h_k(Z)H] = 0$. The lemma with $r = 2$ gives

$$s^4\lambda_k^2(s) \longrightarrow \frac{\mathbb{E}[h_k(Z)H^2]^2}{2} < \infty,$$

including when the limit is zero. This proves the stated upper order. Finally (83) proves the upper bound in (88). If $\lambda_k > 0$, equality holds for $T(Y) = \mathbb{E}[h_k(Z) \mid Y]$. If $\lambda_k = 0$, every numerator is zero, so equality holds in that case as well. □

We can now compare all fixed harmonic orders in the displayed sufficient bound of Damian et al. (2024, Theorem 4.1). For fixed directional accuracy, that bound has the form

$$F_k(D, s) = \frac{(D + D^{k/2})\log^{2k}(3/\lambda_k(s))}{\lambda_k^2(s)},$$

up to constants depending on k and the accuracy; set $F_k = \infty$ when $\lambda_k = 0$. For $s = D^a$, fixed $a > 0$, the exponents of F_1, F_2, F_p are respectively $1 + 6a$, $1 + 4a$, and $p/2 + 2a$, by (87). The logarithms do not alter these exponents. For any other fixed $k \geq 2$, the upper bound on λ_k^2 gives

$$\liminf_{D\to\infty} \frac{\log F_k(D, D^a)}{\log D} \geq \max\{1, k/2\} + 4a \geq 1 + 4a.$$

Thus the smallest exponent among these fixed-order sufficient guarantees is exactly $\min\{1 + 4a, p/2+2a\}$, attained by order two or order p. This comparison allows the optimal square-integrable transformation of a single label at each order. It does not address orders growing with D, combinations of orders beyond that recovery guarantee, or arbitrary algorithms.

G Detection and prediction on the same observations

We compare prediction with a test of angular input-label dependence. The null below preserves the joint law of the label and input radius. We use a transformed-mean statistic to show that a detectable signal need not permit prediction in the stated class.

Fix an odd integer $p = 2m + 1 \geq 3$, an integer $D \geq 1$, and $s \geq 1$. Under the planted law $P_{u,N}$, for a fixed unit vector u, the N observations are independent and satisfy

$$X_i \sim \mathcal{N}(0, I_D), \qquad Y_i = h_p(u \cdot X_i) + sG_i,$$

where the G_i are independent standard Gaussians, independent of the design. To define the null Q_N, independently for each row draw

$$Z_i', G_i' \sim \mathcal{N}(0,1), \qquad W_i' \sim \mathcal{N}(0, I_{D-1}), \qquad \omega_i \sim \mathrm{Unif}(\mathbb{S}^{D-1}),$$

with all variables mutually independent before forming

$$R_i = \sqrt{(Z_i')^2 + \|W_i'\|^2}, \qquad Y_i = h_p(Z_i') + sG_i', \qquad X_i = R_i\omega_i. \tag{89}$$

For $D = 1$, W_i' is the empty vector and ω_i is a fair sign. This null depends only on the known parameters D, p, s.

For independent standard Gaussians Z, G, put $H = h_p(Z)$ and define

$$A_p = \frac{3(p!)^{3/2}}{(m+1)!(m!)^2}, \tag{90}$$

$$M_p(s) = \mathbb{E}(H + sG)^6, \qquad B_p(s) = \mathbb{E}[Z^2(H + sG)^6], \tag{91}$$

$$v_{p,D}(s) = \frac{B_p(s) + (D-1)M_p(s)}{D}.$$

We use the finite constant

$$K_p^{\mathrm{det}} = \max\left\{\begin{matrix}\mathbb{E}H^6 + 15\mathbb{E}H^4 + 60, \\ \mathbb{E}[Z^2H^6] + 15\mathbb{E}[Z^2H^4] + 45(2p+1) + 15\end{matrix}\right\}. \tag{92}$$

It depends only on p and is distinct from the statistical constant in Theorem D.1. Every entry is a finite Gaussian polynomial moment.

For $N \geq 2$, let $\phi_i = Y_i^3 X_i$ and set

$$S_N = \frac{1}{N(N-1)} \sum_{i \neq j} \langle \phi_i, \phi_j \rangle, \qquad \Delta_N = \mathbf{1}\{S_N > A_p^2/2\}. \tag{93}$$

The decision one denotes the planted law. The score and threshold do not use u or s.

Proposition G.1. *For the stated parameters and every integer $N \geq 2$, the null Q_N has standard Gaussian input marginals and the same joint law of $((Y_i, \|X_i\|))_{i=1}^N$ as every $P_{u,N}$. Uniformly over unit vectors u,*

$$\mathbb{E}_{Q_N} S_N = 0, \qquad \mathbb{E}_{P_{u,N}} S_N = A_p^2, \tag{94}$$

$$\mathrm{Var}_{Q_N}(S_N) = \frac{2Dv_{p,D}(s)^2}{N(N-1)},$$

$$\mathrm{Var}_{P_{u,N}}(S_N) = \frac{4A_p^2[B_p(s) - A_p^2]}{N} + \frac{2\{[B_p(s) - A_p^2]^2 + (D-1)M_p(s)^2\}}{N(N-1)}. \tag{95}$$

In particular,

$$Q_N\{\Delta_N = 1\} + \sup_{\|u\|=1} P_{u,N}\{\Delta_N = 0\} \leq \frac{16K_p^{\mathrm{det}}}{A_p^2}\frac{s^6}{N} + \frac{32(K_p^{\mathrm{det}})^2}{A_p^4}\frac{Ds^{12}}{N^2}. \tag{96}$$

The score has label degree six and joint degree eight in the observed inputs and labels. It can be evaluated in $O(ND)$ real-arithmetic operations with $O(D)$ streaming storage.

Proof. Rotate a fixed u to the first coordinate under the planted law and write $X = Zu+W$, with W an independent standard Gaussian in $u^\perp$. Then $\|X\| = \sqrt{Z^2 + \|W\|^2}$ and $Y = h_p(Z) + sG$. This is the same joint label-radius law as (89), independent of the choice of u. The rows are independent under each fixed hypothesis, so the complete lists have the same joint law.

The radius R in (89) has the chi distribution of a standard Gaussian norm. In polar coordinates, the rotationally invariant Gaussian density factors into uniform surface measure and radial density proportional to $r^{D-1}e^{-r^2/2}$. Thus $R\omega$ is standard Gaussian when the independent direction ω is uniform. For $D = 1$, this follows by multiplying the absolute value of a standard Gaussian by an independent fair sign. The event $R = 0$ has probability zero. One measurable construction of ω is $V/\|V\|$ for an independent standard Gaussian vector V, with any fixed value on its null zero event. This proves both assertions about the null. In particular, every test using only the labels and radii has the same distribution under the two hypotheses.

We next compute the mean of $\phi = Y^3X$. The generating function of the normalized Hermite polynomials gives $h_p' = \sqrt{p}\, h_{p-1}$. Integration by parts yields

$$\mathbb{E}[ZH^3] = 3\sqrt{p}\,\mathbb{E}[h_p(Z)^2 h_{p-1}(Z)].$$

The expectation of three Hermite generating functions is $\exp(t_1t_2 + t_1t_3 + t_2t_3)$. For the coefficient of $t_1^p t_2^p t_3^{p-1}$, the three pair multiplicities are $(m+1, m, m)$. Including the Hermite normalizations gives

$$\mathbb{E}[h_p^2 h_{p-1}] = \frac{p!\sqrt{(p-1)!}}{(m+1)!(m!)^2}.$$

Hence $\mathbb{E}[ZH^3] = A_p$. Expanding the independent Gaussian noise, the only possible additional term in $\mathbb{E}[Z(H+sG)^3]$ is $3s^2\mathbb{E}[ZH]$, which is zero by Hermite orthogonality since $p \geq 3$. Therefore

$$\mathbb{E}[Z(H+sG)^3] = A_p > 0. \tag{97}$$

For example, $A_3 = 9\sqrt{6}$.

Under the fixed planted law, the decomposition $X = Zu + W$ gives $\mathbb{E}\phi = A_p u$. Put $P = uu^\top$. Independence and $\mathbb{E}W = 0$ remove every mixed entry between Z and W, while $\mathbb{E}WW^\top = I - P$. It follows that

$$\begin{aligned} \mathbb{E}[\phi\phi^\top] &= B_p(s)P + M_p(s)(I-P), \\ \Sigma_{P_u} := \mathrm{Cov}_{P_u}(\phi) &= [B_p(s) - A_p^2]P + M_p(s)(I-P). \end{aligned} \tag{98}$$

Here $B_p(s) - A_p^2 = \mathrm{Var}(ZY^3) \geq 0$. Under the null, ω is independent of (Y, R), with $\mathbb{E}\omega = 0$ and $\mathbb{E}\omega\omega^\top = I/D$ by symmetry and unit norm, also when $D = 1$. Consequently

$$\mathbb{E}_Q\phi = 0, \qquad \Sigma_Q = \frac{\mathbb{E}[R^2Y^6]}{D} I = v_{p,D}(s)I. \tag{99}$$

The last equality uses $R^2 = Z^2 + \|W\|^2$ in the common label-radius law and the independence of W from (Z, G).

The identities $\mathbb{E}G^2 = 1, \mathbb{E}G^4 = 3$ and $\mathbb{E}G^6 = 15$ give

$$\begin{aligned} M_p(s) &= \mathbb{E}H^6 + 15s^2\mathbb{E}H^4 + 45s^4 + 15s^6, \\ B_p(s) &= \mathbb{E}[Z^2H^6] + 15s^2\mathbb{E}[Z^2H^4] + 45s^4\mathbb{E}[Z^2H^2] + 15s^6. \end{aligned} \tag{100}$$

The Hermite recurrence $Zh_p = \sqrt{p+1}h_{p+1} + \sqrt{p}\, h_{p-1}$, obtained by differentiating the generating function in its parameter, and orthogonality imply $\mathbb{E}[Z^2H^2] = 2p + 1$. All coefficients in (100) are nonnegative. Since $s \geq 1$, each of $1, s^2, s^4$ is at most s^6, so

$$M_p(s), B_p(s), v_{p,D}(s) \leq K_p^{\mathrm{det}} s^6, \qquad 0 \preceq \Sigma_{P_u}, \Sigma_Q \preceq K_p^{\mathrm{det}} s^6 I. \tag{101}$$

In particular, each covariance satisfies $\mathrm{tr}(\Sigma^2) \leq D(K_p^{\mathrm{det}})^2 s^{12}$, and (98) gives

$$(A_p u)^\top \Sigma_{P_u}(A_p u) = A_p^2[B_p(s) - A_p^2] \leq A_p^2 K_p^{\mathrm{det}} s^6. \tag{102}$$

All moments are finite. Their values can be computed using $\mathbb{E}Z^{2j} = (2j-1)\mathbb{E}Z^{2j-2}$, starting at one, and zero odd moments; the recurrence follows from Gaussian integration by parts.

For completeness, consider any iid vectors ϕ_i with finite second moment, mean μ and covariance Σ. Write $\zeta_i = \phi_i - \mu$. Expanding the score gives

$$S_N = \|\mu\|^2 + \frac{2}{N}\sum_i \langle \mu, \zeta_i\rangle + \frac{2}{N(N-1)}\sum_{i<j}\langle \zeta_i, \zeta_j\rangle. \tag{103}$$

Both sums are centered. The linear sum has variance $4\mu^\top\Sigma\mu/N$. Its covariance with the last sum is zero: if its index is absent from a pair, use independence; if it belongs to that pair, condition on all variables except the other member, whose conditional mean is zero.

Two distinct pair terms in the last sum are also uncorrelated. Disjoint pairs are independent and centered. Pairs sharing an index have an unshared centered vector; conditioning on all other vectors gives zero expectation for their product. For a single pair,

$$\mathbb{E}\langle \zeta_i, \zeta_j\rangle^2 = \sum_{a,b}\mathbb{E}[\zeta_{i,a}\zeta_{i,b}]\mathbb{E}[\zeta_{j,a}\zeta_{j,b}] = \operatorname{tr}(\Sigma^2).$$

There are $N(N-1)/2$ pairs. Every product used is integrable: the degree in each independent vector is at most two, and $|\zeta_a\zeta_b| \le (\zeta_a^2+\zeta_b^2)/2$. Thus the finite second moment justifies all expansions and gives

$$\operatorname{Var}(S_N) = \frac{4\mu^\top\Sigma\mu}{N} + \frac{2\operatorname{tr}(\Sigma^2)}{N(N-1)}. \tag{104}$$

Substituting (98) and (99) proves (94) and (95), including $D = 1$, where the perpendicular block is absent.

Chebyshev's inequality at distance $A_p^2/2$ now gives

$$Q_N\{\Delta_N = 1\} \le \frac{8D(K_p^{\text{det}})^2 s^{12}}{A_p^4 N(N-1)},$$
$$P_{u,N}\{\Delta_N = 0\} \le \frac{16K_p^{\text{det}} s^6}{A_p^2 N} + \frac{8D(K_p^{\text{det}})^2 s^{12}}{A_p^4 N(N-1)}.$$

The second error includes equality at the threshold; it still implies an absolute deviation at least $A_p^2/2$ from the planted mean. Since $N(N-1) \ge N^2/2$ for $N \ge 2$, adding these bounds proves (96), uniformly in u. The bound is valid even when its right side exceeds two.

Finally,

$$S_N = \frac{\|\sum_i \phi_i\|^2 - \sum_i \|\phi_i\|^2}{N(N-1)}.$$

Accumulate the vector sum and the scalar sum in one pass. This gives the stated arithmetic and storage counts. Each summand in (93) has label degree six. Its inner product adds two input degrees, giving joint degree eight. □

The decision Δ_N thresholds the score and is not itself a polynomial in the labels. The arithmetic count does not imply a finite-bit bound. If instead the null makes X standard Gaussian and independent of a label with the same marginal law, then $\mathbb{E}_Q\phi = 0$ and $\Sigma_Q = M_p(s)I$ by independence. Substitution in (104) gives null variance $2DM_p(s)^2/[N(N-1)]$, and (101) gives the same error bound (96). We use the stronger null (89) in the following comparison.

G.1 The coexistence interval

The next corollary combines the detector with the two prediction theorems. It gives a sufficient detection regime, not a lower bound for detection.

Corollary G.2. *Fix odd $p \ge 3$, $0 < a < 1/4$, and*

$$\max\{1+2a, 1/2+6a\} < r < 1+4a.$$

Put $s = D^a$, $N_D = \lceil D^r \rceil$, $\eta = (1 + 4a - r)/2$, $\tau = \min\{2a, \eta/2\}$, *and* $L_D = \lfloor D^{\tau/[4(3p+5)]} \rfloor$. *Put* $\gamma = (r - 1 - 2a)/4 > 0$. *For every* $D \geq 2$, *the test* (93) *satisfies*

$$Q_{N_D}\{\Delta_{N_D} = 1\} + \sup_{\|u\|=1} P_{u,N_D}\{\Delta_{N_D} = 0\} \leq \frac{16K_p^{\text{det}}}{A_p^2} D^{6a-r} + \frac{32(K_p^{\text{det}})^2}{A_p^4} D^{1+12a-2r}. \quad (105)$$

The right side tends to zero. For all sufficiently large D, *an unrestricted proper predictor has worst-teacher expected prediction risk at most* $D^{-\gamma}$, *whereas*

$$\mathcal{R}(D, N_D, L_D, p, D^a) \geq 1 - 2D^{-\tau/2}. \quad (106)$$

Proof. First, the interval for r is nonempty. Its upper endpoint exceeds $1+2a$ by $2a > 0$ and exceeds $1/2 + 6a$ by $1/2 - 2a > 0$. Since $N_D \geq D^r$ and $N_D \geq 2$, (96) gives (105). The inequalities $r > 1/2 + 6a$ imply $6a - r < -1/2$ and $1 + 12a - 2r < 0$, so both terms tend to zero.

For every odd $p \geq 3$ and $a < 1/4$,

$$p/2 + 2a - (1 + 4a) = p/2 - 1 - 2a \geq 1/2 - 2a > 0.$$

Thus the exponent b of Corollary A.3 is $1 + 4a$. Our choice of η gives $r = b - 2\eta$. As $D^r \geq 1$, we have $N_D \leq 2D^r$, and for $D \geq 2^{1/\eta}$ this yields

$$N_D \leq D^{r+\eta} = D^{b-\eta}.$$

Apply Corollary A.3 once its stated dimension threshold is also met. Rescaling the known labels from Y_i/s to Y_i preserves label degree and measurable coefficients. That corollary gives the average risk lower over the finite sign prior $U_j \in \{-D^{-1/2}, D^{-1/2}\}$. Every learner's worst fixed-teacher risk is at least this average; taking the infimum proves (106).

For the statistical assertion, the strict inequality $r > 1 + 2a$ allows Corollary D.5 with the displayed value of γ. It supplies a proper predictor with worst-teacher expected risk at most $D^{-\gamma}$. The theorem's dimension requirement, $2^{1/\eta}$ and this statistical threshold can all be met because p, a, r are fixed. □

For comparison with the finite sign prior, first apply the detection bound at each fixed U and then average. Its uniformity gives the same summed error bound. The observations share U after averaging and are not unconditionally independent. In particular, the zero unconditional mean of Y^3X under the symmetric prior must not be substituted into the iid variance formula (104). Its conditional score mean is A_p^2 for every teacher.

Corollary G.3. *Let* $p = 3$, $s = D^{1/8}$ *and* $N_D = \lceil D^{21/16} \rceil$. *The test* (93) *has summed errors at most*

$$\frac{16K_3^{\text{det}}}{A_3^2} D^{-9/16} + \frac{32(K_3^{\text{det}})^2}{A_3^4} D^{-1/8}.$$

For all sufficiently large D, *an unrestricted measurable predictor has worst-teacher expected risk at most* $D^{-1/64}$, *while*

$$\mathcal{R}\left(D, N_D, \lfloor D^{1/1792} \rfloor, 3, D^{1/8}\right) \geq 1 - 2D^{-1/64}.$$

Proof. Substitute $a = 1/8$ and $r = 21/16$ in (105); the two exponents become $-9/16$ and $-1/8$. Since $r - (1 + 2a) = 1/16 > 0$, the statistical conclusion follows from Corollary D.5 with $\gamma = (r - 1 - 2a)/4 = 1/64$.

For the displayed prediction constants, apply Corollary A.3 with the smaller margin $\eta = 1/16$. Then $b = 3/2$, $\tau = 1/32$ and $3p + 5 = 14$. For $D^{1/8} \geq 2$,

$$N_D \leq 2D^{21/16} \leq D^{23/16} = D^{b-\eta}.$$

The label-degree exponent is $(1/32)/(4 \cdot 14) = 1/1792$ and the risk exponent is $\tau/2 = 1/64$. The dimension threshold of that corollary and the finite-prior-to-minimax argument in Corollary G.2 prove the result. □

These are asymptotic statements with fixed p, a, r. The conservative degree guarantee in Corollary G.3 includes degree six only when $D \geq 6^{1792}$. The detector's sufficient exponent is not asserted to be optimal. The prediction lower concerns the specified label-polynomial class, and the statistical upper does not assert efficient computation.

H Gradient iterations and label degree

The prediction class includes a specified number of unnormalized gradient steps for a Hermite ridge. The next proposition gives the precise degree bound. We obtain it by tracking degrees under polynomial composition.

Proposition H.1. *Fix an odd integer $p \geq 3$ and integers $D, N \geq 1$. Let R be learner randomness independent of the observations and teacher. Let the initial vector $w_0 \in \mathbb{R}^D$, the finite real step sizes γ_j, and the nonempty index sets $I_j \subseteq \{1, \dots, N\}$ be jointly measurable functions of the design $(X_1, \dots, X_N)$, the supplied model parameters, and R, with no dependence on the labels. Define, without normalization or projection,*

$$w_{j+1} = w_j - \frac{\gamma_j}{|I_j|} \sum_{i \in I_j} \big(h_p(w_j^\top X_i) - Y_i\big) h_p'(w_j^\top X_i) X_i, \qquad j \geq 0.$$

For every integer $t \geq 1$, each coordinate of w_t is a polynomial in the training labels of total degree at most $(2p-1)^{t-1}$. The predictor $\widehat{f}_t(x) = h_p(w_t^\top x)$ therefore belongs to $\mathcal{P}_{p(2p-1)^{t-1}}$.

In the notation and parameter regime of Theorem 2.1, suppose also that $1 \leq N \leq D^{b-\eta}$ and D is sufficiently large. For every integer

$$1 \leq t \leq 1 + \left\lfloor \frac{\log(L_-/p)}{\log(2p-1)} \right\rfloor,$$

one has

$$\sup_{\|u\|=1} \mathbb{E}_u\big[(\widehat{f}_t(x) - h_p(u^\top x))^2\big] \geq 1 - 2D^{-\tau/2}.$$

The expectation includes the independent fresh input and learner randomness. The displayed iteration range is used only once $L_- \geq p$.

Proof. Condition on the design and R. The initialization, step sizes and index sets are then fixed independently of the label variables. The first update is affine in those variables, so its degree is at most one. If each coordinate of w_j has degree at most $Q \geq 1$, then $h_p(w_j^\top X_i)$ and $h_p'(w_j^\top X_i)$ have degrees at most pQ and $(p-1)Q$, respectively. The two terms in their product with the residual have degrees at most $(2p-1)Q$ and $1 + (p-1)Q$. Both the latter degree and the degree of w_j are at most $(2p-1)Q$. Taking sums and multiplying by label-independent coefficients do not increase this bound. Induction with $Q = (2p-1)^{j-1}$ proves the asserted coordinate bound.

Composing with h_p at the fresh input multiplies the degree bound by at most p. Each coefficient is jointly measurable in the full design, query and seed because only finite sums, products and compositions of the stated measurable functions are used. Thus this is an allowed predictor, whether or not it has finite expected risk. The polynomial formula does not require w_t to have unit norm.

For the stated iteration range, $p(2p-1)^{t-1} \leq L_-$. The lower in Theorem 2.1 applies to every member of $\mathcal{P}_{L_-}$ and gives the displayed worst-teacher risk bound. An infinite risk also satisfies that bound. Since the exponent defining L_- is positive, $L_- \geq p$ for all sufficiently large D. □

At fixed p, a, η, the allowed iteration count is proportional to $\log D$. The proposition does not cover polynomially many iterations, label-dependent stopping or step sizes, normalization, or projection onto the sphere. Such operations need not preserve the displayed label degree. No conclusion about their prediction risk follows from this degree calculation alone.

I Lower bounds for links of fixed parity

The lower bound extends to any fixed mixture of Hermite orders of the same parity. The least order determines the lower-bound exponent, while the largest order enters the allowed polynomial degree.

Fix integers $3 \leq p \leq P$ of the same parity and a known polynomial

$$f(z) = \sum_{q \in I} c_q h_q(z), \qquad I \subseteq \{p, p+2, \dots, P\}, \qquad c_p \neq 0, \qquad \sum_{q \in I} c_q^2 = 1, \tag{107}$$

where I contains precisely the nonzero coefficients. All these parameters are fixed independently of dimension and noise. No nondegeneracy assumption on a transformed-label moment is needed here. Let $D \geq 1$ and $N, L \geq 0$ be integers, and let $s > 0$. As in (12), take a sign-prior teacher $U = \epsilon/\sqrt{D}$, an independent standard Gaussian design $A \in \mathbb{R}^{N\times D}$, and independent $G \sim N(0, I_N)$. Observe

$$V_i = T_i(A, U) + G_i, \qquad T_i(A, U) = f(A_i^{\mathsf{T}} U)/s. \tag{108}$$

The target at an independent $x \sim \gamma_D$ is $F_U(x) = f(U^{\mathsf{T}} x)$. Allowed predictors are

$$\widehat{F}(A, V, \xi; x) = \sum_{|\alpha|\leq L} a_\alpha(A, \xi, x) V^\alpha, \tag{109}$$

with jointly measurable real coefficients and independent learner randomness ξ. The coefficients may depend arbitrarily on the entire design and query. No teacher-dependent side information is supplied. Scaling between $Y = sV$ and V preserves this class.

Put $\mathcal{H} = \bigoplus_{q\in I} \mathcal{H}_q$. Define, at fixed design,

$$\begin{aligned} m_\alpha &= \mathbb{E}_U T^\alpha, & v_\alpha &= \mathbb{E}_U[F_U T^\alpha], \\ \kappa_\alpha &= v_\alpha - \sum_{\substack{\beta\leq\alpha \\ \beta\neq\alpha}} \binom{\alpha}{\beta} \kappa_\beta m_{\alpha-\beta}, & E_k &= \mathbb{E}_A \sum_{|\alpha|=k} \frac{\|\kappa_\alpha(A)\|_{\mathcal{H}}^2}{\alpha!}. \end{aligned} \tag{110}$$

Define

$$\varepsilon_{p,P}(D) = \mathbf{1}_{\{p \text{ even}\}} (P-1)!! D^{-p/2}. \tag{111}$$

Theorem I.1. *In the experiment* (108)*, every predictor* (109) *satisfies*

$$\mathbb{E}[(\widehat{F}(A, V, \xi; x) - F_U(x))^2] \geq 1 - \sum_{k=0}^{L} E_k. \tag{112}$$

All energies are finite, and $E_0 \leq \varepsilon_{p,P}(D)$*. If* p *is odd, the energies vanish for even* k*, including* $k = 0$*. For every* $k \geq 1$*,*

$$\begin{aligned} E_k &\leq (2Pk)^{\Gamma_P k} s^{-2k} \sum_{r=1}^{\min\{N,k\}} N^r D^{-\delta_p(k,r)}, \\ \Gamma_P &= 3P + 6, \qquad \delta_p(k, r) = \max\left\{ r, (p-1)r - \frac{p-2}{2}k + \frac{p}{2} \right\}. \end{aligned} \tag{113}$$

An empty sum is zero. These statements hold uniformly over the coefficients in (107) *and all the displayed finite parameters.*

I.1 The direct-sum certificate

The Hermite identities in (18), applied at each order q, and orthogonality of distinct chaoses give

$$\|F_u\|_2^2 = 1, \qquad \langle F_u, F_w \rangle = \sum_{q\in I} c_q^2 (u^{\mathsf{T}} w)^q. \tag{114}$$

If p is odd, sign symmetry also gives $\mathbb{E}_U F_U = 0$. We check the certificate for the direct sum, including merely measurable coefficients and the possibly nonzero κ_0.

If the risk of (109) is finite, then $\mathbb{E}\|\widehat{F}\|_2^2 \leq 2\mathbb{E}\|\widehat{F} - F_U\|_2^2 + 2 < \infty$. At every fixed design a, the label law is a finite mixture of identity-covariance Gaussian laws. It has a strictly positive density and all polynomial moments. Its monomial Gram matrix

$$J_a(\alpha, \beta) = \mathbb{E}[V^{\alpha+\beta} \mid A = a], \qquad |\alpha|, |\beta| \leq L,$$

is positive definite, since a nonzero polynomial cannot vanish Lebesgue almost everywhere. If $\lambda_a > 0$ is its least eigenvalue, Tonelli's theorem gives, for almost every (a, ξ),

$$\mathbb{E}[\|\widehat{F}(a, V, \xi; \cdot)\|_2^2 \mid a, \xi] \geq \lambda_a \sum_\alpha \|a_\alpha(a, \xi, \cdot)\|_2^2.$$

Thus every coefficient is in $L^2(\gamma_D)$. Project each coefficient onto the finite-dimensional space $\mathcal{H}$. This preserves label degree and cannot increase risk, since every target belongs to $\mathcal{H}$. The projection is measurable by integration against the finite product-Hermite basis; Cauchy–Schwarz makes these integrals finite. Set the projected coefficients to zero on the exceptional design and seed null set.

Fix a, ξ and write the resulting polynomial as

$$\widehat{F}(y) = \sum_{|\alpha| \le L} b_\alpha q_\alpha(y), \qquad q_\alpha(y) = \prod_i \frac{\mathrm{He}_{\alpha_i}(y_i)}{\sqrt{\alpha_i!}}, \qquad b_\alpha \in \mathcal{H}.$$

The shift identity (19) yields

$$\mathbb{E}\langle F_U, \widehat{F}(T+G)\rangle = \sum_\alpha \langle z_\alpha, b_\alpha\rangle, \qquad z_\alpha = v_\alpha/\sqrt{\alpha!}.$$

Jensen's inequality over U at fixed G, followed by Gaussian orthogonality, yields

$$\mathbb{E}\|\widehat{F}(T+G)\|^2 \ge \sum_\beta \|d_\beta\|^2, \qquad d = Mb, \qquad M_{\beta,\alpha} = \mathbf{1}_{\{\beta\le\alpha\}} m_{\alpha-\beta}\sqrt{\frac{\beta!}{\alpha!}}\binom{\alpha}{\beta}.$$

The recursion (110) says $z = M^\mathsf{T} w$ for $w_\beta = \kappa_\beta/\sqrt{\beta!}$, including $\beta = 0$. Completing squares and using (114) therefore gives

$$\mathbb{E}\|\widehat{F}(T+G) - F_U\|^2 \ge 1 + \sum_\beta \|d_\beta - w_\beta\|^2 - \sum_\beta \|w_\beta\|^2 \ge 1 - \sum_{|\alpha|\le L} \frac{\|\kappa_\alpha\|^2}{\alpha!}.$$

Each basis coordinate of $\kappa_\alpha(a)$ is a polynomial in a: the sign prior is finite, f is a polynomial, and the recursion is finite. Thus every energy has finite Gaussian-design expectation. Averaging proves (112), and the inequality is automatic for infinite-risk predictors. For independent sign-prior teachers U, W,

$$E_0 = \|\mathbb{E}_U F_U\|_2^2 = \sum_{q\in I} c_q^2 \mathbb{E}\left(D^{-1}\sum_{j=1}^D \epsilon_j\right)^q,$$

where the ϵ_j are iid uniform signs. For odd p, every moment in this sum vanishes, so $\kappa_0 = 0$. For even q, each surviving coordinate assignment in the expanded moment has even multiplicities and admits a pairing of its q positions. There are $(q-1)!!$ pairings and at most $D^{q/2}$ assignments per pairing. Thus the normalized moment is at most $(q-1)!!D^{-q/2}$. When p is even, all $q \in I$ are even and $p \le q \le P$, giving $E_0 \le (P-1)!!D^{-p/2}$ after summing c_q^2. This proves $E_0 \le \varepsilon_{p,P}(D)$ in both cases. If $N = 0$ or $L = 0$, only E_0 remains in the certificate. The risk bound is $1 - E_0$, attained by the prior-mean predictor $\mathbb{E}_U F_U$; for odd p this is the zero predictor and the bound is one.

I.2 Variable-order tensor expansion

Use the symmetric-tensor isometries J_q of Section A.3, with the order in (22) replaced by q. Distinct orders are orthogonal, so the direct-sum representation is

$$F_U = \bigoplus_{q\in I} c_q J_q(U^{\otimes q}). \tag{115}$$

These identities follow from the same Gaussian generating function: two normal-ordered factors of unequal orders have zero expectation, while equal orders have all cross matchings. Products of more factors pair their coordinate positions with no pair internal to one factor. Independence of the design rows forces each pair to stay in its row. Repeated positions retain their derivative multiplicities.

The cumulant partition formula (20) applies to the expansion of (110) by multilinearity. Every factor is a sign character; for odd p its support is nonempty and odd. For even p, empty support is possible. Its derivation by the target-weighted generating-function ratio in Section A.2 also applies coordinatewise to the finite-dimensional target (115). At positive cumulant order $k \ge 1$, a constant argument makes the cumulant zero: its exponential factors out of the joint moment generating function, adding a separate linear term to its logarithm. We delete these terms. All remaining reduced

characters therefore have nonempty support. In particular, a nonzero cumulant requires empty total symmetric difference and a connected character intersection graph, and its absolute value is at most $(k+1)^{k+1}$ by (21). If p is odd, the total symmetric difference of $k+1$ odd supports has the parity of $k+1$. Thus, in the odd case, every even-order κ_α vanishes.

Fix $k \geq 1$ and a multiindex α of total order k and support size r. Its positive row multiplicities sum to k. Expand its tensor cumulants and square their norms. Orthogonality in (115) pairs only equal target orders q. Fix such an order. The two replicas have $2k$ label factors, whose orders $j_1, \ldots, j_{2k} \in I$ may differ. Put

$$R = \sum_{v=1}^{2k} j_v, \qquad e = R/2 + q.$$

A row of multiplicity m contains $2m$ factors of the same parity, so its total number of positions is even. There are $R/2$ Gaussian pairs. The tensor inner product identifies corresponding target positions in order, adding q target bridges with no further matching factor. The exact scalar factor for these choices of orders is

$$s^{-2k} D^{-e} c_q^2 \prod_{v=1}^{2k} \frac{c_{j_v}}{\sqrt{j_v!}}. \tag{116}$$

An order-j label contributes $D^{-j/2}/(s\sqrt{j!})$; each order-q target tensor contributes $D^{-q/2}$. The tensor isometry removes the target factorial.

Give one initial coordinate variable to every Gaussian pair and target bridge. There are exactly $e \leq P(k+1) \leq 2Pk$ such variables. Their values in $[D]$ induce a partition into h equality classes. Each replica contributes the cumulant of its reduced characters. Remove terms whose cumulant vanishes in either replica.

I.3 Coordinate loss

For a remaining term form a bipartite multigraph B. Its left vertices are the r row groups and the target; its right vertices are the h coordinate classes. Each initial variable supplies one edge between its row or target and its coordinate class. Parallel edges remain distinct. Write $\ell = e - h$.

Lemma I.2. *Every remaining term satisfies $\ell \geq \delta_p(k, r)$.*

Proof. The graph B is connected. The nonzero cumulant in one replica has a connected character intersection graph. Every shared coordinate gives a path through its coordinate vertex in B. Identifying factors in the same row preserves connection among all row groups and the target. A coordinate class absent from the reduced characters still attaches to one of these groups. Consequently

$$\beta(B) = e - (r + 1 + h) + 1 = \ell - r \geq 0. \tag{117}$$

Mark the Gaussian pairs crossing replicas and all q target bridges. Remove isolated vertices to form a subgraph M. Every coordinate vertex in M has positive even degree. Indeed, its number of positions in the left replica is congruent to its marked degree modulo two: left-internal pairs contribute two, right-internal pairs zero, and crossing pairs or target bridges one. Empty total symmetric difference in that replica forces evenness. Such a vertex therefore has degree at least two.

The target has marked degree $q \geq p$. Let r_1, r_o, r_e count singleton rows, other odd-multiplicity rows, and even-multiplicity rows. A singleton has only one normal-ordered factor in each replica. Neither can pair within itself, so all pairs in that row cross replicas. A surviving pairing forces their orders to agree, and the marked degree is that common order, at least p. First suppose p is odd. In any row of multiplicity m, the sum of the left-replica orders has parity m. Removing internal pairs removes an even number of positions. The number of crossing pairs has parity m, regardless of the individual orders. Thus another odd row has marked degree at least one, and an even row is isolated or has degree at least two.

The target ensures that the number c_M of components of M is at least one. The degree-sum identity gives

$$\beta(M) = c_M + \frac{1}{2}\sum_{v\in M}(\deg_M(v) - 2)$$
$$\geq 1 + \frac{(p-2)(r_1+1) - r_o}{2} = \frac{(p-2)r_1 - r_o + p}{2}. \tag{118}$$

Coordinate vertices and nonisolated even rows contribute nonnegatively. The target and each singleton contribute at least $p-2$, while every other odd row contributes at least -1. The bound holds even if its right side is negative.

Cycle rank is monotone under taking subgraphs, including multigraphs, because independent cycle vectors in a subgraph remain independent in the full cycle space. Thus $\beta(B) \geq \beta(M)$. Moreover,

$$k \geq r_1 + 3r_o + 2r_e = 2r - r_1 + r_o,$$

and $p \geq 3$ gives

$$(p-2)r_1 - r_o = (p-2)(r_1 - r_o) + (p-3)r_o \geq (p-2)(2r-k).$$

Equations (117) and (118) prove both branches of $\ell \geq \delta_p(k,r)$.

Now suppose p is even. All factor orders in a row are even. Removing internal pairs removes two positions at a time, so each row has even marked degree. Every nonisolated nonsingleton therefore has degree at least two. The target and each singleton have degree at least p, and coordinate vertices have degree at least two. The target ensures at least one marked component, so the degree-sum identity gives

$$\beta(M) \geq 1 + \frac{(p-2)(r_1+1)}{2} = \frac{(p-2)r_1 + p}{2}.$$

Cycle-rank monotonicity still gives $\beta(B) \geq \beta(M)$. Since $k \geq r_1 + 2(r-r_1)$, we have $r_1 \geq 2r-k$. Together with (117), this gives $\ell \geq (p-1)r - (p-2)k/2 + p/2$ and $\ell \geq r$, proving the same loss bound for even p. □

The proof only uses necessary conditions for a nonzero cumulant. It does not require every connected pairing diagram to contribute.

I.4 Counting and the exponent

Completion of the proof of Theorem I.1. The certificate, finiteness, and parity were proved above. Fix $k \geq 1$ and a multiindex of support size r. At fixed target and label orders, there are at most $(2Pk)^{Pk}$ Gaussian pairings: the $R \leq 2Pk$ positions admit at most $R^{R/2}$ successive partner choices before row restrictions are imposed. There are at most $e^e \leq (2Pk)^{2Pk}$ partitions of the initial coordinate variables, since sending a variable to the least-indexed member of its block injects these partitions into e-valued maps. A partition with h blocks has $(D)_h \leq D^h$ assignments of distinct coordinates, and none if $h > D$. Lemma I.2 gives

$$D^{-e}(D)_h \leq D^{-\delta_p(k,r)}.$$

The product of the two cumulants has absolute value at most

$$(k+1)^{2(k+1)} \leq (2k)^{4k} \leq (2Pk)^{4k}.$$

Drop the factorial denominators in (116). Summing the absolute label coefficients over their orders gives

$$\sum_{j_1,\dots,j_{2k}\in I}\prod_v |c_{j_v}| = \left(\sum_{j\in I}|c_j|\right)^{2k} \leq |I|^k \leq P^k \leq (2Pk)^k,$$

by Cauchy–Schwarz and the normalization in (107). The target coefficients sum as $\sum_q c_q^2 = 1$. There are no cross terms between distinct target chaoses. Taking absolute values termwise therefore proves

$$\mathbb{E}_A\|\kappa_\alpha(A)\|^2 \leq (2Pk)^{(3P+5)k} s^{-2k} D^{-\delta_p(k,r)}.$$

There are $\binom{N}{r}\binom{k-1}{r-1} \le N^r 2^{k-1}$ multiindices of order k and support size r. Drop $1/\alpha! \le 1$, absorb $2^{k-1} \le (2Pk)^k$, and sum over r to obtain (113). All expansions are finite. They include repeated rows and coordinates, unequal label orders across replicas, and $k > D$. For $N = 0$ the positive-order sums are empty. □

Corollary I.3. *Fix the link* (107) *and* $a, \eta > 0$. *Put*

$$b = \min\{1 + 4a, p/2 + 2a\}, \qquad \tau = \min\{2a, \eta/2\}, \qquad L = \left\lfloor D^{\tau/(4\Gamma_P)} \right\rfloor.$$

If $s = D^a$, $N \le D^{b-\eta}$, *and*

$$D \ge \left\lceil \max\{2, (4P)^{4\Gamma_P/\tau}, 2^{2/\tau}\} \right\rceil,$$

then every predictor (109) *has sign-prior risk at least* $1 - \varepsilon_{p,P}(D) - 2D^{-\tau/2}$. *The same lower bound holds for its worst-teacher risk, and hence for minimax risk in this class. It also holds for joint polynomials of total degree at most* L *in all training inputs, labels, and the query.*

Proof. The case $N = 0$ follows from Theorem I.1, so assume $N > 0$. Each term in (113) has dimension exponent at most

$$e_{k,r} = (b - \eta)r - 2ak - \delta_p(k, r), \qquad 1 \le r \le k.$$

Write $q_0 = p - 2$ and $c = b - \eta - 1$. If $c < 0$, the first branch of δ_p gives $e_{k,r} \le cr - 2ak \le -\tau k$. If $0 \le c \le q_0$, take a convex combination of its two branches, with weight c/q_0 on the second:

$$\delta_p(k, r) \ge (1 + c)r - \frac{c}{2}k + \frac{cp}{2q_0}.$$

Since $c \le 4a - \eta$, this implies

$$e_{k,r} \le -(2a - c/2)k - \frac{cp}{2q_0} \le -\eta k/2 \le -\tau k.$$

Finally, if $c > q_0$, then $a > q_0/4$, since otherwise $c \le 4a - \eta < q_0$. Thus $b = p/2 + 2a$. Put $t = 2a - q_0/2$; the inequality $c = q_0 + t - \eta > q_0$ gives $t > \eta$. The second branch of δ_p gives

$$e_{k,r} \le (t - \eta)r - tk - p/2 = -t(k - r) - \eta r - p/2 \le -\eta k - p/2 \le -\tau k.$$

This covers all cases. There are at most k terms in the sum over r, so

$$E_k \le k(2Pk)^{\Gamma_P k} D^{-\tau k} \le (4Pk)^{\Gamma_P k} D^{-\tau k}.$$

For $k \le L$ and the stated lower bound on D,

$$(4Pk)^{\Gamma_P} \le (4P)^{\Gamma_P} D^{\tau/4} \le D^{\tau/2}.$$

Hence $E_k \le D^{-\tau k/2}$. The bound $D^{-\tau/2} \le 1/2$ and a geometric sum give $\sum_{k=1}^{L} E_k \le 2D^{-\tau/2}$. Adding $E_0 \le \varepsilon_{p,P}(D)$ and applying (112) proves the bound, which is exactly $1 - 2D^{-\tau/2}$ when p is odd. For each fixed learner its sign-prior risk is at most its worst-teacher risk, so the latter bound persists after taking the infimum over learners. Finally, a joint polynomial of total degree at most L has label degree at most L, with measurable design and query coefficients, and is included in (109). □

J Upper bounds for links of fixed parity

Fix integers $3 \le p \le P$ of the same parity and a known polynomial

$$f(z) = \sum_{q \in I} c_q h_q(z), \qquad I \subseteq \{p, p + 2, \dots, P\}, \qquad c_p \ne 0, \qquad \sum_{q \in I} c_q^2 = 1, \tag{119}$$

where I consists of the orders with nonzero coefficients. Write $K = |I|$. The link and its coefficients are fixed independently of dimension and noise. The observations and target are

$$X_i \sim N(0, I_D), \qquad Y_i = f(u^\top X_i) + sG_i, \qquad f_u(x) = f(u^\top x), \qquad \|u\| = 1, \tag{120}$$

with independent Gaussian design rows, independent standard Gaussian noises, and an independent standard Gaussian query x. Assume

$$\beta = \mathbb{E}[(Z^2-1)f(Z)^2] \neq 0, \qquad Z \sim N(0,1).$$

The sign of β is unrestricted. Define

$$\begin{aligned} C_f &= \max_{0\le j\le P} \mathbb{E}[f(Z)^2 h_j(Z)^2], \\ \Delta_0 &= C_f + s^2, \qquad \Delta_p = \Delta_0/c_p^2, \\ G_f &= 1 + \beta^{-2} \max_{w\in\{1,z^2,(z^2-1)^2\}} \mathbb{E}[(f(Z)^4 + 2f(Z)^2 + 3)w(Z)]. \end{aligned} \tag{121}$$

These are finite Gaussian polynomial moments; $C_f, G_f \ge 1$. Their values depend on the known link, and C_f, G_f do not depend on dimension or noise.

Theorem J.1. *Let the fixed link satisfy (119) and $\beta \neq 0$. Fix an integer $D \ge 2$, $s \ge 1$, and $0 < \epsilon \le 1$. Put*

$$r = \epsilon/P, \qquad t = \lceil \log_2(100D/r) \rceil, \quad m = \left\lceil 2^{P+1} K \Delta_0/\epsilon \right\rceil, \tag{122}$$

$$\begin{aligned} n_Q &= \left\lceil 100 G_f D s^4 t/r \right\rceil, \\ n_T &= \left\lceil 400\Delta_p D t/r + 20\Delta_p D^{p/2}\sqrt{t/r} \right\rceil, \\ N_Q &= P t n_Q + m, \qquad N_T = 2Ptn_T + m. \end{aligned}$$

There are predictors using N_Q and N_T observations, respectively, with worst-teacher expected squared prediction risk at most ϵ. Both have label degree at most $2Pt+1$. Their total degrees in the training design, labels, and query are at most

$$4Pt + 2P + 1, \qquad 2P(p+1)t + 2P + 1,$$

respectively. For fixed f, they have polynomial real-arithmetic work and storage. Choosing the smaller deterministic sample count gives

$$N = O_f\left(\min\left\{ Ds^4\epsilon^{-1}\log^2(D/\epsilon),\ s^2\left[D\epsilon^{-1}\log^2(D/\epsilon) + D^{p/2}\epsilon^{-1/2}\log^{3/2}(D/\epsilon)\right]\right\}\right). \tag{123}$$

In particular, fix $a, \eta > 0$ and put $b = \min\{1+4a, p/2+2a\}$ and $s = D^a$. For all sufficiently large D, every integer $N \ge \lceil D^{b+\eta} \rceil$ permits expected risk at most $D^{-\eta/4}$ with joint degree at most

$$2P(p+1)\left\lceil \log_2(100PD^{1+\eta/4}) \right\rceil + 2P + 1.$$

Extra observations may be ignored.

J.1 Covariance at each output order

For each positive integer q, let R_q, Π_q, and J_q be the Hermite tensor, symmetric-tensor projection, and isometry defined as in (35), with order q. For a block of $n \ge 1$ observations define

$$C_{q,n} = \frac{1}{\sqrt{q!}\,n}\sum_{i=1}^{n} Y_i R_q(X_i).$$

Lemma J.2. *For $q \in I$,*

$$\mathbb{E}C_{q,n} = c_q u^{\otimes q}, \qquad \mathrm{Cov}(C_{q,n}) \preceq \frac{\Delta_0}{n}\Pi_q.$$

In particular, $C_{p,n}/c_p$ has mean $u^{\otimes p}$ and covariance at most $(\Delta_p/n)I$ on ordered tensors.

Proof. Taking derivatives in the Gaussian exponential identity

$$\mathbb{E}e^{v^\top X - \|v\|^2/2} e^{w^\top X - \|w\|^2/2} = e^{v^\top w}$$

gives the isometry $\mathbb{E}[J_q(B)J_q(H)] = \langle \Pi_q B, \Pi_q H\rangle$. Different Hermite orders are orthogonal. Differentiating along u also gives $J_q(u^{\otimes q})(X) = h_q(u^\top X)$. The expansion of f and the centered independent noise therefore give $\mathbb{E}[Y R_q(X)]/\sqrt{q!} = c_q u^{\otimes q}$.

For an order-q homogeneous chaos function g, rotate u to the first coordinate and write $X = (Z, W)$. Its product-Hermite expansion is

$$g(Z, W) = \sum_{j=0}^{q} h_j(Z) g_{q-j}(W).$$

Distinct orders in W are orthogonal. Integrating W removes cross terms even after multiplication by $f(Z)^2$, and yields

$$\mathbb{E}[f(u^\top X)^2 g(X)^2] = \sum_{j=0}^{q} \mathbb{E}[f(Z)^2 h_j(Z)^2] \|g_{q-j}\|_2^2 \le C_f \|g\|_2^2.$$

Consequently, for every deterministic ordered tensor B,

$$\mathrm{Var}\langle C_{q,n}, B\rangle \le \frac{1}{n}\mathbb{E}[Y^2 J_q(B)(X)^2] \le \frac{\Delta_0}{n}\|\Pi_q B\|^2.$$

The noise contributes $s^2\|J_q(B)\|_2^2$ and has zero cross term. Independence of observations divides the variance by n. This proves the covariance bound; division by the known nonzero c_p gives the final assertion. All moments are finite polynomial Gaussian moments. □

J.2 Matrix blocks

Write $U = uu^\top$, $Q = I - U$, and $d = D - 1$, distinguishing the projector U from the largest allowed Hermite order P. The quadratic block is

$$M_Q = \frac{1}{\beta n_Q}\sum_{i=1}^{n_Q}(Y_i^2 - 1 - s^2)(X_i X_i^\top - I).$$

For the second block, use two independent copies of $C_{p,n_T}/c_p$, flattened to D by L matrices with $L = D^{p-1}$, and put

$$C_j = uv^\top + E_j, \qquad v = u^{\otimes(p-1)}, \qquad M_T = \tfrac{1}{2}(C_1 C_2^\top + C_2 C_1^\top).$$

Every block uses fresh observations.

Lemma J.3. *Both blocks satisfy the hypotheses of Lemma B.3 with the parameters r, t in (122). Their respective values of γ are*

$$\gamma_Q = G_f s^4/n_Q, \qquad \gamma_T = 2\kappa + \kappa^2 L, \qquad \kappa = \Delta_p/n_T.$$

Proof. For the quadratic block write $X = (Z, V)$ along u, and set $H = f(Z)$ and $W_0 = Y^2 - 1 - s^2$. Then $\mathbb{E}[W_0 \mid Z] = H^2 - 1$. The parallel entry has mean $\beta/\beta = 1$; independence and the zero mean of V remove off-diagonal means, and $\mathbb{E}(VV^\top - I_d) = 0$ removes the perpendicular mean. Thus $\mathbb{E}M_Q = U$. Also

$$\mathbb{E}_G W_0^2 = (H^2 - 1)^2 + 4s^2 H^2 + 2s^4 \le s^4(H^4 + 2H^2 + 3).$$

Here the first term can be multiplied by $s^4 \ge 1$, and $s^2 \le s^4$ bounds the second. Using the notation A, B, C of Lemma B.3, independence gives

$$A = 1 + \frac{\mathbb{E}[W_0^2(Z^2 - 1)^2] - \beta^2}{\beta^2 n_Q},$$

$$B = \frac{d\,\mathbb{E}[W_0^2 Z^2]}{\beta^2 n_Q}, \qquad C = \frac{D\,\mathbb{E}W_0^2}{\beta^2 n_Q}.$$

The last identity uses $\mathbb{E}\|VV^\top - I_d\|_F^2 = d(d+1) = dD$, which follows by expanding the norm and using $\mathbb{E}\|V\|^2 = d$ and $\mathbb{E}\|V\|^4 = d(d+2)$. Thus $A \le 1 + \gamma_Q$, $B \le d\gamma_Q$, and $C \le D\gamma_Q$. The chosen n_Q gives $D\gamma_Q \le r/(100t)$.

For the tensor block, Lemma J.2 makes the errors E_1, E_2 centered and independent with covariance at most κI, so $\mathbb{E}M_T = U$. Testing this covariance and summing over perpendicular coordinates or columns gives, for deterministic z,

$$\mathbb{E}\|QE_j z\|^2 \le \kappa d\|z\|^2, \qquad \mathbb{E}\|E_j^\top Q\|_F^2 \le \kappa dL, \qquad \mathrm{Cov}(E_j^\top u) \preceq \kappa I_L.$$

Expand

$$C_1 C_2^\top = U + u(E_2 v)^\top + (E_1 v)u^\top + E_1 E_2^\top.$$

The parallel entry has two linear error terms of variance at most κ and a bilinear term of second moment at most $\kappa^2 L$. Their cross expectations vanish by conditioning on one centered independent error. Symmetrization preserves scalar quadratic forms, giving $A - 1 \le 2\kappa + \kappa^2 L$.

The perpendicular column is $QE_1 v + QE_1(E_2^\top u)$. Its cross expectation vanishes by centering of E_2. Conditioning on E_2 bounds its second moment by $\kappa d + \kappa d\,\mathbb{E}\|E_2^\top u\|^2 \le d(\kappa + \kappa^2 L)$. The transpose has the same bound; convexity of squared norm therefore gives this bound for B. For the perpendicular block, condition on E_2 and apply the first covariance inequality to every column of $E_2^\top Q$:

$$\mathbb{E}\|QE_1 E_2^\top Q\|_F^2 \le \kappa d\,\mathbb{E}\|E_2^\top Q\|_F^2 \le \kappa^2 d^2 L.$$

Convexity under symmetrization gives $C \le \kappa^2 dL$. These inequalities imply the required moment bounds with γ_T. The chosen n_T gives

$$D\gamma_T = \frac{2\Delta_p D}{n_T} + \frac{\Delta_p^2 D^p}{n_T^2} \le \frac{r}{200t} + \frac{r}{400t} \le \frac{r}{100t}.$$

Both matrices are symmetric. If an orthogonal matrix O fixes u, rotating every input by O preserves its law and the labels. The quadratic block transforms to $OM_Q O^\top$. The generating function gives $R_p(OX) = O^{\otimes p} R_p(X)$, so the flattened tensors transform to $OC_j(O^{\otimes(p-1)})^\top$ and their Gram block transforms to $OM_T O^\top$. Thus both laws have the required invariance. The displayed estimates establish finite second moments, completing every hypothesis of Lemma B.3. □

J.3 Output and degree bounds

Proof of Theorem J.1. For either block type, construct P independent products $\mathcal{B}^{(1)}, \dots, \mathcal{B}^{(P)}$ of t blocks each, using disjoint observations. Lemmas J.3 and B.3 give

$$\mathbb{E}\mathcal{B}^{(j)} = U, \qquad K_0 := \mathbb{E}(\mathcal{B}^{(j)})^\top \mathcal{B}^{(j)} = a_t U + b_t Q, \qquad a_t \le e^{r/100}, \qquad \mathrm{tr}\, K_0 < 2.$$

Take an independent final block of m observations and form all $C_{q,m}, q \in I$, from this block. Define

$$T_q = (\mathcal{B}^{(1)} \otimes \cdots \otimes \mathcal{B}^{(q)})C_{q,m}, \qquad \widehat{f}(x) = \sum_{q \in I} J_q(\Pi_q T_q)(x).$$

Independence gives $\mathbb{E}T_q = c_q u^{\otimes q}$. The Hermite isometry and contraction of Π_q bound the order-q prediction risk by $\mathbb{E}\|T_q - c_q u^{\otimes q}\|^2$. For a deterministic tensor B, independence of the q products gives

$$\mathbb{E}\|(\mathcal{B}^{(1)} \otimes \cdots \otimes \mathcal{B}^{(q)})B\|^2 = \langle B, K_0^{\otimes q} B\rangle.$$

Using Lemma J.2 for the final tensor gives

$$\mathbb{E}\|T_q\|^2 \le c_q^2 a_t^q + \frac{\Delta_0}{m}\,\mathrm{tr}(K_0^{\otimes q}\Pi_q) \le c_q^2 a_t^q + \frac{\Delta_0}{m}(\mathrm{tr}\, K_0)^q.$$

For the trace inequality, diagonalize the projection Π_q. The omitted diagonal entries of the positive semidefinite $K_0^{\otimes q}$ are nonnegative. Subtracting the squared mean norm bounds the order-q risk by $c_q^2(a_t^q - 1) + (\Delta_0/m)(\mathrm{tr}\, K_0)^q$. These estimates also establish finiteness of all expectations.

For each realized training sample, the errors at different orders belong to orthogonal query chaoses. Therefore their risks add exactly, even though the same products and final observations are used at several orders. The total expected risk is at most

$$\sum_{q \in I} c_q^2(e^{qr/100} - 1) + \frac{K\Delta_0\, 2^P}{m} \le e^{\epsilon/100} - 1 + \epsilon/2 \le \epsilon/99 + \epsilon/2 < \epsilon.$$

The penultimate inequality follows from $e^z \le (1-z)^{-1}$ for $0 \le z < 1$. All estimates are uniform over unit u. For odd p, the final tensors supply the signed target even though the intermediate matrix mean is $uu^\top$. For even p, $F_u = F_{-u}$, so sign ambiguity does not change the target.

There are Pt matrix blocks and one final block, giving exactly N_Q or N_T observations. Each quadratic block has label degree two and joint training degree four. Each empirical order-p tensor has joint training degree $p+1$; its Gram block has degree $2(p+1)$ and label degree two. The order-q output uses qt matrix factors, a final tensor of joint training degree $q+1$, and a query tensor of degree q. Its total degree is therefore at most $4qt + 2q + 1$ or $2q(p+1)t + 2q + 1$, and its label degree is at most $2qt + 1$. Taking $q \le P$ proves the asserted degree bounds. Symmetrization and query evaluation are fixed linear maps. All divisions and the choice of branch depend only on known parameters; no data-dependent normalization or eigenvector selection is used.

For fixed f, $\Delta_0 \le (C_f+1)s^2$, $\Delta_p \le (C_f+1)s^2/c_p^2$, $r^{-1} = P/\epsilon$, and $t = O_f(\log(D/\epsilon))$. Substitution into (122), with $\lceil z \rceil \le z+1$, proves (123). The additive block counts and $m = O_f(s^2/\epsilon)$ are absorbed by either branch for $D \ge 2$. Setting $\epsilon = D^{-\eta/4}$, the quadratic branch has exponent $1 + 4a + \eta/4$ up to logarithms. The tensor branch has exponents $1 + 2a + \eta/4$ and $p/2 + 2a + \eta/8$ up to logarithms. Since $p/2 > 1$, the branch attaining b uses $o(D^{b+\eta})$ observations. Its count is at most $D^{b+\eta}$ for all sufficiently large D. The tensor degree bound also dominates the quadratic one because $p \ge 3$, proving the claimed logarithmic joint degree.

For fixed P, each order-q tensor has D^q ordered entries, each computable as a polynomial of fixed degree. Forming all final tensors costs $O_f(D^P)$ per observation; forming the order-p tensors costs $O_f(D^p)$ per observation. A Gram product costs $O(D^{p+1})$, a matrix multiplication $O(D^3)$, and applying one matrix to an index of an order-q tensor costs $O(D^{q+1})$. Thus a coarse total work bound is

$$O_f\big(ND^P + Pt(D^{p+1} + D^3) + P^2 D^{P+1}\big).$$

Streaming the blocks and retaining the products and final tensors uses $O_f(D^P + D^2)$ storage. Evaluation on a fresh query costs $O_f(D^P)$. These bounds concern real arithmetic with the link fixed, and make no finite-bit complexity assertion. □

J.4 Adjacent Hermite orders

The condition $\beta \neq 0$ is open in the normalized coefficient vector, since β is a polynomial in those coefficients. It also holds throughout the following fixed one-parameter family.

Proposition J.4. *For every odd $p \ge 3$ and every fixed real t_0, the link*

$$f(z) = \frac{h_p(z) + t_0 h_{p+2}(z)}{\sqrt{1+t_0^2}}$$

satisfies the assumptions of Theorem J.1, with $P = p+2$.

Proof. The coefficient of h_p is $(1+t_0^2)^{-1/2} > 0$, and orthogonality gives $\mathbb{E} f(Z)^2 = 1$. The Hermite recurrence $zh_j = \sqrt{j+1}h_{j+1} + \sqrt{j}h_{j-1}$, obtained by differentiating the generating function, gives

$$(z^2-1)h_j = \sqrt{(j+1)(j+2)}h_{j+2} + 2jh_j + \sqrt{j(j-1)}h_{j-2}.$$

Orthogonality therefore yields

$$\beta = \frac{2[p + \sqrt{(p+1)(p+2)}t_0 + (p+2)t_0^2]}{1+t_0^2} > 0.$$

The quadratic in brackets has positive leading coefficient, value $p > 0$ at zero, and discriminant $(p+2)(1-3p) < 0$, so it is positive on the real line. This proves the remaining hypothesis. □

Defilippis et al. (2026, Appendix C) use a tridiagonal quadratic Hermite form to study even links. The following proposition gives an exact certificate for the odd-order family considered here.

Proposition J.5. *Every link $f = \sum_{q \in \{3,5,\dots,15\}} c_q h_q$ with $\sum_q c_q^2 = 1$ satisfies $\mathbb{E}[(Z^2-1)f(Z)^2] \ge 1/8$ for standard Gaussian Z. There is a normalized link supported on $\{3,5,\dots,17\}$ with nonzero coefficients of h_3 and h_{17} for which this expectation is zero.*

Proof. The Hermite recurrence in Proposition J.4 gives $\mathbb{E}[(Z^2-1)f(Z)^2] = c^\top A_m c$ on the first m orders $3, 5, \dots, 2m+1$, where A_m is symmetric tridiagonal with

$$(A_m)_{ii} = 4i+2, \qquad (A_m)_{i,i+1} = \sqrt{(2i+2)(2i+3)}.$$

Let $E_m = \det(8A_m - I_m)$ and set $E_0 = 1$. Expansion along the last row gives $E_1 = 47$ and

$$E_m = (32m+15)E_{m-1} - 64(2m)(2m+1)E_{m-2} \qquad (m \geq 2).$$

Successive substitution gives

$$(E_1, \dots, E_7) = (47, 2433, 143727, 9341697, 622958895, 35684988417, 156144682863).$$

These are the leading principal minors of $8A_7 - I_7$ and are positive. Sylvester's criterion therefore gives $A_7 \succ I_7/8$, proving the first assertion, including links of smaller degree by padding their coefficient vectors with zeros.

For the second assertion, write $A_8 = \left(\begin{smallmatrix} A_7 & b \\ b^\top & 34 \end{smallmatrix}\right)$, where $b = \sqrt{272}\, e_7$, and put $w = A_7^{-1}b$. The same determinant expansion, now without the shift, gives

$$D_0 = 1, \quad D_1 = 6, \quad D_m = (4m+2)D_{m-1} - (2m)(2m+1)D_{m-2}, \qquad D_m = \det A_m.$$

It yields $D_7 = 1627440$ and $D_8 = -6504608$. Eliminating the last off-diagonal block gives $S := 34 - b^\top w = D_8/D_7 = -406538/101715$. Set $z = \sqrt{-S/6} = \sqrt{203269/305145}$ and $v = (w + ze_1, -1) \in \mathbb{R}^8$. Since $A_7 w = b$ and $(A_7)_{11} = 6$,

$$v^\top A_8 v = S + 6z^2 = 0.$$

The cofactor for $(A_7^{-1})_{1,7}$ is the positive product of the six intervening off-diagonal entries. Hence $w_1 = \sqrt{17!/6}/D_7 > 0$, so $v_1 = w_1 + z > 0$ and $v_8 = -1$. The coefficient vector $c = v/\|v\|_2$ consequently gives the claimed normalized degree-17 link. □

The constants may deteriorate when the lowest coefficient or β approaches zero. The conclusions fix the link; they do not assert uniform bounds for dimension-dependent coefficients.

J.5 Proof of the fixed-link theorem

Proof. Corollary I.3, with $\Gamma_P = 3P + 6$, gives the first inequality, including arbitrary measurable design and query coefficients and independent learner randomization. Its prior-mean term is zero for odd p and at most $(P-1)!!D^{-p/2}$ for even p. Rescaling labels by the known positive s preserves label degree. Theorem J.1 gives the second inequality with the displayed joint degree and full expected-risk control. Its assumptions are exactly (7) and $\beta_f \neq 0$. Since $J_f = O_{f,\eta}(\log D)$ and L_f grows as a fixed positive power of D, eventually $J_f \leq L_f$. The containment $\mathcal{Q}_J \subseteq \mathcal{P}_J$ and monotonicity in degree give both conclusions throughout that interval for both classes. □

K Prediction without the link coefficients or noise level

The known-link construction divides its matrix statistics by a nonzero signal coefficient. We replace that division by a polynomial in independent trace estimates. The resulting matrix mean is a controlled multiple of $uu^\top$. We first prove the scalar calibration and the matrix-product bound needed for this change.

K.1 Polynomial scalar calibration

Lemma K.1. *Let $0 < \mu_0 \leq B$ and $0 \leq v \leq \mu_0^2/4$ be known constants, and let $\mu_0 \leq |\mu| \leq B$. For an integer $m \geq 1$, suppose $T_0, \dots, T_{2m-2}$ are independent real random variables with*

$$\mathbb{E}T_j = \mu, \qquad \mathrm{Var}(T_j) \leq v.$$

Put $H = 2B$ and

$$R_m = \frac{T_0}{H^2} \sum_{j=0}^{m-1} \prod_{i=1}^{j} \left(1 - \frac{T_{2i-1}T_{2i}}{H^2}\right), \tag{124}$$

where an empty product is one. Set $x = \mu^2/H^2$, $z = v/\mu^2$, *and* $\alpha = \mu\mathbb{E}R_m$. *Then*

$$\alpha = 1 - (1-x)^m \in [0,1], \tag{125}$$

$$1 - \alpha \le \exp\left(-\frac{m\mu_0^2}{4B^2}\right), \tag{126}$$

$$\mu^2\mathbb{E}R_m^2 \le \frac{1+z}{(1-z/2)^2} \le 1 + 3z, \tag{127}$$

$$\mathbb{E}(\mu R_m - 1)^2 \le \frac{3v}{\mu_0^2} + 2\exp\left(-\frac{m\mu_0^2}{4B^2}\right). \tag{128}$$

If each T_j *is a polynomial of label degree at most* d_{lab} *and joint degree at most* d_{joint} *in its own independent sample block, then* R_m *has these degrees at most* $(2m-1)d_{\text{lab}}$ *and* $(2m-1)d_{\text{joint}}$.

Proof. Independence gives $\mathbb{E}(T_{2i-1}T_{2i}) = \mu^2$. Taking expectations in (124) gives a finite geometric sum, proving (125). Since $1 - x \le e^{-x}$ and $x \ge \mu_0^2/(4B^2)$, it also proves (126).

Write $A_i = 1 - T_{2i-1}T_{2i}/H^2$. We have $0 < x \le 1/4$ and $0 \le z \le 1/4$. Thus independence gives

$$\begin{aligned}\mathbb{E}A_i^2 &= 1 - 2x + \frac{\mathbb{E}T_{2i-1}^2\,\mathbb{E}T_{2i}^2}{H^4} \\ &\le (1-x)^2 + x^2(2z+z^2) \le (1-x)^2 + 3x^2 z =: \rho.\end{aligned}$$

For $a > 0$ and $b \ge 0$, one has $\sqrt{a^2+b} \le a + b/(2a)$ by squaring the right side. Since $1 - x \ge 3/4$, this gives

$$\sqrt{\rho} \le 1 - x + \frac{3x^2 z}{2(1-x)} \le 1 - x + 2x^2 z \le 1 - x(1 - z/2) < 1. \tag{129}$$

The variables A_i are independent of one another and of T_0. Each prefix product has finite second moment, with

$$\left\|\prod_{i=1}^{j} A_i\right\|_2 \le \rho^{j/2}.$$

The triangle inequality in L^2 therefore yields

$$\begin{aligned}\|R_m\|_2 &\le \frac{\sqrt{\mu^2+v}}{H^2}\sum_{j=0}^{m-1}\rho^{j/2} \\ &\le \frac{\sqrt{\mu^2+v}}{H^2(1-\sqrt{\rho})} \le \frac{\sqrt{1+z}}{|\mu|(1-z/2)}.\end{aligned}$$

This uses no independence between different prefix products. For $0 \le z \le 1/4$,

$$(1+3z)(1-z/2)^2 - (1+z) = z\left(1 - \frac{11}{4}z + \frac{3}{4}z^2\right) \ge 0,$$

since the parenthesis is at least $1 - 11/16 > 0$. This proves (127). Finally,

$$\mathbb{E}(\mu R_m - 1)^2 = \mu^2\mathbb{E}R_m^2 - 2\alpha + 1 \le 3z + 2(1-\alpha),$$

which gives (128). The longest term in (124) multiplies $2m-1$ trace estimates, proving both degree bounds. Every sum in the construction is finite. □

K.2 Matrix products with a separate parallel variance

Scalar calibration adds error in the direction of u. The next lemma keeps this error separate from the dimension-dependent perpendicular terms.

Lemma K.2. *Let* $D \ge 2$, *let* u *be a unit vector, and put* $U = uu^\top$, $Q = I - U$, *and* $d = D - 1$. *Let* M *be a symmetric random matrix with finite second moments whose law is invariant under*

$M \mapsto OMO^\top$ for every orthogonal matrix O fixing u. Suppose $\mathbb{E}M = \alpha U$, where $0 \le \alpha \le 1$, and define

$$A = \mathbb{E}(u^\top M u)^2, \qquad B_1 = \mathbb{E}\|QMu\|^2, \qquad C = \frac{1}{d}\mathbb{E}\|QMQ\|_F^2.$$

Let $\zeta, \gamma \ge 0$, let $0 < r \le 1$, and let $t \ge \lceil \log_2(100D/r) \rceil$ be an integer. Suppose

$$A \le 1 + \zeta, \qquad B_1 \le d\gamma, \qquad C \le D\gamma, \qquad h := \zeta + D\gamma \le \frac{r}{100t}. \tag{130}$$

For independent copies $M_1, \dots, M_t$, set $\mathcal{B} = M_t \cdots M_1$. Then

$$\mathbb{E}\mathcal{B} = \alpha^t U, \qquad K_t := \mathbb{E}\mathcal{B}^\top \mathcal{B} = a_t U + b_t Q, \tag{131}$$

$$a_t \le e^{r/100}, \qquad b_t \le h^t + \frac{\gamma e^{r/100}}{1-h}, \qquad \operatorname{tr} K_t < 2. \tag{132}$$

Proof. A symmetric deterministic matrix invariant under every orthogonal map fixing u has the form $aU + bQ$: reflections remove the cross terms, and rotations make its restriction to $u^\perp$ scalar. This invariance holds for the product's second-moment matrix. Starting from $a_0 = b_0 = 1$, independence and symmetry give

$$\begin{pmatrix} a_{j+1} \\ b_{j+1} \end{pmatrix} = \begin{pmatrix} A & B_1 \\ B_1/d & C \end{pmatrix} \begin{pmatrix} a_j \\ b_j \end{pmatrix}. \tag{133}$$

Indeed, a product of $j+1$ independent copies has the same law as $\mathcal{B}'_j M$, where $\mathcal{B}'_j$ is independent of M and distributed as a length-j product. Hence $K_{j+1} = \mathbb{E}[MK_jM]$. Its quadratic form at u is $Aa_j + B_1 b_j$. Its trace on $u^\perp$, divided by d, is $(B_1/d)a_j + Cb_j$, using symmetry of M.

The first row sum in (133) is at most $1+\zeta+d\gamma \le 1+h$. The second is at most $\gamma + D\gamma \le 2h \le 1+h$. All entries are nonnegative. Thus, for $j \le t$,

$$0 \le a_j, b_j \le (1+h)^j \le e^{r/100}.$$

The second recursion separately gives $b_{j+1} \le \gamma e^{r/100} + hb_j$. Iterating from $b_0 = 1$ proves the bound on b_t. Since $d\gamma \le h$, $h \le 1/100$, and $D2^{-t} \le r/100$,

$$\begin{aligned} \operatorname{tr} K_t &\le e^{r/100}\left(1 + \frac{d\gamma}{1-h}\right) + dh^t \\ &\le \frac{e^{r/100}}{1-h} + D2^{-t} \\ &\le \frac{e^{1/100}}{1-1/100} + \frac{1}{100} < 2. \end{aligned}$$

Successive conditioning and independence give finite second moments of every product. Finally, $\mathbb{E}\mathcal{B} = (\alpha U)^t = \alpha^t U$. □

K.3 Calibrated matrix blocks

The next lemma combines the scalar estimate with the matrix-product bound. The division by μ in its hypothesis is used only in the analysis of the raw matrix; the construction does not use this division.

Lemma K.3. *Let D, u, U, Q, d, r, t be as in Lemma K.2. Let $\widetilde{M}$ be a symmetric random matrix with finite second moments, invariant under orthogonal maps fixing u, and with $\mathbb{E}\widetilde{M} = \mu U$. Suppose $0 < \mu_0 \le |\mu| \le B$. For the conceptual normalization $\overline{M} = \widetilde{M}/\mu$, suppose the moments defined as in Lemma K.2 satisfy*

$$A_0 \le 1 + \gamma_0, \qquad B_0 \le d\gamma_0, \qquad C_0 \le D\gamma_0$$

for some $\gamma_0 \ge 0$. Independently of $\widetilde{M}$, construct R_m from Lemma K.1 using trace estimates of common mean μ and variance at most v. Assume

$$\frac{v}{\mu_0^2} \le \frac{r}{1000t}, \qquad D\gamma_0 \le \frac{r}{1000t}, \qquad m \ge \left\lceil \frac{4B^2}{\mu_0^2} \log \frac{1000t}{r} \right\rceil. \tag{134}$$

Then $M = R_m \widetilde{M}$ satisfies Lemma K.2 with

$$\mathbb{E}M = \alpha U, \qquad 0 \le 1-\alpha \le \frac{r}{1000t}, \qquad \zeta = 3z + 2\gamma_0, \quad \gamma = 2\gamma_0, \quad z = \frac{v}{\mu^2}. \tag{135}$$

The construction does not use μ or its sign.

Proof. The first condition in (134) implies $v/\mu_0^2 \le 1/4$, so Lemma K.1 applies. Independence gives $\mathbb{E}M = \mu \mathbb{E}R_m\, U = \alpha U$ and

$$A = (\mu^2 \mathbb{E}R_m^2)A_0, \qquad B_1 = (\mu^2 \mathbb{E}R_m^2)B_0, \qquad C = (\mu^2 \mathbb{E}R_m^2)C_0.$$

Because $z \le 1/4$, (127) gives $\mu^2 \mathbb{E}R_m^2 \le 1 + 3z < 2$. Therefore

$$A \le 1 + 3z + 2\gamma_0, \qquad B_1 \le 2d\gamma_0, \qquad C \le 2D\gamma_0.$$

The proposed values of ζ, γ satisfy

$$\zeta + D\gamma = 3z + 2\gamma_0 + 2D\gamma_0 \le 3z + 4D\gamma_0 \le \frac{7r}{1000t} < \frac{r}{100t}.$$

Thus (130) holds. The last condition in (134), together with (126), gives the bound on $1-\alpha$. Symmetry is preserved by scalar multiplication, independence of the scalar preserves the matrix law's invariance, and independence with finite second moments gives finite second moments for M. These verify every hypothesis of Lemma K.2. Formula (124) uses only the observations and deterministic constants, proving the last assertion. □

K.4 Uniform raw moments and the finite construction

Fix integers $3 \le p \le P$ of the same parity, put $I = \{p, p+2, \dots, P\}$ and $K = |I|$, and suppose

$$f = \sum_{q \in I} c_q h_q, \qquad \sum_{q \in I} c_q^2 = 1, \qquad c_p \ne 0.$$

The coefficients are not supplied to the learner. The observations have the law (120), and the prediction target is $f(u^\top x)$. Write $\beta = \mathbb{E}[(Z^2-1)f(Z)^2]$. Define the known constants

$$\begin{aligned} H_*(z) &= \sum_{q \in I} h_q(z)^2, & C &= \max_{0 \le j \le P} \mathbb{E}[H_*(Z) h_j(Z)^2], \\ G &= 1 + \max_{w \in \{1, z^2, (z^2-1)^2\}} \mathbb{E}[(H_*(Z)^2 + 6H_*(Z) + 3)w(Z)], & & \\ B &= 4P + 3, & \Delta &= C + s^2. \end{aligned} \tag{136}$$

All these Gaussian polynomial moments are finite, and $C, G \ge 1$. Cauchy–Schwarz gives $f^2 \le H_*$ and $f^4 \le H_*^2$ pointwise. The Hermite recurrence $z h_j = \sqrt{j+1}\, h_{j+1} + \sqrt{j}\, h_{j-1}$ and the triangle inequality give $\|Zf(Z)\|_2 \le \sqrt{P+1} + \sqrt{P}$. Consequently $|\beta| = |\|Zf(Z)\|_2^2 - 1| \le B$, while $c_p^2 \le 1$.

Theorem K.4. *Let $D \ge 2$, $s \ge 1$, $0 < \epsilon \le 1$, and $0 < \mu_0 \le 1$. Suppose $|\beta| \ge \mu_0$ and $c_p^2 \ge \mu_0$. Define*

$$\begin{aligned} r &= \epsilon/P, & t &= \lceil \log_2(100D/r) \rceil, & \nu &= \frac{r}{1000t}, \\ m &= \left\lceil \frac{4B^2}{\mu_0^2} \log \frac{1000Pt}{\epsilon} \right\rceil, & m_{\text{out}} &= \left\lceil \frac{2^{P+1} K \Delta}{\epsilon} \right\rceil, & & \\ n_Q &= \left\lceil \frac{2GDs^4}{\mu_0^2 \nu} \right\rceil, & & & & \\ n_T &= \left\lceil \frac{4\Delta D}{\mu_0^2 \nu} + \frac{2\Delta D^{p/2}}{\mu_0 \sqrt{\nu}} \right\rceil. & & & & \end{aligned} \tag{137}$$

There are two predictors, independent of the coefficients of f, using respectively

$$N_Q = 2mPt\, n_Q + m_{\text{out}}, \qquad N_T = 4mPt\, n_T + m_{\text{out}}, \tag{138}$$

whose worst-unit-teacher expected squared prediction risks are less than ϵ. Their label degrees are at most $4mPt+1$ and their respective joint degrees are at most

$$8mPt+2P+1, \qquad 4mP(p+1)t+2P+1. \tag{139}$$

The guarantees are uniform over the displayed coefficient family. The predictor formulas use s only through the deterministic block sizes. Increasing the common raw block size or the final block size preserves the risk bounds.

We first verify the raw statistics to which Lemma K.3 will be applied. Use R_q, Π_q, J_q from Appendix J, and set

$$C_{q,n} = \frac{1}{\sqrt{q!}\,n}\sum_{i=1}^{n} Y_i R_q(X_i).$$

Lemma J.2 applies without using knowledge of the coefficients in this formula. Since $f^2 \le H_*$, its constant C_f is at most C, and hence

$$\mathbb{E}C_{q,n} = c_q u^{\otimes q}, \qquad \mathrm{Cov}(C_{q,n}) \preceq (\Delta/n)\Pi_q \quad (q \in I). \tag{140}$$

Lemma K.5. *Put $U = uu^\top$, $Q = I - U$, and $d = D - 1$. From n observations form*

$$A_Q = \frac{1}{n}\sum_i Y_i^2 (X_i X_i^\top - I).$$

Alternatively, flatten two independent copies of $C_{p,n}$ to D by D^{p-1} matrices C_1, C_2 and form $A_T = (C_1 C_2^\top + C_2 C_1^\top)/2$. For either type write A for the raw matrix and $T = \operatorname{tr} A$. The matrix is symmetric, has finite second moments, is invariant under orthogonal conjugation fixing u, and satisfies

$$\begin{aligned} \mathbb{E}A &= \mu U, & \mathbb{E}(u^\top A u)^2 &\le \mu^2 + \gamma_{\text{raw}}, \\ \mathbb{E}\|QAu\|^2 &\le d\gamma_{\text{raw}}, & d^{-1}\mathbb{E}\|QAQ\|_F^2 &\le D\gamma_{\text{raw}}, \end{aligned} \tag{141}$$

with the following choices:

$$\begin{aligned} \mu_Q &= \beta, & \gamma_{\text{raw},Q} &= Gs^4/n, & \mathrm{Var}(T_Q) &\le 2GDs^4/n, \\ \mu_T &= c_p^2, & \gamma_{\text{raw},T} &= 2\Delta/n + \Delta^2 D^{p-1}/n^2, & \mathrm{Var}(T_T) &\le 2\Delta/n + \Delta^2 D^p/n^2. \end{aligned} \tag{142}$$

In both cases $\mathbb{E}T = \mu$.

Proof. Write $X = (Z, V)$ along u, where V is standard Gaussian in $\mathbb{R}^d$ and independent of (Z, Y). Since $\mathbb{E}[Y^2 \mid Z] = f(Z)^2 + s^2$, the parallel mean of A_Q is β, and its other means vanish. The constant s^2 integrates to zero against $Z^2 - 1$. Also

$$\mathbb{E}[Y^4 \mid Z] = f(Z)^4 + 6s^2 f(Z)^2 + 3s^4 \le s^4(H_*(Z)^2 + 6H_*(Z) + 3).$$

Independence of rows and of V bounds the three variance numerators in (141) by $\mathbb{E}[Y^4(Z^2-1)^2]$, $d\mathbb{E}[Y^4 Z^2]$, and $D\mathbb{E}Y^4$, respectively. For the last, use $\mathbb{E}\|VV^\top - I_d\|_F^2 = dD$. These are bounded by Gs^4, dGs^4, and DGs^4. For the trace, write $\|X\|^2 - D = (Z^2-1) + (\|V\|^2 - d)$. The cross term has expectation zero and $\mathbb{E}(\|V\|^2 - d)^2 = 2d$. Thus

$$\mathrm{Var}(T_Q) \le \frac{\mathbb{E}[Y^4(Z^2-1)^2] + 2d\mathbb{E}Y^4}{n} \le \frac{2GDs^4}{n}.$$

For the tensor block, $A_T = c_p^2 M_T$ for the normalized block in Lemma J.3. The moment calculation in that lemma uses only the covariance bound and is valid for every block size. Its covariance parameter obeys $\kappa = (C_f + s^2)/(c_p^2 n) \le \Delta/(c_p^2 n)$. Multiplying each second-moment bound there by c_p^4 gives

$$c_p^4(2\kappa + \kappa^2 D^{p-1}) \le 2c_p^2\Delta/n + \Delta^2 D^{p-1}/n^2 \le \gamma_{\text{raw},T}.$$

Its mean becomes $c_p^2 U$, proving all matrix bounds.

For completeness the trace needs a separate estimate. Write $C_j = c_p u^{\otimes p} + E_j$ before flattening. The centered independent errors have covariance at most $(\Delta/n)I$ on the ordered tensor space. The trace is

$$T_T = \langle C_1, C_2\rangle = c_p^2 + c_p\langle u^{\otimes p}, E_1 + E_2\rangle + \langle E_1, E_2\rangle.$$

Its three centered terms have zero cross expectations by conditioning on one error. Each linear term has variance at most $c_p^2\Delta/n$. The bilinear second moment is the trace of the product of the two covariance matrices, at most $\Delta^2 D^p/n^2$. This proves (142). Rotating all inputs by an orthogonal O fixing u leaves their joint law with the labels unchanged. The quadratic matrix transforms to $OA_QO^\top$. The identity $R_p(OX) = O^{\otimes p}R_p(X)$ makes the tensor block transform to $OA_TO^\top$. This proves invariance. Symmetry is part of the definitions, and the displayed bounds give finite second moments. □

Proof of Theorem K.4. For each calibrated matrix take one raw matrix A and, independently, $2m-1$ raw traces of the same type, each with the corresponding block size in (137). Form R_m by (124) and set $M = R_mA$. The formula uses neither μ nor its sign. Put $v = \mu_0^2\nu$ and $\gamma_0 = \gamma_{\text{raw}}/\mu_0^2$. The quadratic bounds give $\mathrm{Var}(T_Q) \le v$ and $D\gamma_0 \le \nu/2$. For the tensor type, the two summands in n_T give

$$2\Delta/n_T \le v/(2D), \qquad \Delta^2 D^p/n_T^2 \le v/4.$$

Thus $\mathrm{Var}(T_T) \le v/2$, since $D \ge 2$, and $D\gamma_0 \le \nu/2 + \nu/4$. The conceptual normalization A/μ consequently satisfies the moment hypotheses of Lemma K.3 with this γ_0. All conditions in (134) hold. Lemmas K.3 and K.2 therefore show, for t independent calibrated matrices, that

$$\mathbb{E}\mathcal{B} = \alpha^t U, \qquad K_0 := \mathbb{E}\mathcal{B}^\top\mathcal{B} = a_tU + b_tQ, \qquad a_t \le e^{r/100}, \quad \mathrm{tr}\, K_0 < 2, \quad 1-\alpha \le \frac{\epsilon}{1000Pt}. \tag{143}$$

Take P independent products $\mathcal{B}^{(1)}, \dots, \mathcal{B}^{(P)}$ and an independent final block of size m_{out}. For every $q \in I$, including orders with zero unknown coefficient, define

$$T_q = (\mathcal{B}^{(1)} \otimes \cdots \otimes \mathcal{B}^{(q)})C_{q,m_{\text{out}}}, \qquad \widehat{f}(x) = \sum_{q\in I} J_q(\Pi_q T_q)(x). \tag{144}$$

The final empirical tensors already contain each coefficient in their means. Independence and (140) give

$$\begin{aligned}\mathbb{E}T_q &= c_q\alpha^{qt}u^{\otimes q},\\ \mathbb{E}\|T_q\|_F^2 &\le c_q^2a_t^q + \frac{\Delta}{m_{\text{out}}}\,\mathrm{tr}(K_0^{\otimes q}\Pi_q)\\ &\le c_q^2a_t^q + \frac{\Delta}{m_{\text{out}}}(\mathrm{tr}\, K_0)^q.\end{aligned}$$

The last inequality uses positivity of $K_0^{\otimes q}$ and $0 \preceq \Pi_q \preceq I$. The chaos isometry and contraction of Π_q bound the order-q prediction risk by $c_q^2(a_t^q - 2\alpha^{qt} + 1) + (\Delta/m_{\text{out}})(\mathrm{tr}\, K_0)^q$. For every realized training sample, errors in distinct query chaoses are orthogonal. Their risks therefore add even though the products in (144) are shared across orders. Using $\sum c_q^2 = 1, 1-\alpha^{qt} \le qt(1-\alpha)$ and (143), the total expected risk is at most

$$e^{\epsilon/100} - 1 + 2Pt(1-\alpha) + K\Delta 2^P/m_{\text{out}} \le \left(\frac{1}{99} + \frac{1}{500} + \frac{1}{2}\right)\epsilon < \epsilon.$$

Here $e^{\epsilon/100} - 1 \le \epsilon/99$ follows from $e^z \le (1-z)^{-1}$ for $0 \le z < 1$. The bound is uniform in the unit vector u.

Each calibrated matrix uses $2m$ independent raw blocks in the quadratic construction and $2m$ pairs of tensor blocks in the tensor construction. There are Pt such matrices and one final block, giving (138). Every raw matrix and trace has label degree two; its joint training degree is at most four for the quadratic type and $2(p+1)$ for the tensor type. Equation (124) has degree at most $2m-1$ in the traces. Multiplication by A makes the calibrated matrix's label degree at most $4m$ and its joint training degree at most $8m$ or $4m(p+1)$, respectively. The order-q output uses qt such factors, an empirical tensor of training degree $q+1$, and a query tensor of degree q. This proves (139) and the label-degree bound. Every denominator is deterministic.

All moment estimates decrease when their sample sizes increase. The same argument consequently permits a larger common raw block size or a larger final block size, completing the proof. □

The scalar sum is evaluated by successive products in $O(m)$ operations. There is no need to expand its monomials. A coarse bound for the full construction on N samples is

$$O_{p,P}\big(ND^P + mPt(D^{p+1} + D^3) + P^2D^{P+1}\big)$$

real arithmetic operations. Indeed forming tensors costs $O_P(D^P)$ per observation, a Gram product costs $O(D^{p+1})$, a matrix product costs $O(D^3)$, and applying a matrix to an order-q tensor index costs $O(D^{q+1})$. Streaming the blocks requires $O_P(D^P + D^2)$ storage; a new query costs $O_P(D^P)$. These are real-arithmetic, not bit-complexity, bounds.

K.5 A common predictor without the noise level

The following allocation removes both the link coefficients and the noise level from the learner's inputs. Its dimension threshold may depend on the fixed link and noise exponent.

Theorem K.6. *Fix integers $3 \le p \le P$ of the same parity and $\eta > 0$. There is a collection of predictors $\widehat{f}_{p,P,D,N,\eta}$, indexed by integers $D \ge 2$ and $N \ge 0$, whose coefficients depend only on these indices. Their joint degrees are $O_{p,P,\eta}((\log(eD))^3)$, and they have polynomial real-arithmetic work and storage for fixed p, P, η. For every fixed normalized link $f = \sum_{q=p,p+2,\dots,P} c_q h_q$ with $c_p \neq 0$ and $\beta = \mathbb{E}[(Z^2-1)f(Z)^2] \neq 0$, and every fixed $a > 0$, put $s = D^a$ and $b = \min\{1 + 4a, p/2 + 2a\}$. For all sufficiently large D and every integer $N \ge \lceil D^{b+\eta} \rceil$,*

$$\sup_{\|u\|=1} \mathbb{E}\big(\widehat{f}_{p,P,D,N,\eta}(x) - f(u^\top x)\big)^2 \le D^{-\eta/8}.$$

The learner does not receive f, a, s, or positive lower bounds on $|c_p|$ and $|\beta|$.

Proof. Set $\epsilon = D^{-\eta/8}$ and $\mu_0 = \log^{-1/2}(eD)$. Define r, t, ν, m by (137); these do not use s or the coefficients. Put

$$\mathcal{H} = D^{p-1}, \qquad L_0 = 4mPt + 1, \qquad n = \lfloor N/L_0 \rfloor.$$

If $n = 0$, output zero. If $n \ge 1$ and $N \le \mathcal{H}$, use the quadratic construction with $2mPt$ disjoint raw blocks of size n and one final block of size n. If $N > \mathcal{H}$, use the tensor construction with $4mPt$ disjoint tensor blocks of size n and one final block of size n. The quadratic allocation uses $(2mPt + 1)n \le N$ observations; the tensor allocation uses $L_0 n \le N$. Unused observations are ignored. The branch depends only on known integers. Neither raw statistic, the scalar calibration, nor the output (144) uses the unknown noise level. This defines the stated collection for all D, N.

Fix f and a as in the theorem. Eventually $\mu_0 \le \min\{|\beta|, c_p^2\}$. Let

$$S_Q = Ds^4, \qquad S_T = s^2 D^{p/2}, \qquad S = \min\{S_Q, S_T\} = D^b.$$

Then $S_T^2 = \mathcal{H} S_Q$. If $N \ge \lceil D^{b+\eta} \rceil$, then $S \le ND^{-\eta}$. When $N \le \mathcal{H}$, the possibility $S_Q > \mathcal{H}$ would imply $S_T > \mathcal{H}$ and $S > N$, a contradiction. Thus $S_Q \le \mathcal{H}$, which implies $S_Q \le S_T$, so $S_Q = S \le ND^{-\eta}$. When $N > \mathcal{H}$ and $S_T \le S_Q$, we have $S_T = S \le ND^{-\eta}$. In the remaining case $N > \mathcal{H}$ and $S_Q < S_T$,

$$S_T = \sqrt{\mathcal{H} S_Q} < \sqrt{N S_Q} \le ND^{-\eta/2}.$$

Thus the chosen tensor branch always satisfies $S_T \le ND^{-\eta/2}$.

Since $s \ge 1$, $\Delta \le (C+1)s^2$, and since $p \ge 3$, $Ds^2 \le S_T$ and $Ds^2 \le S$. Directly from (137), $t = O_{P,\eta}(\log(eD))$, $m = O_{P,\eta}((\log(eD))^2)$, and, for constants A_0, c_0 depending only on p, P, η,

$$n_Q \le A_0 S_Q D^{\eta/8} (\log(eD))^{c_0},$$
$$n_T \le A_0 S_T D^{\eta/8} (\log(eD))^{c_0},$$
$$m_{\text{out}} \le A_0 S D^{\eta/8}.$$

The ceilings are absorbed because the factors on the right are at least one. In the quadratic branch, n_Q/N is at most $A_0 D^{-7\eta/8} (\log(eD))^{c_0}$. In the tensor branch, n_T/N is at most

$A_0 D^{-3\eta/8}(\log(eD))^{c_0}$. Also $m_{\text{out}}/N \le A_0 D^{-7\eta/8}$. Because $L_0 = O_{p,P,\eta}((\log(eD))^3)$, all relevant ratios are eventually at most $1/(2L_0)$, uniformly over the allowed integers N. We also eventually have $N \ge 2L_0$, so $n = \lfloor N/L_0 \rfloor \ge N/(2L_0)$. The common raw and final block sizes therefore meet all requirements of Theorem K.4, proving the risk bound.

Equation (139) bounds the joint degree in both branches by $4mP(p+1)t + 2P + 1 = O_{p,P,\eta}((\log(eD))^3)$. The preceding work and storage bounds prove the computational assertion. All choices made by the learner depend only on p, P, D, N, η. If fixed positive lower bounds on $|c_p|$ and $|\beta|$ are imposed, the same proof gives a dimension threshold uniform over that coefficient family, for fixed a, η. □

K.6 An upper bound with a finite noise fourth moment

The upper construction uses Gaussian inputs, but its independent blocks require only four moments of the additive noise.

Proposition K.7. *Fix integers $3 \le p \le P$ of the same parity, $\eta > 0$, and a fixed link $f = \sum_{q=p,p+2,\dots,P} c_q h_q$ with $\sum_q c_q^2 = 1$, $c_p \ne 0$, and $\beta = \mathbb{E}[(Z^2-1)f(Z)^2] \ne 0$ for standard Gaussian Z. For $M \ge 1$, let $\mathcal{N}_M$ consist of the probability laws with mean zero, variance one, and fourth moment at most M. Consider observations*

$$X_i \sim N(0, I_D), \qquad Y_i = f(u^\top X_i) + s\xi_i,$$

where the X_i are iid and the ξ_i are iid with a law in $\mathcal{N}_M$ and are independent of all inputs. The query is independent standard Gaussian. Use exactly the collection $\widehat{f}_{p,P,D,N,\eta}$ in Theorem K.6. For every fixed $a > 0$ and finite $M \ge 1$, put $s = D^a$ and $b = \min\{1+4a, p/2+2a\}$. For all sufficiently large D and every integer $N \ge \lceil D^{b+\eta} \rceil$,

$$\sup_{\mathcal{L}(\xi)\in\mathcal{N}_M} \sup_{\|u\|=1} \mathbb{E}_{\xi,u}\big(\widehat{f}_{p,P,D,N,\eta}(x) - f(u^\top x)\big)^2 \le D^{-\eta/8}.$$

The learner receives neither the noise law nor M. The dimension threshold may depend on p, P, η, a, M, f. Its joint degree remains $O_{p,P,\eta}(\log^3(eD))$.

Proof. Retain $H_*, C, \Delta = C + s^2$ and $B = 4P + 3$ from (136). Replace only the analysis constant G by

$$G_M = 1 + 8 \max_{w\in\{1,z^2,(z^2-1)^2\}} \mathbb{E}[(H_*(Z)^2 + M)w(Z)]. \tag{145}$$

It is finite and depends only on p, P, M.

The tensor mean and covariance in (140) use only the first two noise moments. Indeed, for an order-q Gaussian chaos g, independence gives

$$\mathbb{E}[Yg(X)] = \mathbb{E}[f(u^\top X)g(X)], \qquad \mathbb{E}[Y^2 g(X)^2] = \mathbb{E}[f(u^\top X)^2 g(X)^2] + s^2\|g\|_2^2 \le \Delta\|g\|_2^2.$$

The last inequality is the coordinate-chaos calculation in Lemma J.2, with $C_f \le C$. Thus (140) is unchanged. The tensor Gram calculations in Lemma K.5 use only these covariances and independent centered errors. They give the same signal $\mu_T = c_p^2$ and bounds

$$\gamma_{\text{raw},T} = 2\Delta/n + \Delta^2 D^{p-1}/n^2, \qquad \operatorname{Var}(\operatorname{tr} A_T) \le 2\Delta/n + \Delta^2 D^p/n^2.$$

For the quadratic block, write $X = (Z, V)$ along u, and put $U = uu^\top$, $Q = I - U$, $d = D - 1$. The identity $\mathbb{E}[Y^2 \mid Z] = f(Z)^2 + s^2$ still gives $\mathbb{E}A_Q = \beta U$ for $A_Q = n^{-1}\sum_i Y_i^2(X_i X_i^\top - I)$. Convexity gives $|a+b|^4 \le 8(|a|^4 + |b|^4)$ for real a, b. Consequently, since $s \ge 1$ and $f^2 \le H_*$,

$$\mathbb{E}[Y^4 \mid Z] \le 8(f(Z)^4 + s^4 M) \le 8s^4(H_*(Z)^2 + M). \tag{146}$$

This inequality does not assume symmetric noise. The three quadratic matrix variance numerators are bounded by $\mathbb{E}[Y^4(Z^2-1)^2]$, $d\mathbb{E}[Y^4 Z^2]$, and $D\mathbb{E}Y^4$. These follow from row independence, the independence of V from (Z, ξ), and $\mathbb{E}\|VV^\top - I_d\|_F^2 = dD$. Equations (145)–(146) therefore give all matrix bounds in (141) with $\mu_Q = \beta$ and $\gamma_{\text{raw},Q} = G_M s^4/n$. For the trace, decompose

$\|X\|^2 - D = (Z^2 - 1) + (\|V\|^2 - d)$. The cross expectation is zero and the latter term has variance $2d$, independently of (Z, Y). Thus

$$\mathrm{Var}(\mathrm{tr}\, A_Q) \le \frac{\mathbb{E}[Y^4(Z^2-1)^2] + 2d\mathbb{E}Y^4}{n} \le 2DG_M s^4/n.$$

Both raw matrices retain conjugation invariance under orthogonal maps fixing u, since the Gaussian input and independent scalar noise have the required joint invariance.

For the parameters r, t, ν, m in (137), the finite proof consequently requires only

$$\begin{aligned} n_Q &\ge \left\lceil \frac{2G_M D s^4}{\mu_0^2 \nu} \right\rceil, \\ n_T &\ge \left\lceil \frac{4\Delta D}{\mu_0^2 \nu} + \frac{2\Delta D^{p/2}}{\mu_0 \sqrt{\nu}} \right\rceil, \\ n_{\text{out}} &\ge \left\lceil \frac{2^{P+1} K \Delta}{\epsilon} \right\rceil. \end{aligned} \tag{147}$$

Here $0 < \mu_0 \le \min\{|\beta|, c_p^2, 1\}$. These inequalities give trace variance at most $\mu_0^2\nu$ and normalized matrix error $D\gamma_0 \le \nu$, precisely the hypotheses checked in Theorem K.4. The signal bound $|\mu| \le B$ is unchanged. Lemmas K.1, K.3 and K.2 concern independent random objects with the specified second moments and invariance; they make no Gaussian noise assumption. They give the same product bounds in (143). Applying them to the final tensors in (144) gives full expected risk at most

$$e^{\epsilon/100} - 1 + 2Pt(1-\alpha) + K\Delta 2^P/n_{\text{out}} \le (1/99 + 1/500 + 1/2)\epsilon < \epsilon.$$

Different output errors remain orthogonal in the Gaussian query for every realized training sample.

No noise moments above order four are needed for these growing-degree polynomials. Each raw quadratic matrix and trace has finite second moment by (146), and the tensor blocks need only the covariance bound. Each monomial of the scalar calibrator uses an independent trace block at most once; squaring uses it at most twice. Lemma K.1 therefore needs only its stated second moments. The raw matrix is independent of that calibrator, and successive matrix factors use fresh blocks. The matrix-product lemma gives finite second moments by successive conditioning. The common final-tensor block is independent of all matrix products. Sharing products and this final block across orders introduces no cross-risk terms because of query orthogonality.

Finally use the exact blind allocation in Theorem K.6: $\epsilon = D^{-\eta/8}$, $\mu_0 = \log^{-1/2}(eD)$, $L_0 = 4mPt + 1$, and $n = \lfloor N/L_0 \rfloor$, choosing the quadratic type when $N \le D^{p-1}$ and the tensor type otherwise. These choices use neither G_M nor M. Write $\ell = \log(eD)$, $S_Q = Ds^4$, $S_T = s^2 D^{p/2}$, and $S = \min\{S_Q, S_T\} = D^b$. Then $t = O_{P,\eta}(\ell)$, $m = O_{P,\eta}(\ell^2)$, and $L_0 = O_{p,P,\eta}(\ell^3)$. The three requirements in (147) are at most

$$AS_Q D^{\eta/8}\ell^2, \qquad AS_T D^{\eta/8}\ell^2, \qquad ASD^{\eta/8},$$

respectively, for A depending only on p, P, η, M. Indeed $\Delta \le (C+1)s^2$, $(\mu_0^2\nu)^{-1} = O_{P,\eta}(D^{\eta/8}\ell^2)$, and $(\mu_0\sqrt{\nu})^{-1} = O_{P,\eta}(D^{\eta/16}\ell)$. Since $p \ge 3$, $Ds^2 \le S_T$ and $s^2 \le S$.

The deterministic comparison in the proof of Theorem K.6 uses only $S_T^2 = D^{p-1}S_Q$ and $S \le ND^{-\eta}$. It gives $S_Q \le ND^{-\eta}$ in the selected quadratic branch and $S_T \le ND^{-\eta/2}$ in the selected tensor branch. Thus the selected raw requirement divided by N is at most $AD^{-7\eta/8}\ell^2$ or $AD^{-3\eta/8}\ell^2$, respectively; the final-block ratio is at most $AD^{-7\eta/8}$. Eventually all are at most $1/(2L_0)$ and $N \ge 2L_0$, uniformly over the stated integers N. Hence $n \ge N/(2L_0)$ meets (147). The signal cutoff is also eventually admissible because the link is fixed. The finite risk bound proves the assertion uniformly over $\mathcal{N}_M$. Fixed positive lower bounds on $|c_p|$ and $|\beta|$ make this dimension threshold uniform over the corresponding coefficient family. The predictor itself is unchanged, so its degree bound is unchanged as well. □

L A vanishing first Hermite coefficient

We retain the coefficient of each linear label factor in the coordinate-identification bound, allowing any fixed normalized mixture of higher odd orders.

Theorem L.1. *Fix odd integers $3 \le p \le P$ and real numbers $a, b, \eta > 0$. Let $I \subseteq \{p, p+2, \dots, P\}$ contain p, and fix real coefficients $(c_q)_{q\in I}$, independently of dimension, such that $c_p \neq 0$ and $\sum_{q\in I} c_q^2 = 1$. Define*

$$f_0(z) = \sum_{q\in I} c_q h_q(z), \qquad h_j = \mathrm{He}_j / \sqrt{j!}.$$

Put

$$B = \min\left\{1 + 4a, \frac{p}{2} + 2a, 1 + 2a + \frac{2b(p-2)}{p-1}\right\},$$
$$\Gamma_P = 3P + 6, \qquad \tau = \min\{2a, \eta/2\}, \qquad L = \left\lfloor D^{\tau/(4\Gamma_P)} \right\rfloor. \tag{148}$$

Let D, N be integers satisfying

$$D \ge \left\lceil \max\{2, (4P)^{4\Gamma_P/\tau}, 2^{2/\tau}\} \right\rceil, \qquad 0 \le N \le D^{B-\eta}.$$

Set $\varepsilon = D^{-b}$, $s = D^a$, and

$$f_D(z) = \frac{f_0(z) + \varepsilon h_1(z)}{\sqrt{1+\varepsilon^2}}.$$

Take U uniformly from $\{-1,1\}^D/\sqrt{D}$, an independent standard Gaussian design $A \in \mathbb{R}^{N\times D}$, independent $G \sim N(0, I_N)$, and an independent query $x \sim N(0, I_D)$. Observe $Y_i = f_D(A_i^\mathsf{T} U) + sG_i$, and write $V = Y/s$. Every predictor of the form

$$\widehat{F}(A, V, \xi; x) = \sum_{|\alpha|\le L} a_\alpha(A, \xi, x) V^\alpha \tag{149}$$

with jointly measurable real coefficients and a seed ξ independent of U, A, G, x satisfies

$$\mathbb{E}[(\widehat{F}(A, V, \xi; x) - f_D(U^\mathsf{T} x))^2] \ge \frac{1}{1 + D^{-2b}} - 2D^{-\tau/2} \ge 1 - D^{-2b} - 2D^{-\tau/2}. \tag{150}$$

The coefficients may depend on the known parameters and on the entire design and query, but not on teacher-dependent side information. The same bound holds for worst-teacher risk over unit vectors and for the minimax risk within this class. It also holds for predictors of joint polynomial degree at most L in the design, labels and query.

Proof. Let $c = (1+\varepsilon^2)^{-1/2}$ and let $\mathcal{H} = \bigoplus_{q\in I} \mathcal{H}_q$ be the direct sum of the nonlinear query chaoses. Project the predictor and target onto $\mathcal{H}$. The projected target is

$$F_U^{\mathrm{nl}}(x) = c \sum_{q\in I} c_q h_q(U^\mathsf{T} x), \qquad \|F_U^{\mathrm{nl}}\|_2^2 = c^2 \sum_{q\in I} c_q^2 = c^2,$$

by (18) and orthogonality of distinct chaoses. Projection cannot increase risk. The order-one component remains in the observations but is not included in this target.

The conditional certificate. Put $T_i = f_D(A_i^\mathsf{T} U)/s$. At fixed design define

$$m_\alpha = \mathbb{E}_U T^\alpha, \qquad v_\alpha = \mathbb{E}_U[F_U^{\mathrm{nl}} T^\alpha],$$
$$\kappa_\alpha = v_\alpha - \sum_{\substack{\beta\le\alpha \\ \beta\neq\alpha}} \binom{\alpha}{\beta} \kappa_\beta m_{\alpha-\beta}, \qquad E_k = \mathbb{E}_A \sum_{|\alpha|=k} \frac{\|\kappa_\alpha(A)\|_{\mathcal{H}}^2}{\alpha!}. \tag{151}$$

The projection and certificate argument of Section I.1 applies to this target. Here are its hypotheses in the present setting. Finite risk implies an integrable squared norm of the predictor, since the target has norm one before projection. At each fixed design, the label law is a finite mixture of identity-covariance Gaussian laws. Its monomial Gram matrix is positive definite, because its density is strictly positive and a nonzero polynomial cannot vanish Lebesgue almost everywhere. The coefficient integrability argument there therefore puts every coefficient in $L^2(\gamma_D)$ for almost every

design and seed. Projection onto the finite-dimensional space $\mathcal{H}$ is measurable and preserves label degree. Exceptional null sets can be assigned the zero coefficients.

For completeness, the last certificate step uses (19). Expand the projected predictor in the orthonormal product-Hermite basis of the labels, with coefficients $b_\alpha \in \mathcal{H}$. Define the triangular matrix

$$M_{\beta,\alpha} = \mathbf{1}_{\{\beta\le\alpha\}} m_{\alpha-\beta}\sqrt{\frac{\beta!}{\alpha!}}\binom{\alpha}{\beta}, \qquad z_\alpha = \frac{v_\alpha}{\sqrt{\alpha!}}, \quad w_\alpha = \frac{\kappa_\alpha}{\sqrt{\alpha!}}.$$

The shift identity and Jensen's inequality over U give the cross term $\sum_\alpha \langle z_\alpha, b_\alpha\rangle$ and a predictor squared norm at least $\sum_\beta \|(Mb)_\beta\|^2$, as established in that section. Recursion (151) gives $z = M^{\mathsf{T}} w$. Completing squares at fixed design and seed and then averaging proves

$$\text{Risk} \ge c^2 - \sum_{k=0}^{L} E_k. \tag{152}$$

The initial constant is $c^2 = (1+\varepsilon^2)^{-1}$. Every cumulant coordinate is a polynomial in A, because the teacher prior is finite. Thus all energies are finite. The argument also covers dimension-dependent ε, since D is fixed in each certificate. Infinite-risk predictors satisfy the lower bound without projection.

The cumulant partition formula (20) and its consequences in Section A.2 apply coordinatewise. Every target or label tensor factor reduces to a nonempty odd sign character. Thus $\kappa_0 = 0$ and every even-order κ_α vanishes: empty total symmetric difference requires $|\alpha| + 1$ to be even. Nonzero cumulants also require a connected character intersection graph and have absolute value at most $(k+1)^{k+1}$ when $|\alpha| = k$.

Orders and coefficient weights. Fix odd $k \ge 1$ and a multiindex α of order k with r occupied design rows. Expand its tensor cumulant using (22) at every order in $\{1\}\cup I$, and square its norm. The tensor isometry and normal-order pairing identities in Section A.3 apply at each order, as explained in Section I.2. Distinct query chaoses are orthogonal, so only equal target orders pair across replicas. Fix such an order $q \in I$. The $2k$ label factors have orders $j_v \in \{1\}\cup I$. Let m count their order-one factors, and put

$$R = \sum_{v=1}^{2k} j_v, \qquad e = R/2 + q \le P(k+1).$$

Within a design row there are twice its multiplicity many odd-order factors, so its number of coordinate positions is even. Every surviving Gaussian pairing stays inside a row. There is one initial coordinate variable for each of the $R/2$ Gaussian pairs and one for each of the q target bridges. The exact scalar factor for these choices of orders is

$$s^{-2k} D^{-e} (1+\varepsilon^2)^{-(k+1)} \varepsilon^m c_q^2 \prod_{\substack{1\le v\le 2k\\ j_v>1}} \frac{c_{j_v}}{\sqrt{j_v!}}. \tag{153}$$

Indeed, an order-j_v label contributes $D^{-j_v/2}/(s\sqrt{j_v!})$, multiplied by $c\varepsilon$ when $j_v = 1$ and by cc_{j_v} otherwise. The target isometry removes the target factorial. The two order-q target coefficients contribute $c^2 c_q^2$, and their coordinate factors contribute D^{-q}. Combining these factors proves (153); no extra target matching factor occurs.

Partition the e coordinate variables into h equality classes, and put $\ell = e - h$. The number of distinct coordinate assignments is $(D)_h \le D^h$, and is zero when $h > D$. Consequently

$$D^{-e}(D)_h \varepsilon^m \le D^{-(\ell+bm)}. \tag{154}$$

Remove terms with a zero cumulant in either replica.

The weighted graph bound. Form the bipartite multigraph of Section I.3, whose left vertices are the r rows and the target and whose right vertices are the h coordinate classes. Each initial variable is an edge. The nonzero-cumulant intersection graph connects the factor groups through shared coordinates. Identifying factors within a row preserves this connection. A coordinate class absent from the reduced characters still attaches to one of these groups. Thus the graph B is connected, as in that section, and

$$\beta(B) = e - (r + 1 + h) + 1 = \ell - r \ge 0. \tag{155}$$

Mark the cross-replica Gaussian pairs and the target bridges, and delete isolated vertices to obtain M. Every coordinate vertex of M has positive even degree. Its marked degree has the parity of its number of positions in one replica, because internal pairs contribute two positions and crossing pairs or bridges contribute one. Empty symmetric difference forces that parity to be even. Each such vertex therefore has degree at least two. The target has marked degree $q \geq p$.

Let r_1 count singleton rows, r_o other odd-multiplicity rows, and r_e even-multiplicity rows. A singleton has one normal-ordered factor in each replica. No pair can stay inside either factor, so all its pairs cross replicas. Unequal orders in that row admit no pairing. Hence every surviving singleton has either two factors of a common nonlinear order at least p, or two order-one factors. Its marked degree is that common order. Let j count the linear singleton rows. Then $m \geq 2j$.

In a nonsingleton row, the sum of the left-replica orders has the parity of the row multiplicity, since every order is odd. Removing internal pairs preserves this parity. Therefore an odd-multiplicity row has marked degree at least one, and an even-multiplicity row is isolated or has marked degree at least two. This reasoning allows unequal orders and arbitrary numbers of linear factors in the two replicas.

The marked target ensures that M has at least one component. The degree identity for cycle rank gives

$$\begin{aligned}\beta(M) &\geq 1 + \frac{(p-2)(r_1 - j + 1) - j - r_o}{2} \\ &= \frac{(p-2)r_1 - r_o + p - (p-1)j}{2}.\end{aligned}$$

The coordinate vertices and nonisolated even rows contribute nonnegatively. Cycle rank is monotone under taking subgraphs, since independent cycle vectors remain independent in the full cycle space. Moreover,

$$k \geq r_1 + 3r_o + 2r_e = 2r - r_1 + r_o, \qquad (p-2)r_1 - r_o = (p-2)(r_1 - r_o) + (p-3)r_o \geq (p-2)(2r-k).$$

Together with (155), these inequalities give

$$\ell \geq \max\left\{r,\ (p-1)r - \frac{p-2}{2}k + \frac{p}{2} - \frac{p-1}{2}j\right\}. \tag{156}$$

Set

$$\theta = \min\left\{1, \frac{4b}{p-1}\right\}, \qquad \chi = \theta(p-2) > 0.$$

Take the convex combination of the two branches of (156), with weight θ on the second, and use $bm \geq 2bj$. This yields

$$\ell + bm \geq (1+\chi)r - \frac{\chi}{2}k + \frac{\theta p}{2} + \left(2b - \frac{\theta(p-1)}{2}\right) j \geq (1+\chi)r - \frac{\chi}{2}k + \frac{\theta p}{2}.$$

The last coefficient is nonnegative. Since also $\ell + bm \geq r$,

$$\ell + bm \geq \Delta_b(k, r) := \max\left\{r, (1+\chi)r - \frac{\chi}{2}k + \frac{\theta p}{2}\right\}. \tag{157}$$

Linear factors in nonsingleton rows add a nonnegative penalty and have not been excluded.

Counting. For fixed target and label orders there are at most $(2Pk)^{Pk}$ Gaussian pairings. There are at most $e^e \leq (2Pk)^{2Pk}$ partitions of the initial variables, because $e \leq P(k+1) \leq 2Pk$. The product of the absolute cumulants is at most $(k+1)^{2(k+1)} \leq (2Pk)^{4k}$ by (21).

Apply (154) and (157) before summing the label orders, so that the factor ε^m supplies its full coordinate penalty. Put $K = |I|$. The remaining absolute coefficient mass of the label assignments is at most

$$\left(1 + \sum_{q \in I} |c_q|\right)^{2k} \leq (1 + \sqrt{K})^{2k} \leq \big(2(K+1)\big)^k \leq (2P)^k \leq (2Pk)^k.$$

The first inequality is Cauchy–Schwarz and uses $\sum_q c_q^2 = 1$; the next uses $(1+\sqrt{K})^2 \le 2(K+1)$. Finally $K \le (P-1)/2$, since $p \ge 3$. The normalization $(1+\varepsilon^2)^{-(k+1)}$ and the factorials in (153) are at most one. Summing the target orders contributes $\sum_q c_q^2 = 1$. Thus, for a fixed multiindex,

$$\mathbb{E}_A \|\kappa_\alpha(A)\|^2 \le (2Pk)^{(3P+5)k} s^{-2k} D^{-\Delta_b(k,r)}.$$

The number of multiindices of order k and support size r is $\binom{N}{r}\binom{k-1}{r-1} \le N^r 2^{k-1}$. Drop $1/\alpha! \le 1$ and absorb $2^{k-1} \le (2Pk)^k$ to obtain

$$E_k \le (2Pk)^{\Gamma_P k} s^{-2k} \sum_{r=1}^{\min\{N,k\}} N^r D^{-\Delta_b(k,r)}. \tag{158}$$

All expansions are finite and include repeated rows and coordinates. For $N = 0$, this sum is empty for positive k, and (152) already proves the result.

The exponent and degree range. Assume $N > 0$. By the definition of χ,

$$B = \min\{1 + 4a, 1 + \chi/2 + 2a\}.$$

Each term in (158) has dimension exponent at most

$$e_{k,r} = (B - \eta)r - 2ak - \Delta_b(k, r).$$

Put $\omega = B - \eta - 1$. If $\omega < 0$, the first branch gives $e_{k,r} \le \omega r - 2ak \le -2ak$. If $0 \le \omega \le \chi$, combine the two branches of Δ_b with weight ω/χ on the second. Since $\omega \le 4a - \eta$,

$$e_{k,r} \le -(2a - \omega/2)k - \frac{\omega\theta p}{2\chi} \le -\eta k/2.$$

If $\omega > \chi$, then $a > \chi/4$: otherwise $\omega \le 4a - \eta < \chi$. Hence $B = 1 + \chi/2 + 2a$. Write $t = 2a - \chi/2$. The identity $\omega = \chi + t - \eta > \chi$ implies $t > \eta$, and the second branch gives

$$e_{k,r} \le (t - \eta)r - tk - \theta p/2 = -t(k - r) - \eta r - \theta p/2 \le -\eta k - \theta p/2.$$

Thus $e_{k,r} \le -\tau k$ in all cases. There are at most k terms in the sum over r, so

$$E_k \le (4Pk)^{\Gamma_P k} D^{-\tau k}.$$

For $k \le L$ and the stated lower bound on D,

$$(4Pk)^{\Gamma_P} \le (4P)^{\Gamma_P} D^{\tau/4} \le D^{\tau/2}.$$

Consequently $E_k \le D^{-\tau k/2}$. Since $D^{-\tau/2} \le 1/2$, a geometric sum bounds the total energy by $2D^{-\tau/2}$. Apply (152), and use $(1 + D^{-2b})^{-1} \ge 1 - D^{-2b}$, to obtain (150).

For each fixed learner, its sign-prior risk is at most its worst-teacher risk. Taking the infimum over learners preserves the lower bound. A joint polynomial of degree at most L has label degree at most L and measurable design and query coefficients, so it is included in (149). Finally, scaling labels by the fixed $s > 0$ preserves label degree. This completes every stated case. □

M A POLYNOMIAL UPPER BOUND WITH A WEAK LINEAR COMPONENT

Let h_j denote the normalized probabilists' Hermite polynomial. Fix odd integers $3 \le p \le P$ and a polynomial

$$f_0(z) = \sum_{q \in I} c_q h_q(z), \qquad I \subseteq \{p, p+2, \dots, P\}, \qquad c_p \ne 0, \qquad \sum_{q \in I} c_q^2 = 1,$$

where I consists of the orders with nonzero coefficients. The coefficients and endpoint orders are fixed independently of dimension. Assume

$$\beta_0 = \mathbb{E}[(Z^2 - 1) f_0(Z)^2] \ne 0, \qquad Z \sim N(0, 1).$$

For $0 < \varepsilon \le 1$, put

$$f(z) = \frac{f_0(z) + \varepsilon z}{\sqrt{1+\varepsilon^2}}, \qquad J = \{1\} \cup I, \qquad \vartheta_1 = \lambda = \frac{\varepsilon}{\sqrt{1+\varepsilon^2}},$$

and put $\vartheta_q = c_q/\sqrt{1+\varepsilon^2}$ for $q \in I$. Thus $f = \sum_{q\in J} \vartheta_q h_q$ and $\sum_{q\in J} \vartheta_q^2 = 1$. Write $\kappa = \vartheta_p$; this scalar is nonzero and may be negative. The observations are independent copies of $X \sim N(0, I_D)$ and $Y = f(u^\top X) + s\xi$, where the noises are independent and identically distributed, independent of all design rows, and u is an unknown unit vector. For $1 \le M < \infty$, let $\mathcal{N}_M$ be the set of probability laws ν on $\mathbb{R}$ satisfying

$$\mathbb{E}_\nu \xi = 0, \qquad \mathbb{E}_\nu \xi^2 = 1, \qquad \mathbb{E}_\nu \xi^4 \le M.$$

The common noise law belongs to $\mathcal{N}_M$; it need not be symmetric. Prediction risk is measured against $f(u^\top x)$ at an independent query $x \sim N(0, I_D)$.

Warm-start tensor amplification in Lemma F.4 of Damian et al. (2024), arXiv version 2, supplies prior ancestry for the hybrid recovery scale. We prove the upper here in the exact joint-polynomial class with full expected prediction risk.

Theorem M.1. *Fix f_0, p, P as above and a real $\eta > 0$. There is a collection of predictors $\widehat{f}_{f_0,D,N,\varepsilon,\eta}$, indexed by integers $D \ge 2$, $N \ge 0$ and reals $0 < \varepsilon \le 1$, whose coefficients depend only on f_0, p, P, η and these indices. Their joint degrees in the training design, labels and query are $O_{p,P,\eta}(\log D)$. For every fixed $a, b > 0$ and $1 \le M < \infty$, put $s = D^a$, $\varepsilon = D^{-b}$ and*

$$B = \min\left\{1 + 4a, \frac{p}{2} + 2a, 1 + 2a + \frac{2b(p-2)}{p-1}\right\}.$$

There is $D_0 = D_0(f_0, p, P, a, b, \eta, M)$ such that, for every integer $D \ge D_0$ and every integer $N \ge \lceil D^{B+\eta} \rceil$,

$$\sup_{\nu\in\mathcal{N}_M} \sup_{\|u\|=1} \mathbb{E}_\nu\left[(\widehat{f}_{f_0,D,N,\varepsilon,\eta}(x) - f(u^\top x))^2\right] \le D^{-\eta/4}.$$

The learner receives f_0, p, P and ε, but not a, b, s, M or the noise law as separate inputs. It has polynomial real-arithmetic work and storage for fixed P.

The hybrid construction contracts a fresh tensor against independent linear pilots. Independence makes the matrix mean exactly $uu^\top$ using only the pilots' first two moments. The final output retains every target order.

M.1 Uniform tensor and matrix bounds

Let $R_q(X)$ be the normal-ordered Gaussian tensor, let Π_q be orthogonal projection onto symmetric order-q tensors, and set $J_q(H) = \langle H, R_q(X)\rangle/\sqrt{q!}$. The generating function $\exp(v^\top X - \|v\|^2/2)$ gives

$$\mathbb{E}[J_q(H)J_q(K)] = \langle \Pi_q H, \Pi_q K\rangle, \qquad J_q(u^{\otimes q}) = h_q(u^\top X). \tag{*}$$

Indeed, multiply two generating functions, integrate to obtain $\exp(v^\top w)$, and equate coefficients; coefficients of different total orders are orthogonal. Define

$$H_P(z) = \sum_{\substack{1\le j\le P\\ j \text{ odd}}} h_j(z)^2.$$

Cauchy–Schwarz and the normalized coefficients give $f^2 \le H_P$. Define the finite constants

$$C_P^{\mathrm{cov}} = \max\left\{1, \max_{0\le j\le P} \mathbb{E}[H_P(Z)h_j(Z)^2]\right\},$$

$$G_{f_0,M}^{\mathrm{quad}} = 1 + \frac{32}{\beta_0^2} \max_{w\in\{1,z^2,(z^2-1)^2\}} \mathbb{E}[(H_P(Z)^2 + M)w(Z)],$$

$$\Delta = C_P^{\mathrm{cov}} + s^2, \qquad Z \sim N(0,1).$$

The first constant depends only on P, and the second only on P, β_0, M. Neither depends on ε, D, s, u, or the choice of law in $\mathcal{N}_M$. These constants are used only in the analysis. Below, generic constants

depend only on their indicated subscripts and may change between occurrences. All moment estimates below concern $s \geq 1$, as in the asserted range. For a block of size n, put

$$C_{q,n} = \frac{1}{\sqrt{q!}\, n} \sum_{i=1}^{n} Y_i R_q(X_i), \qquad q \in J.$$

Then

$$\mathbb{E} C_{q,n} = \vartheta_q u^{\otimes q}, \qquad \mathrm{Cov}(C_{q,n}) \preceq \frac{\Delta}{n} \Pi_q, \qquad q \in J. \tag{159}$$

For the covariance assertion, rotate u to the first coordinate and expand an order-q chaos as $g(Z, W) = \sum_{j=0}^{q} h_j(Z) g_{q-j}(W)$. Orthogonality of the distinct orders in W gives

$$\mathbb{E}[f(Z)^2 g(Z, W)^2] = \sum_{j=0}^{q} \mathbb{E}[f(Z)^2 h_j(Z)^2] \|g_{q-j}\|_2^2 \leq C_P^{\mathrm{cov}} \|g\|_2^2,$$

since $q \leq P$ and $f^2 \leq H_P$. The independent noise contributes $s^2 \|g\|_2^2$, and averaging independent rows divides variance by n. The mean assertions follow from (*) and orthogonality.

Write $U = uu^\top$, $Q = I - U$, and $d = D - 1$. Put $c_3 = 0$ if $3 \notin I$. Direct Hermite orthogonality gives

$$\beta = \mathbb{E}[(Z^2 - 1) f(Z)^2] = \frac{\beta_0 + 2\sqrt{6}\, c_3 \varepsilon + 2\varepsilon^2}{1 + \varepsilon^2}. \tag{160}$$

Indeed, $(Z^2 - 1)Z = \sqrt{6} h_3(Z) + 2h_1(Z)$ and $\mathbb{E}[(Z^2 - 1)Z^2] = 2$. Thus $|\beta| \leq C_P$ for $0 < \varepsilon \leq 1$, and $|\beta| \geq |\beta_0|/2$ for all sufficiently small ε. Also $|\kappa| \geq |c_p|/\sqrt{2}$.

When $\beta \neq 0$, the first matrix block is

$$M_Q = \frac{1}{\beta n} \sum_{i=1}^{n} Y_i^2 (X_i X_i^\top - I).$$

It has mean U: conditional expectation of Y^2 is $f(Z)^2 + s^2$, and the constant noise term vanishes against $XX^\top - I$. To check uniform second moments, write $X = (Z, V)$ along u and $W_0 = Y^2$. The first two noise moments and the inequality $|x + y|^4 \leq 8(|x|^4 + |y|^4)$ give

$$\begin{aligned} \mathbb{E}[W_0 \mid Z] &= f(Z)^2 + s^2, \\ \mathbb{E}[W_0^2 \mid Z] &\leq 8(f(Z)^4 + s^4 M) \leq 8 s^4 (H_P(Z)^2 + M). \end{aligned}$$

The scalar inequality follows from convexity of $x \mapsto x^4$ applied to $(x + y)/2$. The last bound uses $s \geq 1$ and $f^2 \leq H_P$ and holds without a symmetry assumption on the noise. Consequently, whenever $|\beta| \geq |\beta_0|/2$, set $\gamma_Q = G^{\mathrm{quad}}_{f_0, M} s^4 / n$ to obtain

$$\begin{aligned} \mathbb{E}(u^\top M_Q u)^2 &= 1 + \frac{\mathbb{E}[W_0^2 (Z^2 - 1)^2] - \beta^2}{\beta^2 n} \leq 1 + \gamma_Q, \\ \mathbb{E}\|Q M_Q u\|^2 &= \frac{d\, \mathbb{E}[W_0^2 Z^2]}{\beta^2 n} \leq d \gamma_Q, \\ d^{-1} \mathbb{E}\|Q M_Q Q\|_F^2 &= \frac{D\, \mathbb{E} W_0^2}{\beta^2 n} \leq D \gamma_Q. \end{aligned} \tag{161}$$

The last identity uses $\mathbb{E}\|VV^\top - I_d\|_F^2 = d(d + 1)$, obtained by expanding the norm and using $\mathbb{E}\|V\|^2 = d, \mathbb{E}\|V\|^4 = d(d + 2)$.

The second matrix uses two independent copies of $C_{p,n}/\kappa$. Flatten them into D by L matrices with $L = D^{p-1}$, and write $A_j = uv^\top + E_j$, $v = u^{\otimes(p-1)}$. Set

$$M_T = \tfrac{1}{2}(A_1 A_2^\top + A_2 A_1^\top), \qquad \omega = \Delta/(n\kappa^2), \qquad \gamma_T = 2\omega + \omega^2 L.$$

The centered independent errors have covariance at most ωI on ordered entries by (159), and $\mathbb{E} M_T = U$. The expansion

$$A_1 A_2^\top = U + u(E_2 v)^\top + (E_1 v) u^\top + E_1 E_2^\top$$

gives

$$\mathbb{E}(u^\top M_T u)^2 \le 1+\gamma_T, \qquad \mathbb{E}\|QM_T u\|^2 \le d\gamma_T, \qquad d^{-1}\mathbb{E}\|QM_T Q\|_F^2 \le D\gamma_T. \tag{162}$$

Here are the covariance calculations. The two parallel linear terms have variance at most ω each, and their bilinear term has second moment at most $\omega^2 L$. All their cross expectations vanish by conditioning on a centered independent error. The unsymmetrized perpendicular column is $QE_1 v + QE_1 E_2^\top u$. Its cross expectation vanishes and its second moment is at most $d\omega + d\omega\,\mathbb{E}\|E_2^\top u\|^2 \le d(\omega+\omega^2 L)$. For the perpendicular block, conditioning on E_2 gives

$$\mathbb{E}\|QE_1E_2^\top Q\|_F^2 \le \omega d\,\mathbb{E}\|E_2^\top Q\|_F^2 \le \omega^2 d^2 L.$$

Convexity of squared norm under symmetrization proves (162).

M.2 Independent pilot contractions

Put $k = p-2$. Take k disjoint pilot blocks, each of size n, and define

$$w_j = \frac{1}{n\lambda}\sum_{i \text{ in block } j} Y_i X_i, \qquad S_j = u^\top w_j, \qquad \rho = \frac{s^2}{n\lambda^2}, \qquad 1 \le j \le k.$$

The pilots are independent and identically distributed. Conditioning on $Z = u^\top X$, orthogonality of the Hermite orders and the first two noise moments give

$$\mathbb{E}[YX] = \lambda u, \qquad \mathbb{E}[Y^2 XX^\top] = (1+s^2)I + \beta U.$$

Indeed, $\mathbb{E}[Y \mid X] = f(Z)$ and $\mathbb{E}[Y^2 \mid X] = f(Z)^2 + s^2$; Gaussian symmetry makes all cross entries zero, while the parallel entry is $1+s^2+\beta$. Averaging independent rows therefore gives

$$\begin{aligned} \mathbb{E}w_j = u, \qquad \mathbb{E}S_j = 1, \qquad v := \operatorname{Var} S_j = \frac{1+s^2+\beta-\lambda^2}{n\lambda^2}, \\ \mathbb{E}\|Qw_j\|^2 = \frac{(D-1)(1+s^2)}{n\lambda^2}. \end{aligned} \tag{163}$$

Since $|\beta| \le C_P$ and $s \ge 1$, it follows that

$$0 \le v \le C_P\rho, \qquad \mathbb{E}\|w_j\|^2 = 1+v+\mathbb{E}\|Qw_j\|^2 \le 1 + C_P D\rho. \tag{164}$$

These estimates use no fourth or higher pilot moments.

Set $\mathcal{T} = \prod_{j=1}^k S_j$. Independence gives the exact identities and bound

$$\begin{aligned} \mathbb{E}\mathcal{T} = \prod_{j=1}^k \mathbb{E}S_j = 1, \qquad \mathbb{E}\mathcal{T}^2 = \prod_{j=1}^k \mathbb{E}S_j^2 = (1+v)^k, \\ \operatorname{Var}\mathcal{T} = (1+v)^k - 1, \\ \mathbb{E}\prod_{j=1}^k \|w_j\|^2 = \prod_{j=1}^k \mathbb{E}\|w_j\|^2 \le (1+C_P D\rho)^k. \end{aligned} \tag{165}$$

All these expectations are finite by (164). For $\rho \le 1$, the binomial expansion and $k \le P$ give

$$(1+v)^k - 1 = \sum_{\ell=1}^k \binom{k}{\ell} v^\ell \le \left(\sum_{\ell=1}^k \binom{k}{\ell} C_P^\ell\right)\rho \le C_P'\rho.$$

Also $(1+C_P D\rho)^k \le C_P'(1+D\rho)^k$, after increasing the constant, because $1 + C_P x \le \max\{1, C_P\}(1+x)$.

Take a fresh $C_{p,n}$ independent of every pilot, and define

$$A = C_{p,n}[\,\cdot,\cdot,w_1,\dots,w_k\,], \qquad M_H = A/\kappa.$$

The tensor is symmetric in all indices, so A is symmetric in its two free indices even though the contracted vectors differ. For any deterministic matrix F, (159) gives

$$\mathbb{E}[A \mid w_1, \dots, w_k] = \kappa \mathcal{T} U,$$
$$\mathrm{Var}(\langle A, F\rangle \mid w_1, \dots, w_k) \le \frac{\Delta}{n} \|F\|_F^2 \prod_{j=1}^{k} \|w_j\|^2. \tag{166}$$

The tested tensor is $\Pi_p(F \otimes w_1 \otimes \cdots \otimes w_k)$. Orthogonal projection contracts its norm, and the norm of the tensor product is $\|F\|_F \prod_j \|w_j\|$, proving the variance bound. Equations (165) and (166) show that M_H has finite second moments and exact mean $\mathbb{E}M_H = U$. Put

$$\gamma_H = \frac{\Delta}{n\kappa^2} \mathbb{E} \prod_{j=1}^{k} \|w_j\|^2, \qquad \zeta_H = \mathrm{Var}\, \mathcal{T} + \gamma_H.$$

For the parallel entry, conditional variance and $\mathbb{E}\mathcal{T}^2 = 1 + \mathrm{Var}\,\mathcal{T}$ give $\mathbb{E}(u^\top M_H u)^2 \le 1 + \zeta_H$. Complete u to an orthonormal basis $u, e_1, \dots, e_d$. Each entry $e_i^\top M_H u$ has conditional mean zero and unconditional second moment at most γ_H by (166). Summing over i gives the second bound below. Each $e_i^\top M_H e_j$ has the same bound; summing over i, j and dividing by d gives $d\gamma_H \le D\gamma_H$. Thus

$$\mathbb{E}(u^\top M_H u)^2 \le 1 + \zeta_H, \quad \mathbb{E}\|QM_H u\|^2 \le d\gamma_H, \quad d^{-1}\mathbb{E}\|QM_H Q\|_F^2 \le D\gamma_H. \tag{167}$$

Using (165), $\kappa^{-2} \le 2/c_p^2$, and $\Delta \le (C_P^{\mathrm{cov}} + 1)s^2$, we obtain, for $\rho \le 1$,

$$\zeta_H + D\gamma_H \le C_{f_0,P} \left\{ \rho + \frac{Ds^2}{n} (1 + D\rho)^{p-2} \right\}. \tag{168}$$

M.3 Products and full prediction risk

Every matrix just defined is symmetric, has finite second moments, and has a law invariant under $M \mapsto OMO^\top$ for orthogonal O fixing u. To check the latter assertion, rotate every design row by O while retaining its scalar noise. The joint observation law is unchanged because the design is Gaussian, O fixes u, and the noise is independent of the design. Each pilot becomes Ow_j, and $R_p(OX) = O^{\otimes p} R_p(X)$ by its generating function. Each of the three matrices consequently becomes $OMO^\top$. No invariance of the scalar noise under sign changes is needed.

We use the following exact interface of Lemma K.2. Suppose $\mathbb{E}M = U$, and its three second moments are bounded by $1 + \zeta, d\gamma, D\gamma$ as in (167). For $0 < e \le 1$, let

$$t = \lceil \log_2(100DP/e) \rceil.$$

If $\zeta + D\gamma \le e/(100Pt)$, the product $\mathcal{B}$ of t independent copies satisfies

$$\mathbb{E}\mathcal{B} = U, \qquad K = \mathbb{E}\mathcal{B}^\top \mathcal{B} = a_t U + b_t Q, \qquad a_t \le \exp(e/(100P)), \qquad \mathrm{tr}\, K < 2. \tag{169}$$

This is that lemma with $r = e/P$ and $\alpha = 1$. All three matrices have mean U. Equations (161), (162), and (167) verify its moment hypotheses with $(\zeta, \gamma) = (\gamma_Q, \gamma_Q)$, (γ_T, γ_T), and (ζ_H, γ_H), respectively.

Take P independent products $\mathcal{B}^{(1)}, \dots, \mathcal{B}^{(P)}$ and an independent final block of n observations. For $q \in J$ put

$$T_q = (\mathcal{B}^{(1)} \otimes \cdots \otimes \mathcal{B}^{(q)}) C_{q,n}, \qquad \widehat{f}(x) = \sum_{q \in J} J_q(\Pi_q T_q)(x).$$

Independence, (159), and (169) give

$$\mathbb{E}T_q = \vartheta_q u^{\otimes q}, \qquad \mathbb{E}\|T_q\|_F^2 \le \vartheta_q^2 a_t^q + \frac{\Delta}{n} \mathrm{tr}(K^{\otimes q} \Pi_q) \le \vartheta_q^2 a_t^q + \frac{\Delta}{n} 2^q.$$

Projection is a contraction, and the target is symmetric, so (*) bounds the expected order-q error by

$$\vartheta_q^2 (a_t^q - 1) + 2^q \Delta / n.$$

The distinct query chaoses are orthogonal for every realized training sample. Their errors therefore add even though the final data block is shared. Since $\sum_{q\in J}\vartheta_q^2 = 1$ and $q \le P$, the full risk is at most

$$\exp(e/100) - 1 + \frac{\Delta}{n}V_J, \qquad V_J = \sum_{q\in J} 2^q \le P2^P. \tag{170}$$

Here $\exp(e/100) - 1 \le e/99$ for $0 < e \le 1$. Thus the risk is less than e if $n \ge 2V_J\Delta/e$. All expectations are finite: the final tensor has finite second moments, each product has finite second moments by Lemma K.2, and their blocks are independent.

M.4 Allocation and degree

Proof of Theorem M.1. For every $D \ge 2$, $N \ge 0$ and $0 < \varepsilon \le 1$, set

$$e = D^{-\eta/4}, \qquad t = \lceil \log_2(100DP/e) \rceil,$$
$$n = \lfloor N/((p-1)Pt + 1) \rfloor.$$

If $n = 0$, output zero. For $n \ge 1$, form the known quantities

$$\Lambda = \min\{D^{p/2}, D\varepsilon^{-2(p-2)/(p-1)}\}, \qquad \mathcal{H} = \Lambda^2/D.$$

When $N \le \mathcal{H}$, use the quadratic matrix if $\beta \ne 0$, and output zero if $\beta = 0$. When $N > \mathcal{H}$, use the tensor matrix if $D^{p/2} \le D\varepsilon^{-2(p-2)/(p-1)}$, and use the hybrid matrix otherwise. This fixes both equality conventions. A quadratic matrix uses one block, a tensor matrix two blocks, and a hybrid matrix $k + 1 = p - 1$ blocks, each of size n. Since $p \ge 3$, every matrix uses at most $(p-1)n$ observations. Every matrix copy, every pilot within a copy, and the final tensor block use disjoint observations. The P independent products and final block therefore use at most $((p-1)Pt+1)n \le N$ observations. This specifies a predictor collection using only $f_0, p, P, D, N, \varepsilon, \eta$, not a, b, s, M, ν or u.

Fix $a, b > 0$ and $1 \le M < \infty$, and put $s = D^a, \varepsilon = D^{-b}$ for the analysis. For all sufficiently large D, $|\beta| \ge |\beta_0|/2$, so the quadratic zero-output fallback never occurs in this range. Write

$$S_Q = Ds^4, \qquad S_T = s^2D^{p/2}, \qquad S_H = Ds^2\varepsilon^{-2(p-2)/(p-1)},$$
$$S_R = s^2\Lambda, \qquad S = \min\{S_Q, S_R\} = D^B.$$

Then $S_R = \min\{S_T, S_H\}$ and $S_R^2 = \mathcal{H}S_Q$. For every asserted N, $S \le ND^{-\eta}$. If $N \le \mathcal{H}$, the possibility $S_Q > \mathcal{H}$ would imply $S_R > \mathcal{H}$ and $S > N$, a contradiction. Thus $S_Q \le \mathcal{H}$, so $S_Q \le S_R$ and $S_Q = S \le ND^{-\eta}$. If $N > \mathcal{H}$ and $S_R \le S_Q$, then $S_R = S \le ND^{-\eta}$. In the remaining case,

$$S_R = \sqrt{\mathcal{H}S_Q} < \sqrt{NS_Q} \le ND^{-\eta/2}.$$

Hence the selected quadratic scale is at most $ND^{-\eta}$, and the selected tensor or hybrid scale is at most $ND^{-\eta/2}$, for every allowed integer N.

Since $t = O_{p,P,\eta}(\log D)$ and $S \ge Ds^2$, for all sufficiently large D these integers satisfy

$$n \ge \frac{N}{C_{f_0,p,P,\eta}\log D} \ge 1, \qquad \frac{\Delta}{n} \le C_{f_0,p,P,\eta}D^{-1-\eta}\log D. \tag{**}$$

Indeed, $N \ge D^{1+2a+\eta}$ eventually exceeds $2((p-1)Pt+1)$, so the floor loses at most a factor two. The displayed final-block ratio makes its condition following (170) hold.

In the quadratic branch,

$$(D+1)\gamma_Q \le C_{f_0,p,P,\eta,M}D^{-\eta}\log D.$$

In the tensor branch, $S_T = S_R \le ND^{-\eta/2}$ and $\kappa^{-2} \le 2/c_p^2$ give

$$(D+1)\gamma_T \le C_{f_0,p,P,\eta}\left[D^{1-p/2-\eta/2}\log D + D^{-\eta}(\log D)^2\right].$$

Both are at most $e/(100Pt)$ for all sufficiently large D.

In the hybrid branch, $S_H < S_T$ gives $b < (p-1)/4$. Using $S_H = S_R \le ND^{-\eta/2}$ and $\lambda^2 \ge \varepsilon^2/2$,

$$\rho \le C_{f_0,p,P,\eta}D^{-1+2b/(p-1)-\eta/2}\log D \le C_{f_0,p,P,\eta}D^{-1/2-\eta/2}\log D.$$

Thus $\rho \le 1$ eventually. Using $(1+x)^{p-2} \le C_P(1+x^{p-2})$ in (168) now gives

$$\zeta_H + D\gamma_H \le C_{f_0,p,P,\eta}\big[D^{-1/2-\eta/2}\log D + D^{-2b(p-2)/(p-1)-\eta/2}\log D + D^{-(p-1)\eta/2}(\log D)^{p-1}\big].$$

This also is at most $e/(100Pt)$ eventually. All bounds are uniform in u, the laws $\nu \in \mathcal{N}_M$, and the allowed integers N: the law enters only through its stated moment bounds, and the estimates remain valid when the block size increases. Taking the maximum of the finitely many dimension thresholds gives $D_0(f_0, p, P, a, b, \eta, M)$. Equations (169) and (170) give the stated $D^{-\eta/4}$ risk bound.

For completeness, the matrix label degrees are at most $2, 2, p-1$ for Q, T, H, respectively. Their joint training degrees are at most $4, 2(p+1), 3p-3$. In the last case each pilot contributes joint degree two and the fresh tensor contributes degree $p+1$; the divisor κ is deterministic. With $d_p = \max\{4, 2(p+1), 3p-3\}$, the final predictor has label degree at most $P(p-1)t+1$ and joint degree at most

$$Pd_pt + 2P + 1 = O_{p,P,\eta}(\log D).$$

Normal-ordered tensors of order at most P have $O_P(D^P)$ ordered entries. Forming them, taking the displayed contractions and products, and evaluating the query therefore uses polynomial real-arithmetic work and storage for fixed P. The scalar coefficients and branch comparisons use only $f_0, p, P, D, N, \varepsilon, \eta$. Neither the noise law nor its scale or fourth-moment bound is used. This completes the proof. □

N Deduction of the pure-Hermite theorem

Proof. Corollary A.3 proves (5) averaged over the finite prior $u_j \in \{-D^{-1/2}, D^{-1/2}\}$ with independent signs. For every learner, its worst unit-teacher risk is at least its average over this prior. Taking the infimum gives the required minimax bound. Rescaling the labels by known $s > 0$ preserves (3) and label degree.

Apply Theorem B.1 with $\epsilon = D^{-\eta/4}$. If $b = 1 + 4a$, its quadratic construction requires at most

$$C_{p,\eta}D^{b+\eta/4}(\log D)^2$$

observations for all sufficiently large D. If $b = p/2 + 2a$, its tensor construction requires at most a constant times

$$D^{1+2a+\eta/4}(\log D)^2 + D^{p/2+2a+\eta/8}(\log D)^{3/2}.$$

Since $p/2 \ge 1$, both terms are at most $C_{p,\eta}D^{b+\eta/4}(\log D)^2$. This includes equality of the two branches. The ceilings and final tensor block are included in the sample bound of Theorem B.1. The ratio $D^{3\eta/4}/(\log D)^2$ tends to infinity, so either required sample count is eventually at most $D^{b+\eta}$. Extra observations can be ignored. The degree in that theorem is exactly L_+ and its expected risk is at most $D^{-\eta/4}$, proving (6). □

O Fixed links with a quadratic-label linear signal

Let $3 \le p \le P$ be integers and let

$$f(z) = \sum_{q=p}^{P} c_q h_q(z), \qquad c_p \ne 0, \qquad \sum_{q=p}^{P} c_q^2 = 1, \qquad \alpha = \mathbb{E}[Zf(Z)^2] \ne 0, \quad Z \sim N(0,1). \tag{171}$$

The coefficients are fixed independently of dimension and noise. The orders may have either parity. The observations are independent copies of $X \sim N(0, I_D)$ and $Y = f(u^\top X) + s\xi$, where u is an unknown unit vector. The noises are iid and independent of all inputs. The query $x \sim N(0, I_D)$ is independent. The target is $F_u(x) = f(u^\top x)$. For $M \ge 1$, write $\mathcal{N}_M$ for the laws with mean zero, variance one and fourth moment at most M.

The recurrence $zh_q = \sqrt{q+1}h_{q+1} + \sqrt{q}h_{q-1}$ gives $\alpha = 2\sum_{q=p}^{P-1}\sqrt{q+1}c_qc_{q+1}$. Thus nonzero α requires both parities, but both parities alone do not ensure it. Warm-start tensor contraction in

Lemma F.4 of Damian et al. (2024), arXiv version 2, supplies prior ancestry for the hybrid recovery scale below. The assertion here concerns polynomial prediction and its matching label-degree lower bound.

Theorem O.1. *Fix* (171) *and* $\eta > 0$. *For* $a > 0$ *put*

$$B = \min\left\{1 + \left(4 - \frac{2}{p-1}\right)a, \frac{p}{2} + 2a\right\}, \quad \tau = \min\{2a, \eta/2\},$$
$$\Gamma_P = 3P + 6, \quad L = \left\lfloor D^{\tau/(4\Gamma_P)} \right\rfloor.$$

Let $C_P^0 = \max\{1, (2\lfloor P/2 \rfloor - 1)!!\}$. *For Gaussian noise,* $s = D^a$, *integers* $0 \le N \le D^{B-\eta}$, *and*

$$D \ge \left\lceil \max\{2, (4P)^{4\Gamma_P/\tau}, 2^{2/\tau}\} \right\rceil,$$

every predictor polynomial of degree at most L *in the training labels, with arbitrary jointly measurable design, query and independent-seed coefficients, has sign-prior expected risk at least*

$$1 - C_P^0 D^{-\lceil p/2 \rceil} - 2D^{-\tau/2}. \tag{172}$$

The teacher in this prior is uniform on $\{-1, 1\}^D/\sqrt{D}$. *The same lower holds for worst-teacher risk and for joint polynomials of total degree at most* L.

There is also a collection $\widehat{f}_{f,p,P,D,N,\eta}$, *indexed by integers* $D \ge 2$, $N \ge 0$, *whose coefficients depend only on the displayed indices and supplied link. For every fixed* $a > 0$ *and finite* $M \ge 1$, *there is* $D_0 = D_0(f, p, P, a, \eta, M)$ *such that, for* $D \ge D_0$, $s = D^a$ *and every integer* $N \ge \lceil D^{B+\eta} \rceil$,

$$\sup_{\nu \in \mathcal{N}_M} \sup_{\|u\|=1} \mathbb{E}_{\nu,u}\big(\widehat{f}_{f,p,P,D,N,\eta}(x) - F_u(x)\big)^2 \le D^{-\eta/4}. \tag{173}$$

Its total degree in the training design, labels and query is $O_{p,P,\eta}(\log D)$, *and its real-arithmetic work and storage are polynomial in* N, D *for fixed* P. *The learner receives neither* a, s, M *nor the noise law. No uniformity as* c_p *or* α *approaches zero is asserted.*

Proof. We first prove the lower bound and then construct the upper.

O.1 The finite conditional certificate

For the lower bound, let A be the Gaussian design, let U be the sign-prior teacher, and set $V = Y/s = T + G$, where $T_i = f(A_i^\top U)/s$ and $G \sim N(0, I_N)$ is independent. Rescaling labels preserves their polynomial degree. Put $\mathcal{H} = \bigoplus_{q=p}^P \mathcal{H}_q$, allowing zero coefficients, and define

$$m_\nu = \mathbb{E}_U T^\nu, \qquad v_\nu = \mathbb{E}_U[F_U T^\nu],$$
$$\kappa_\nu = v_\nu - \sum_{\substack{\omega \le \nu \\ \omega \ne \nu}} \binom{\nu}{\omega} \kappa_\omega m_{\nu-\omega}, \qquad E_k = \mathbb{E}_A \sum_{|\nu|=k} \frac{\|\kappa_\nu\|_{\mathcal{H}}^2}{\nu!}. \tag{174}$$

In particular $\kappa_0 = \mathbb{E}_U F_U$ need not vanish. Gaussian chaos orthogonality gives

$$\|F_u\|_2^2 = 1, \qquad \langle F_u, F_v \rangle = \sum_{q=p}^P c_q^2 (u^\top v)^q. \tag{175}$$

Every allowed predictor has risk at least

$$1 - \sum_{k=0}^L E_k. \tag{176}$$

Here is the certificate with its nonzero constant term retained. For a finite-risk predictor, conditioning on a fixed design gives a finite mixture of identity-covariance Gaussian label laws. Its monomial Gram matrix is positive definite, since a nonzero polynomial cannot vanish Lebesgue almost everywhere. Its positive least eigenvalue and Tonelli's theorem imply that every coefficient belongs to

$L^2(\gamma_D)$ for almost every design and seed. Project these coefficients onto $\mathcal{H}$. The target belongs to $\mathcal{H}$, so this does not increase risk. Integration against its finite product-Hermite basis makes the projection measurable; set its coefficients to zero on the exceptional null set.

At a fixed design and seed, expand this polynomial as $\sum_{|\nu|\le L} b_\nu q_\nu(V)$, where $q_\nu(y) = \prod_i \mathrm{He}_{\nu_i}(y_i)/\sqrt{\nu!}$ and $b_\nu \in \mathcal{H}$. The finite Hermite shift identity (19) gives

$$\mathbb{E}\langle F_U, \widehat{F}(T+G)\rangle = \sum_\nu \langle z_\nu, b_\nu\rangle, \qquad z_\nu = v_\nu/\sqrt{\nu!}.$$

Jensen's inequality in U at fixed G, then Gaussian orthogonality, gives

$$\mathbb{E}\|\widehat{F}(T+G)\|^2 \ge \sum_\omega \|d_\omega\|^2, \quad d = Jb, \quad J_{\omega,\nu} = \mathbf{1}_{\{\omega\le\nu\}} m_{\nu-\omega}\sqrt{\frac{\omega!}{\nu!}}\binom{\nu}{\omega}.$$

The recursion (174) says $z = J^\top w$ for $w_\omega = \kappa_\omega/\sqrt{\omega!}$, including $\omega = 0$. Completing squares therefore bounds conditional risk below by $1 + \sum_\omega \|d_\omega - w_\omega\|^2 - \sum_\omega \|w_\omega\|^2$. Averaging proves (176). Every coordinate of $\kappa_\nu(A)$ is a polynomial in A, by the finite sign prior and recursion; hence all energies are finite. This also handles infinite-risk predictors.

For two independent sign teachers, their inner product is $D^{-1}\sum_{j=1}^D \epsilon_j$ with iid uniform signs. It follows from (175) that

$$E_0 = \sum_{\substack{p\le q\le P\\ q \text{ even}}} c_q^2 \mathbb{E}\left(D^{-1}\sum_{j=1}^D \epsilon_j\right)^q \le C_P^0 D^{-\lceil p/2\rceil}. \tag{177}$$

For even q, expand the moment. A surviving coordinate assignment has even multiplicity at each coordinate and therefore admits a pairing of its q positions. There are $(q-1)!!$ pairings and at most $D^{q/2}$ assignments per pairing, giving the bound $(q-1)!!D^{-q/2}$ after normalization. Odd moments vanish by sign symmetry. Summing the normalized coefficients proves (177).

O.2 Mixed-order diagrams and coordinate loss

For every integer $k \ge 1$ we claim

$$E_k \le (2Pk)^{\Gamma_P k} s^{-2k} \sum_{r=1}^{\min\{N,k\}} N^r D^{-\delta(k,r)}, \qquad \delta(k,r) = \max\left\{r, \left(p-\frac{1}{2}\right)r - \frac{p-1}{2}k + \frac{p}{2}\right\}. \tag{178}$$

An empty sum is zero. We verify all changes caused by the absence of a parity restriction.

Write $J_q(H) = \langle H, R_q\rangle/\sqrt{q!}$ for the normal-ordered Gaussian tensor isometry. Then $F_U = \bigoplus_q c_q J_q(U^{\otimes q})$. Differentiating the Gaussian generating functions proves that a product of normal-ordered factors is a sum over coordinate pairings with no pair internal to a factor. Independence confines pairs to a design row. This is the tensor expansion in (22), separately at each order. Products of coordinates of the sign teacher reduce to sign characters, possibly with empty support.

The joint-cumulant partition identity (20) holds for these characters of either parity, including constants: it follows by extracting the coefficient using each variable once from the logarithm of their joint moment generating function. Its target derivative gives exactly the recursion (174). At order $k \ge 1$, a constant argument has zero cumulant: its exponential factors out, adding a separate linear term to the logarithm. We may therefore delete all terms with an empty reduced support. Nonzero cumulants require empty total symmetric difference and a connected character-intersection graph, by the partition formula and independence of disjoint sets of signs, respectively. Their absolute value is at most $(k+1)^{k+1}$ by (21). There is no remaining restriction to odd k.

Fix $|\nu| = k$ with r occupied rows. Expand the squared tensor cumulant. Orthogonality pairs only equal target orders q. For fixed label orders $j_1, \dots, j_{2k}$ in the two replicas, put $R = \sum_v j_v$ and $e = R/2 + q$. If a row has an odd total number of positions its term vanishes. Otherwise there are $R/2$ Gaussian pairs and q target bridges identifying corresponding target tensor positions. Their

scalar normalization is exactly

$$s^{-2k} D^{-e} c_q^2 \prod_{v=1}^{2k} \frac{c_{j_v}}{\sqrt{j_v!}}. \tag{179}$$

Each Gaussian pair or target bridge initially receives one coordinate variable. These $e \le P(k+1) \le 2Pk$ variables form h equality classes. After deleting zero cumulants, form a bipartite multigraph B_{full} with left vertices the r rows and target, right vertices the h coordinate classes, and one edge for every initial variable. Put $\ell = e - h$.

The character-intersection graph in either replica is connected. Shared coordinates give paths between its factors in B_{full}, and identifying factors in the same row preserves connection. Coordinates erased by sign reduction still attach to a row or target. Thus B_{full} is connected and its cycle rank is $\ell - r \ge 0$.

Mark crossing-replica Gaussian pairs and all target bridges, and remove isolated vertices to obtain B_{mark}. Each of its coordinate vertices has positive even degree: its marked degree is congruent to the number of occurrences in one replica, whose total reduced character is empty. The target has marked degree $q \ge p$. A singleton row contains one normal-ordered factor in each replica. Every pair must cross replicas, forcing equal orders and marked degree at least p. Any other nonisolated row has degree at least one. If r_1 is the number of singleton rows, the degree-sum formula and at least one marked component give

$$\operatorname{rank}_{\text{cyc}}(B_{\text{mark}}) \ge 1 + \frac{(p-2)(r_1+1) - (r-r_1)}{2} = \frac{(p-1)r_1 - r + p}{2}.$$

An isolated nonsingleton has no contribution, so charging it -1 in this inequality remains valid. Cycle rank is monotone under taking subgraphs, because independent cycle vectors remain independent in the larger graph. Also $k \ge r_1 + 2(r - r_1)$, so $r_1 \ge 2r - k$. Consequently $\ell - r$ is at least the last display, and

$$\ell \ge r, \qquad \ell \ge \left(p - \frac{1}{2}\right) r - \frac{p-1}{2} k + \frac{p}{2}.$$

This proves the asserted coordinate loss.

For completeness, at fixed orders there are at most $(2Pk)^{Pk}$ Gaussian pairings: at most $R^{R/2}$ partner choices suffice before imposing row constraints. There are at most $e^e \le (2Pk)^{2Pk}$ equality partitions, by mapping each variable to its block's first element. A partition with h classes has $(D)_h \le D^h$ coordinate assignments, or zero if $h > D$. Thus its dimension factor is at most $D^{-\delta(k,r)}$. The product of the two cumulants is at most $(k+1)^{2(k+1)} \le (2Pk)^{4k}$. Drop factorial denominators in (179). Summing absolute label coefficients costs at most $(\sum_q |c_q|)^{2k} \le P^k \le (2Pk)^k$; target coefficients sum to one. For each ν, the result is bounded by $(2Pk)^{(3P+5)k} s^{-2k} D^{-\delta(k,r)}$. Finally there are at most $\binom{N}{r}\binom{k-1}{r-1} \le N^r 2^{k-1}$ multiindices with r occupied rows. Drop $1/\nu! \le 1$ and absorb this last factor into $(2Pk)^k$. This proves (178). All sums are finite and include repeated coordinates, repeated rows and unequal label orders in the two replicas.

O.3 The lower exponent

Suppose $N > 0$ and $N \le D^{B-\eta}$. Put $q_0 = p - 3/2$ and $c = B - \eta - 1$. A term in (178) has dimension exponent at most

$$e_{k,r} = (B-\eta)r - 2ak - \delta(k,r).$$

If $c < 0$, the first branch gives $e_{k,r} \le cr - 2ak \le -\tau k$. If $0 \le c \le q_0$, combine the branches with weight c/q_0 on the second to obtain

$$\delta(k,r) \ge (1+c)r - \frac{c(p-1)}{2p-3} k + \frac{cp}{2p-3}.$$

Since $c \le (4 - 2/(p-1))a - \eta$,

$$2a - \frac{c(p-1)}{2p-3} \ge \frac{\eta(p-1)}{2p-3} \ge \eta/2,$$

and again $e_{k,r} \le -\tau k$. If $c > q_0$, then $a > (p-1)/4$; otherwise the preceding upper bound on c would be less than q_0. In this case $B = p/2 + 2a$ and $v = 2a - (p-1)/2 > \eta$. The second branch gives

$$e_{k,r} \le (v-\eta)r - vk - p/2 = -v(k-r) - \eta r - p/2 \le -\eta k - p/2 \le -\tau k.$$

Thus $E_k \le (4Pk)^{\Gamma_P k} D^{-\tau k}$. For $1 \le k \le L$ and the theorem's stated dimension bound, $(4Pk)^{\Gamma_P} \le D^{\tau/2}$. Hence $E_k \le D^{-\tau k/2}$, and their sum is at most $2D^{-\tau/2}$ since $D^{-\tau/2} \le 1/2$. Equations (176) and (177) prove (172). If $N = 0$ the positive-order energies vanish, giving the same conclusion. A prior average is no greater than worst-teacher risk for each learner. A joint polynomial of total degree L has label degree at most L. These observations give both stated extensions.

O.4 Tensors and independent quadratic-label pilots

We now prove the upper. Constants with subscript f, M below may also depend on the fixed endpoint orders p, P, and may increase between occurrences. Define the fixed covariance constant

$$C_f^{\mathrm{cov}} = \max_{0 \le j \le P} \mathbb{E}[f(Z)^2 h_j(Z)^2], \qquad \Delta = C_f^{\mathrm{cov}} + s^2.$$

For an independent block of $n \ge 1$ observations let $C_{q,n} = n^{-1} \sum_i Y_i R_q(X_i)/\sqrt{q!}$. At every order $p \le q \le P$,

$$\mathbb{E}C_{q,n} = c_q u^{\otimes q}, \qquad \mathrm{Cov}(C_{q,n}) \preceq (\Delta/n)\Pi_q. \tag{180}$$

To verify the covariance without a parity assumption, rotate u to the first coordinate and write an order-q chaos as $g(Z,W) = \sum_{j=0}^{q} h_j(Z) g_{q-j}(W)$. Distinct orders in W are orthogonal, so $\mathbb{E}[f(Z)^2 g(Z,W)^2] = \sum_j \mathbb{E}[f(Z)^2 h_j(Z)^2] \|g_{q-j}\|_2^2 \le C_f^{\mathrm{cov}} \|g\|_2^2$. Independent centered noise adds $s^2\|g\|_2^2$. The Gaussian tensor isometry gives the mean and the covariance bound, and independent rows divide variance by n. Only the noise's first two moments were used.

Set $U = uu^\top$, $Q = I - U$, $d = D - 1$, and $k = p - 2$. For k independent pilot blocks put

$$w_j = \frac{1}{n\alpha} \sum_{i \in B_j} Y_i^2 X_i, \qquad S_j = u^\top w_j, \qquad \rho = s^4/n.$$

The identities $\mathbb{E}[Y^2 \mid Z] = f(Z)^2 + s^2$ and $\mathbb{E}[X \mid Z] = Zu$ give $\mathbb{E}w_j = u$ and $\mathbb{E}S_j = 1$. More explicitly, the parallel mean is $\mathbb{E}[Y^2 Z]/\alpha = 1$, and perpendicular means vanish by independence and centering of the other Gaussian coordinates. For $s \ge 1$ and any law in $\mathcal{N}_M$, convexity gives

$$\mathbb{E}[Y^4 \mid Z] \le 8(f(Z)^4 + s^4 M) \le 8s^4(f(Z)^4 + M).$$

Define $A_{f,M} = 8\alpha^{-2} \max\{\mathbb{E}[(f(Z)^4 + M)Z^2], \mathbb{E}[f(Z)^4 + M]\}$. Parallel variance and perpendicular second moments therefore obey

$$0 \le v := \mathrm{Var}\, S_j \le A_{f,M}\rho, \quad \mathbb{E}\|Qw_j\|^2 \le dA_{f,M}\rho, \quad \mathbb{E}\|w_j\|^2 \le 1 + DA_{f,M}\rho.$$

Independence gives, for $\mathcal{T} = \prod_{j=1}^{k} S_j$,

$$\mathbb{E}\mathcal{T} = 1, \qquad \mathrm{Var}\,\mathcal{T} = (1+v)^k - 1, \qquad \mathbb{E}\prod_j \|w_j\|^2 \le (1 + DA_{f,M}\rho)^k.$$

For $\rho \le 1$, binomial expansion bounds the variance by $C_{f,M}\rho$ and the last expression by $C_{f,M}(1 + D\rho)^k$.

Take a fresh $C_{p,n}$, independent of all these pilots, and put $M_H = c_p^{-1} C_{p,n}[\cdot, \cdot, w_1, \dots, w_k]$. It is symmetric in its two free indices. Conditionally on the pilots, its mean is $\mathcal{T}U$. For any deterministic matrix F, (180) bounds its conditional tested variance by

$$\frac{\Delta}{nc_p^2} \|F\|_F^2 \prod_j \|w_j\|^2,$$

because symmetrizing $F \otimes w_1 \otimes \cdots \otimes w_k$ cannot increase tensor norm. Define

$$\gamma_H = \frac{\Delta}{nc_p^2} \mathbb{E}\prod_j \|w_j\|^2, \qquad \zeta_H = \mathrm{Var}\,\mathcal{T} + \gamma_H.$$

The unconditional mean is U. Testing the parallel entry and the entries of the perpendicular column and block gives

$$\mathbb{E}(u^\top M_H u)^2 \le 1+\zeta_H, \quad \mathbb{E}\|QM_H u\|^2 \le d\gamma_H, \quad d^{-1}\mathbb{E}\|QM_H Q\|_F^2 \le D\gamma_H. \tag{181}$$

Indeed the latter entries have conditional mean zero, and their numbers are d and d^2, respectively. Since $\Delta \le (C_f^{\mathrm{cov}}+1)s^2$, for $\rho \le 1$,

$$\zeta_H + D\gamma_H \le C_{f,M}\left[\rho + \frac{Ds^2}{n}(1+D\rho)^{p-2}\right]. \tag{182}$$

The other matrix uses two independent copies of $C_{p,n}/c_p$, flattened as $A_j = uv^\top + E_j$ with $v = u^{\otimes(p-1)}$, and is $M_T = (A_1A_2^\top + A_2A_1^\top)/2$. Put $\omega = \Delta/(nc_p^2)$ and $\gamma_T = 2\omega + \omega^2 D^{p-1}$. The errors are centered and independent, with covariance at most ωI on ordered entries. Expand $A_1A_2^\top = U + u(E_2v)^\top + (E_1v)u^\top + E_1E_2^\top$. Parallel linear variances sum to at most 2ω and the bilinear variance is at most $\omega^2 D^{p-1}$. Their cross expectations vanish by conditioning on a centered error. The perpendicular column has second moment at most $d\omega + d\omega^2 D^{p-1}$; its cross term vanishes for the same reason. Conditioning on E_2 bounds the perpendicular block's squared norm by $\omega d\,\mathbb{E}\|E_2^\top Q\|_F^2 \le \omega^2 d^2 D^{p-1}$. Convexity under symmetrization therefore gives

$$\mathbb{E}M_T = U, \quad \mathbb{E}(u^\top M_T u)^2 \le 1+\gamma_T, \quad \mathbb{E}\|QM_T u\|^2 \le d\gamma_T, \quad d^{-1}\mathbb{E}\|QM_T Q\|_F^2 \le D\gamma_T. \tag{183}$$

O.5 Full prediction risk and allocation

Both matrices have finite second moments and are invariant under conjugation by orthogonal maps fixing u. For this invariance, rotate all Gaussian design rows while retaining their independent scalar noises. Each pilot rotates as a vector, and the normal-ordered tensors transform tensorially by their generating function. The observation law is unchanged, with no symmetry assumption on the noise.

For $0 < e \le 1$ put $t = \lceil \log_2(100DP/e)\rceil$. Lemma K.2, with its mean parameter equal to one and $r = e/P$, applies whenever $\zeta + D\gamma \le e/(100Pt)$. Equations (181) and (183) verify its three moment hypotheses, with $(\zeta,\gamma) = (\zeta_H,\gamma_H)$ or (γ_T,γ_T). For a product $\mathcal{B}$ of t fresh matrix copies it gives

$$\mathbb{E}\mathcal{B} = U, \quad K = \mathbb{E}\mathcal{B}^\top\mathcal{B} = a_t U + b_t Q, \quad a_t \le \exp(e/(100P)), \quad \operatorname{tr} K < 2.$$

Take P independent such products and a fresh final block of size n. For every q with $c_q \ne 0$ put $T_q = (\mathcal{B}^{(1)} \otimes \cdots \otimes \mathcal{B}^{(q)})C_{q,n}$ and output $\sum_q J_q(\Pi_q T_q)(x)$. Independence gives

$$\mathbb{E}T_q = c_q u^{\otimes q}, \qquad \mathbb{E}\|T_q\|_F^2 \le c_q^2 a_t^q + (\Delta/n)\operatorname{tr}(K^{\otimes q}\Pi_q) \le c_q^2 a_t^q + 2^q\Delta/n.$$

Here the covariance trace inequality follows from (180); $\Pi_q \preceq I$ and $K^{\otimes q} \succeq 0$ give the last bound. Symmetric projection contracts norm and fixes the target. Distinct query chaoses are orthogonal on every training outcome, regardless of parity and shared final data. Thus full expected risk is at most

$$\exp(e/100) - 1 + V_f\Delta/n, \qquad V_f = \sum_{q:c_q\ne 0} 2^q \le P2^P. \tag{184}$$

This is below e if $n \ge 2V_f\Delta/e$, since $\exp(e/100) - 1 \le e/99$.

To define the promised learner at every index, put

$$e = D^{-\eta/4}, \quad t = \lceil \log_2(100DP/e)\rceil, \quad n = \left\lfloor \frac{N}{(p-1)Pt+1} \right\rfloor, \quad H = D^{p-1/2}.$$

If $n = 0$, output zero. Otherwise choose hybrid matrices when $N^2 \le D^{2p-1}$ and tensor Gram matrices otherwise. These conditions are equivalent to $N \le H$ and $N > H$. A hybrid matrix uses $p-1$ blocks, and a Gram matrix two. All matrices, all pilots within them, and the final tensor block use disjoint observations. Hence this allocation uses at most $((p-1)Pt+1)n \le N$ observations and depends only on the claimed inputs.

For the analysis fix a, M and write

$$S_H = Ds^{4-2/(p-1)}, \quad S_T = s^2 D^{p/2}, \quad d_0 = \frac{2p-3}{p-1}, \quad S = \min\{S_H, S_T\} = D^B.$$

Then $1 < d_0 < 2$ and $S_H/H = (S_T/H)^{d_0}$. At every asserted sample count $S \le ND^{-\eta}$. If $N \le H$, the possibility $S_T > H$ would force $S_H > H$ and $S > N$, a contradiction. Thus $S_H \le S_T$ and $S_H = S \le ND^{-\eta}$. It also follows that $a \le (p-1)/4$ in this branch. If $N > H$ and $S_T \le S_H$, then $S_T \le ND^{-\eta}$. Otherwise $S_H = S$ and

$$S_T = H^{1-1/d_0} S_H^{1/d_0} < N^{1-1/d_0}(ND^{-\eta})^{1/d_0} = ND^{-\eta/d_0}.$$

Thus the selected tensor scale is at most $ND^{-\eta/2}$. These conclusions hold for all stated N, not only the smallest.

Since $S \ge Ds^2$, eventually the integer floor gives $n \ge N/(C_{p,P,\eta} \log D)$ and $\Delta/n \le C_{f,p,P,\eta} D^{-1-\eta} \log D$. This meets the final-block requirement after (184). In the hybrid branch, $n \ge S_H D^{\eta}/(C_{p,P,\eta} \log D)$, so

$$\rho \le C_{p,P,\eta} D^{-1+2a/(p-1)-\eta} \log D \le C_{p,P,\eta} D^{-1/2-\eta} \log D.$$

In particular $\rho \le 1$ eventually. Using $(1+x)^{p-2} \le C_p(1+x^{p-2})$ in (182) yields

$$\zeta_H + D\gamma_H \le C_{f,p,P,\eta,M}\big[D^{-1/2-\eta} \log D + s^{-2(p-2)/(p-1)} D^{-\eta} \log D + D^{-(p-1)\eta} (\log D)^{p-1}\big].$$

It is at most $e/(100Pt)$ for all sufficiently large D. In the tensor branch, $S_T \le ND^{-\eta/2}$ gives

$$(D+1)\gamma_T \le C_{f,p,P,\eta}[D^{1-p/2-\eta/2} \log D + D^{-\eta}(\log D)^2],$$

which is also at most $e/(100Pt)$ eventually. All constants are uniform in the teacher and laws in $\mathcal{N}_M$, and increasing n preserves the moment bounds. Taking a common dimension threshold proves (173) uniformly over all the required sample counts.

Each pilot has label degree two and joint training degree three. A hybrid matrix therefore has respective degrees $2p-3$ and $4p-5$. The Gram matrix has degrees two and $2(p+1)$. With $d_p = \max\{4p-5, 2(p+1)\}$, the final predictor has label degree at most $P(2p-3)t+1$ and joint degree at most

$$Pd_p t + 2P + 1 = O_{p,P,\eta}(\log D).$$

Its only normalizations involving the link are deterministic division by α and c_p. No moment above the fourth is needed: every pilot has finite second moment, independent pilot products use only these moments, fresh tensor blocks have the verified covariance, and the matrix lemma uses independent factors with finite second moments. The final block is independent of all products, and query orthogonality removes cross-order risk terms. Tensors of order at most P have $O_P(D^P)$ ordered entries; the finite sums, contractions and matrix products therefore have polynomial real-arithmetic work and storage in N, D for fixed P. This completes the proof of Theorem O.1. □

O.6 Prediction without the link coefficients

The coefficient divisions in the preceding upper bound can be replaced by polynomial calibration. The collection below is common to the fixed links in (171); the lowest nonzero order is still supplied.

Theorem O.2. *Fix integers $3 \le p \le P$ and $\eta > 0$. There are predictors $\widehat{f}^{\mathrm{imp}}_{p,P,D,N,\eta}$ for every integer $D \ge 2, N \ge 0$, whose coefficients depend only on these indices. Their joint degrees in the training design, labels and query are $O_{p,P,\eta}(\log^3(eD))$, and their real-arithmetic work and storage are polynomial in N, D for fixed p, P, η. For every fixed link (171), fixed $a > 0$ and finite $M \ge 1$, put $s = D^a$ and*

$$b = \min\{1 + (4 - 2/(p-1))a,\ p/2 + 2a\}.$$

There is $D_0 = D_0(f, p, P, a, \eta, M)$ such that, for $D \ge D_0$ and every integer $N \ge \lceil D^{b+\eta} \rceil$,

$$\sup_{\mathcal{L}(\xi) \in \mathcal{N}_M} \sup_{\|u\|=1} \mathbb{E}\big(\widehat{f}^{\mathrm{imp}}_{p,P,D,N,\eta}(x) - f(u^\top x)\big)^2 \le D^{-\eta/8}.$$

The learner receives neither the coefficients, a, s, M, nor the noise law or positive lower bounds on $|c_p|, |\alpha|$. The threshold need not be uniform as these fixed nonzero signals approach zero.

We first separate the parallel variance in the calibration step.

Lemma O.3. *Use D, u, U, Q, d, r, t as in Lemma K.2, and put $\nu = r/(1000t)$. Let $\widetilde{M}$ be symmetric, have finite second moments, be invariant under orthogonal conjugations fixing u, and have mean μU, where $0 < \mu_0 \le |\mu| \le B_*$. Suppose the three moments of $\widetilde{M}/\mu$ satisfy $A_0 \le 1+\zeta_0$, $B_0 \le d\gamma_0$, $C_0 \le D\gamma_0$, for $\zeta_0, \gamma_0 \ge 0$ with $\zeta_0 + D\gamma_0 \le \nu$. Independently take $2m-1$ independent trace estimates, each of mean μ and variance at most $v \ge 0$, where*

$$v/\mu_0^2 \le \nu, \qquad m \ge \left\lceil 4B_*^2\mu_0^{-2}\log(1/\nu)\right\rceil.$$

Form R_m by (124) with $H = 2B_*$. Then $M = R_m\widetilde{M}$ satisfies Lemma K.2 with mean θU, $0 \le 1-\theta \le \nu$, and $\zeta = 3v/\mu^2 + 2\zeta_0$, $\gamma = 2\gamma_0$.

Proof. Since $\nu \le 1/1000$, Lemma K.1 applies and gives $\theta = \mu\mathbb{E}R_m \in [1-\nu, 1]$ and $\mu^2\mathbb{E}R_m^2 \le 1 + 3z < 2$, where $z = v/\mu^2 \le \nu$. Independence multiplies each normalized matrix second moment by $\mu^2\mathbb{E}R_m^2$. Thus $A \le 1 + 3z + 2\zeta_0$, $B_1 \le 2d\gamma_0$ and $C \le 2D\gamma_0$, while $\zeta + D\gamma \le 3z + 2(\zeta_0 + D\gamma_0) \le 5\nu < r/(100t)$. Independence also gives the stated mean, preserves the invariance, and gives finite second moments; scalar multiplication preserves symmetry. These verify all hypotheses of the matrix-product lemma. □

Proof of Theorem O.2. All parameters used by the learner are the following deterministic functions of its indices:

$$\begin{aligned}
&\ell = \log(eD), && \epsilon = D^{-\eta/8}, && r = \epsilon/P, && t = \lceil \log_2(100D/r)\rceil,\\
&\mu_0 = \ell^{-1/2}, && B_* = (2P)^P, && \nu = r/(1000t), && m = \lceil 4B_*^2\mu_0^{-2}\log(1/\nu)\rceil,\\
&L_0 = 2(p-1)mPt + 1, && n = \lfloor N/L_0\rfloor.
\end{aligned}$$

If $n = 0$, output zero. Otherwise choose the hybrid type when $N^2 \le D^{2p-1}$, and the tensor Gram type otherwise. A primitive block contains n observations. Define raw matrices using disjoint primitive blocks by

$$\widetilde{M}_H = C_{p,n}[\cdot,\cdot,\widetilde{w}_1,\dots,\widetilde{w}_{p-2}], \quad \widetilde{w}_j = n^{-1}\sum_{i\in B_j} Y_i^2 X_i, \qquad \widetilde{M}_T = (C_1C_2^\top + C_2C_1^\top)/2.$$

Here C_1, C_2 are independent copies of $C_{p,n}$ flattened as D by D^{p-1} matrices. The types use $p-1$ and two primitive blocks, respectively. For each calibrated matrix take one raw matrix and $2m-1$ independent raw traces of the selected type, form R_m as above, and multiply the raw matrix by R_m. Take P independent products $\mathcal{B}^{(j)}$ of t such calibrated matrices, and a fresh final primitive block. Output

$$\widehat{f}(x) = \sum_{q=p}^{P} J_q(\Pi_q T_q)(x), \qquad T_q = (\mathcal{B}^{(1)}\otimes\cdots\otimes\mathcal{B}^{(q)})C_{q,n}. \tag{185}$$

Every order is included, even when its unknown coefficient is zero. All raw objects and the final block use disjoint observations. Their total count is at most $L_0 n \le N$, since $p-1 \ge 2$. Unused observations are ignored. No formula uses the unknown parameters.

We verify its moments. Retain $\Delta = C_f^{\mathrm{cov}} + s^2$ from (180). The raw means are $\mu_H U$ and $\mu_T U$, where $\mu_H = c_p\alpha^{p-2}$ and $\mu_T = c_p^2$. The recurrence for α and Cauchy–Schwarz give $|\alpha| \le 2\sqrt{P}$; hence $|\mu_H|, |\mu_T| \le B_*$. Eventually $\mu_0 \le \min\{|\mu_H|, \mu_T\}$ for every fixed link in the theorem.

Conceptual division of the raw hybrid by μ_H gives exactly M_H in (181); this division is not performed by the learner. Put $h_H = \zeta_H + D\gamma_H$ there. Conditionally on its normalized pilots, the trace has mean $\mathcal{T} = \prod_j u^\top w_j$. The conditional tested covariance proved before (181), with $F = I$ and $\|I\|_F^2 = D$, gives conditional trace variance whose expectation is at most $D\gamma_H$. Total variance therefore gives

$$\mathrm{Var}(\mathrm{tr}\,\widetilde{M}_H) \le \mu_H^2(\mathrm{Var}\,\mathcal{T} + D\gamma_H) \le \mu_H^2 h_H.$$

All three normalized matrix moments have parameters $(\zeta_0, \gamma_0) = (\zeta_H, \gamma_H)$. For the Gram type they have $(\zeta_0, \gamma_0) = (\gamma_T, \gamma_T)$ by (183); put $h_T = (D+1)\gamma_T$. For its trace write $C_j = c_p u^{\otimes p} + E_j$ before flattening. The centered independent errors have covariance at most $(\Delta/n)I$. Then

$$\mathrm{tr}\,\widetilde{M}_T = c_p^2 + c_p\langle u^{\otimes p}, E_1 + E_2\rangle + \langle E_1, E_2\rangle.$$

The three centered terms have zero cross expectations by conditioning on a centered error. The two linear variances are at most $c_p^2\Delta/n$ each, and the bilinear second moment is the trace of the covariance product, at most $\Delta^2 D^p/n^2$. Consequently

$$\mathrm{Var}(\mathrm{tr}\,\widetilde{M}_T) \le 2c_p^2\Delta/n + \Delta^2 D^p/n^2 \le \mu_T^2 h_T.$$

The last inequality follows by substituting $\gamma_T = 2\Delta/(nc_p^2) + \Delta^2 D^{p-1}/(n^2 c_p^4)$. Both raw types inherit symmetry, finite second moments and stabilizer invariance from their normalized versions.

We now check the bounds uniformly over the required sample counts. Write $H = D^{p-1/2}$, $d_* = (2p-3)/(p-1)$, $S_H = Ds^{4-2/(p-1)}$, $S_T = s^2 D^{p/2}$ and $S = \min\{S_H, S_T\} = D^b$. Thus $1 < d_* < 2$, $S_H/H = (S_T/H)^{d_*}$, and $S \le ND^{-\eta}$. If $N \le H$, having $S_T > H$ would force $S_H > H$ and $S > N$. Hence $S_T \le H$, so $S_H \le S_T$ and $S_H \le ND^{-\eta}$; this also implies $a \le (p-1)/4$. If $N > H$ and $S_T \le S_H$, then $S_T \le ND^{-\eta}$. Otherwise $S_T = H^{1-1/d_*} S_H^{1/d_*} < ND^{-\eta/d_*} \le ND^{-\eta/2}$. Thus the selected scale has the same bounds for every admissible N.

Here $t = O_{P,\eta}(\ell)$, $m = O_{p,P,\eta}(\ell^2)$ and $L_0 = O_{p,P,\eta}(\ell^3)$. Eventually $N \ge 2L_0$, and the floor gives $n \ge N/(2L_0)$. Constants C in the remaining estimates may depend on f, p, P, a, η, M, but not on N, u or the noise law in $\mathcal{N}_M$. In the hybrid branch, $\rho = s^4/n \le CD^{-1/2-\eta}\ell^3 \le 1$ eventually. Expanding $(1+D\rho)^{p-2} \le C_p[1 + (D\rho)^{p-2}]$ in (182) yields

$$h_H \le C\big[D^{-\eta}\ell^3 + D^{-(p-1)\eta}\ell^{3(p-1)}\big].$$

Indeed the linear term Ds^2/n is at most $Cs^{-2(p-2)/(p-1)}D^{-\eta}\ell^3$, and the highest term $D^{p-1}s^{4p-6}/n^{p-1}$ is at most $CD^{-(p-1)\eta}\ell^{3(p-1)}$. In the tensor branch the displayed formula for γ_T gives

$$h_T \le C\big[D^{1-p/2-\eta/2}\ell^3 + D^{-\eta}\ell^6\big].$$

Both are at most $CD^{-\eta/2}\ell^{3(p-1)}$. Since $\nu = \epsilon/(1000Pt)$, eventually the selected h satisfies $h \le \nu$ and $B_*^2\ell h \le \nu$. Its raw trace variance is at most $B_*^2 h \le \mu_0^2\nu$. Lemma O.3 therefore applies, with $v = \mu_0^2\nu$; this bound is used only in the analysis. Also $S \ge Ds^2$ and $\Delta \le (C_f^{\mathrm{cov}} + 1)s^2$ imply $\Delta/n \le CD^{-1-\eta}\ell^3$. With $K = P - p + 1$, this is eventually at most $\epsilon/(2^{P+1}K)$.

For each product, Lemma K.2 now gives

$$\mathbb{E}\mathcal{B} = \theta^t U, \qquad K_t = \mathbb{E}\mathcal{B}^\top\mathcal{B} = a_t U + b_t Q, \qquad a_t \le e^{r/100}, \quad \mathrm{tr}\,K_t < 2, \quad 1-\theta \le \nu.$$

Independence and (180) give $\mathbb{E}T_q = c_q\theta^{qt}u^{\otimes q}$ and $\mathbb{E}\|T_q\|_F^2 \le c_q^2 a_t^q + 2^q\Delta/n$. The latter bound follows from $\mathrm{tr}(K_t^{\otimes q}\Pi_q) \le (\mathrm{tr}\,K_t)^q$. Symmetric projection contracts norm and fixes the target tensor. The Gaussian query chaoses are orthogonal for each realized training sample, including across the shared products and final block. Using $\sum_q c_q^2 = 1$ and $1 - \theta^{qt} \le qt(1-\theta)$, the full expected risk is therefore at most

$$e^{\epsilon/100} - 1 + 2Pt\nu + K2^P\Delta/n \le (1/99 + 1/500 + 1/2)\epsilon < \epsilon.$$

A common dimension threshold for the finitely many eventual bounds proves the claimed uniformity over N, teachers and noise laws.

Finally put $d_p = \max\{4p-5, 2(p+1)\}$. A raw matrix or trace has joint training degree at most d_p and label degree at most $2p-3$, by the primitive definitions. The scalar formula has degree at most $2m-1$ in its traces. Thus the final joint degree is at most $2mPtd_p + 2P + 1$, and its label degree is at most $2mPt(2p-3)+1$, for every D, N. The zero output meets these bounds. Every denominator is deterministic. Each pilot needs only Y^4 for its second moment, and a tensor block needs only Y^2. Each monomial of the scalar formula uses each independent raw trace at most once; squaring uses it at most twice. Independent raw objects, successive matrix factors and the final block consequently have all the finite second moments used above, even before the dimension threshold. No higher noise moment is required. The scalar sum uses $O(m)$ successive products, and all remaining operations are finite sums, tensor contractions and matrix products with tensors of order at most P. Since $mPt = O_{p,P,\eta}(\ell^3)$, their work and storage are polynomial in N, D for fixed p, P, η. □

P A CUBIC-LABEL PILOT AT HERMITE RANK THREE

A cubic-label pilot can replace the quadratic signal at Hermite rank three. Warm-start tensor contraction in Damian et al. (2024, arXiv version 2, Lemma F.4) supplies the recovery ancestry; the following construction is polynomial and controls full expected prediction risk.

Proposition P.1. *Fix an odd integer $P \geq 3$ and a known link $f = \sum_{q=3,5,\dots,P} c_q h_q$ with $\sum_q c_q^2 = 1$, $c_3 \neq 0$, and $\gamma_f := \mathbb{E}[Zf(Z)^3] \neq 0$ for $Z \sim N(0,1)$. Observe iid pairs $X_i \sim N(0, I_D)$, $Y_i = f(u^\top X_i) + s\xi_i$, and an independent Gaussian query x. For every $\eta > 0$ there is a predictor collection indexed by integers $D \geq 2, N \geq 0$, depending only on f, P, D, N, η, of joint degree $O_{P,\eta}(\log D)$ in the training inputs, labels and query. For every fixed $a > 0, M_6 \geq 1$, eventually in D, uniformly over unit teachers, iid noise laws independent of the iid Gaussian inputs with mean zero, variance one and sixth moment at most M_6, and all integer sample counts satisfying*

$$s = D^a, \qquad B_0 = \min\{1 + 4a, 3/2 + 2a\}, \qquad N \geq \lceil D^{B_0+\eta} \rceil,$$

its expected squared error for $f(u^\top x)$ is at most $D^{-\eta/4}$. The dimension threshold may depend on f, P, a, η, M_6; the learner receives no noise parameters. The degree-17 link constructed in Proposition J.5 has $\gamma_f \neq 0$.

Proof. All blocks below are disjoint, of size n. Put $U = uu^\top$, $Q = I - U$, $d = D - 1$. Constants denoted by C may depend on f, P, M_6 and increase between occurrences. The pilot $L = (n\gamma_f)^{-1} \sum_i Y_i^3 X_i$ has mean u: indeed $\mathbb{E}[Y^3 \mid Z] = f(Z)^3 + 3s^2 f(Z) + s^3 \mathbb{E}\xi^3$ and $\mathbb{E}[Zf(Z)] = \mathbb{E}Z = 0$. Since $\mathbb{E}[Y^6 \mid Z] \leq 32 s^6 (f(Z)^6 + M_6)$ for $s \geq 1$, parallel and perpendicular coordinates give

$$v := \mathrm{Var}(u^\top L) \leq Cs^6/n, \qquad \mathbb{E}\|QL\|^2 \leq dCs^6/n, \qquad \mathbb{E}\|L\|^2 \leq 1 + DCs^6/n.$$

For example, $C = 32\gamma_f^{-2} \max\{\mathbb{E}[(f(Z)^6 + M_6)Z^2], \mathbb{E}[f(Z)^6 + M_6]\}$ suffices. The raw tensors $C_{q,n}$ in (180) have mean $c_q u^{\otimes q}$ and covariance at most $(\Delta/n)\Pi_q$, where $\Delta = C_f^{\mathrm{cov}} + s^2 \leq Cs^2$. That identity uses only the first two noise moments and applies at every target order here.

For a fresh order-three tensor define $M_H = c_3^{-1} C_{3,n}[\cdot, \cdot, L]$ and put $g_H = \Delta \mathbb{E}\|L\|^2/(nc_3^2)$, $z_H = v + g_H$. Conditionally on L, its mean is $(u^\top L)U$, and the variance tested against a deterministic matrix F is at most $\Delta \|F\|_F^2 \|L\|^2/(nc_3^2)$, by symmetric projection of $F \otimes L$ in (180). The parallel entry, d perpendicular-column entries and d^2 perpendicular-block entries therefore give the three bounds $1 + z_H, dg_H, Dg_H$ in (130). The matrix is symmetric with mean U, and

$$h_H := z_H + Dg_H \leq C\left(\frac{s^6}{n} + \frac{Ds^2}{n} + \frac{D^2 s^8}{n^2}\right). \tag{186}$$

Alternatively, use the order-three Gram matrix in (183). With $\omega = \Delta/(nc_3^2)$ and $g_T = 2\omega + D^2\omega^2$, its corresponding bounds are $1 + g_T, dg_T, Dg_T$ and $h_T = (D+1)g_T \leq C(Ds^2/n + D^3 s^4/n^2)$. Each matrix uses two blocks. Rotating every Gaussian design row by an orthogonal map fixing u rotates the pilot and tensors and preserves their joint law. Thus both matrices have the invariance required by Lemma K.2.

Set $e = D^{-\eta/4}$, $t = \lceil \log_2(100DP/e) \rceil$ and $n = \lfloor N/(2Pt+1) \rfloor$. Output zero if $n = 0$; otherwise select M_H if $N \leq D^2$ and M_T if $N > D^2$. For the asserted sample counts, eventually $n \geq N/(C_{P,\eta} \log D)$. In the first branch, necessarily $a < 1/4$ and $B_0 = 1 + 4a$. The three terms of (186) are then at most

$$C_{f,P,\eta,M_6}\left(D^{2a-1-\eta} \log D, D^{-2a-\eta} \log D, D^{-2\eta} (\log D)^2\right),$$

respectively. For the Gram branch with $a \geq 1/4$, its two costs are at most $C_{f,P,\eta}[D^{-1/2-\eta} \log D + D^{-2\eta}(\log D)^2]$. When $a < 1/4$, the two inequalities $N > D^2$ and $N \geq D^{1+4a+\eta}$ imply $N^2 > D^{3+4a+\eta}$, so those costs are at most $C_{f,P,\eta}[D^{-2a-\eta} \log D + D^{-\eta}(\log D)^2]$. Every displayed bound gives $h \leq e/(100Pt)$ eventually, uniformly over all required N and noise laws.

Apply Lemma K.2 with mean parameter one and $r = e/P$. Take P independent length-t products $\mathcal{B}^{(j)}$ and one fresh final tensor block. Set $T_q = (\bigotimes_{j=1}^q \mathcal{B}^{(j)}) C_{q,n}$ and output $\sum_{q: c_q \neq 0} J_q(\Pi_q T_q)(x)$, as in (184). The products have mean U and second-moment matrix K with $u^\top K u \leq e^{e/(100P)}$ and

tr $K < 2$. The final covariance bound and orthogonality of the target chaoses therefore give full risk at most

$$e^{e/100} - 1 + \sum_{q:c_q \neq 0} 2^q \Delta/n \leq e/99 + P2^P \Delta/n \leq e$$

eventually, since $B_0 \geq 1+2a$ implies $\Delta/n \leq C_{f,P,\eta} D^{-1-\eta} \log D$. This uses at most $(2Pt+1)n \leq N$ observations. The pilot has label and joint training degrees three and four; M_H has degrees four and eight, and M_T has degrees two and eight. Including the final tensor and query, the degrees are at most $4Pt+1$ and $8Pt+2P+1$. Each block enters independently, so the second-moment product bound needs no noise moment beyond the sixth, regardless of t.

For the last assertion write $H_j = \mathrm{He}_j$ and $R = \sum_{q=3,5,\ldots,17} a_q H_q$, where $a_1 = 0$, $a_{17} = -1$, and

$$2qa_q + a_{q-2} + (q+1)(q+2)a_{q+2} = 0 \quad (q = 3, 5, \ldots, 15).$$

The invertible matrix A_7 of Proposition J.5 makes these coefficients unique and rational. Its witness before positive normalization is $R + \theta H_3$, where $\theta > 0$ and $\theta^2 = 267778621561651200/6781$: this is the rescaling $\theta = z\sqrt{17!/6}$ of that proof. For $j = 2, 4$, the Hermite product formula gives the finite sum

$$\mathbb{E}[H_j R^2] = j! \sum_{q,r \in \{3,5,\ldots,17\}} a_q a_r \binom{q}{\ell} \binom{r}{\ell} \ell!, \qquad \ell = (q + r - j)/2,$$

with a summand zero unless ℓ is an integer in $[0, \min(q, r)]$. Substitution of the recurrence gives

$$\mathbb{E}[H_2 R^2] = -36\theta^2, \qquad \mathbb{E}[H_4 R^2] = \frac{2301022004490445101465600}{45981961} > 0.$$

The $q = 3$ recurrence also gives $\mathbb{E}[H_2 R H_3] = 0$. Thus $\mathbb{E}[H_2 (R + \theta H_3)^2] = 0$, using $\mathbb{E}[H_2 H_3^2] = 36$. Since $H_1 H_3 = H_4 + 3H_2$ and $\mathbb{E}[Z H_3^3] = 324$,

$$\mathbb{E}[Z(R + \theta H_3)^3] = A + 3\mathbb{E}[H_4 R^2]\theta \quad \text{for some } A \in \mathbb{Q}.$$

The fraction defining θ^2 is in lowest terms and $82^2 < 6781 < 83^2$, so θ is irrational. The last display is nonzero. Moreover, $a_3 = 247005158400/6781 > 0$ and $a_{17} = -1$, so normalization preserves nonzero coefficients at both orders and preserves the nonzero cubic signal. □

For Gaussian noise, Corollary I.3 with $p = 3$ gives the matching lower exponent B_0 for the arbitrary-design label-polynomial class; it requires no assumption on β_f. This lower statement does not assert hardness for each of the other noise laws in Proposition P.1.

Corollary P.2. *Fix an odd integer $P \geq 3$ and $\eta > 0$. There are predictors $\widehat{f}_{P,D,N,\eta}$ for integers $D \geq 2, N \geq 0$, with coefficients depending only on these indices and joint degree $O_{P,\eta}(\log^3(eD))$. For every fixed normalized link $f = \sum_{q=3,5,\ldots,P} c_q h_q$ with $c_3 \gamma_f \neq 0$, and every fixed $a > 0$ and finite $M_6 \geq 1$, under the observation and noise hypotheses of Proposition P.1, put $s = D^a$ and $B_0 = \min\{1 + 4a, 3/2 + 2a\}$. Eventually in D, for every integer $N \geq \lceil D^{B_0+\eta} \rceil$,*

$$\sup_{\|u\|=1} \mathbb{E}\big(\widehat{f}_{P,D,N,\eta}(x) - f(u^\top x)\big)^2 \leq D^{-\eta/8}.$$

The bound is uniform over those noise laws. The learner receives no coefficients or noise parameters; its dimension threshold may depend on f, P, a, η, M_6.

Proof. Put $H_P = \sum_{q=3,5,\ldots,P} h_q^2$, $M_P = \mathbb{E} H_P(Z)^3$ and $B_* = 1 + M_P$, explicit constants from Gaussian polynomial moments. Cauchy–Schwarz gives $f^2 \leq H_P$ and $|\gamma_f|^2 \leq \mathbb{E} f(Z)^6 \leq M_P$; Jensen gives $M_P \geq 1$. For each D, N choose

$$\ell = \log(eD), \quad \epsilon = D^{-\eta/8}, \quad r = \epsilon/P, \quad t = \lceil \log_2(100D/r) \rceil, \quad \mu_0 = \ell^{-1/2},$$
$$\nu = r/(1000t), \quad m = \lceil 4B_*^2 \ell \log(1/\nu) \rceil, \quad L_0 = 4mPt + 1, \quad n = \lfloor N/L_0 \rfloor.$$

Output zero when $n = 0$. Otherwise use the raw hybrid $\widetilde{M}_H = C_{3,n}[\cdot, \cdot, n^{-1} \sum_i Y_i^3 X_i]$ if $N \leq D^2$, and the raw Gram $\widetilde{M}_T$ from the proof of Theorem O.2 at order three otherwise. Each uses

two independent size-n blocks. Calibrate each raw matrix by (124), with $H = 2B_*$ and $2m-1$ independent raw traces of the selected type. Form P independent length-t products and use (185), restricted to all odd orders $3, 5, \dots, P$, with one fresh final block. This uses $L_0 n \le N$ observations and no unknown parameters.

The raw means are $\mu_H U = c_3 \gamma_f U$ and $\mu_T U = c_3^2 U$. They have absolute value at most B_*, and eventually both exceed μ_0 in absolute value. Their conceptual normalizations are exactly the matrices of the preceding proof, with its invariance, finite second moments and parameters (z_H, g_H) or (g_T, g_T). Using that proof's v, g_H, h_H, h_T, its conditional covariance with $F = I$ and total variance give the first bound below; the Gram trace expansion in the proof of Theorem O.2 gives the second:

$$\mathrm{Var}(\mathrm{tr}\,\widetilde{M}_H) \le \mu_H^2 (v + D g_H) \le \mu_H^2 h_H,$$
$$\mathrm{Var}(\mathrm{tr}\,\widetilde{M}_T) \le 2c_3^2 \Delta/n + \Delta^2 D^3/n^2 \le \mu_T^2 h_T.$$

Here and below constants may depend on f, P, a, η, M_6, but not on N, u or the noise law. Since $L_0 = O_{P,\eta}(\ell^3)$, eventually $N \ge 2L_0$ and $n \ge N/(2L_0)$ at the required counts. For $N \le D^2$, necessarily $a < 1/4$ and $B_0 = 1 + 4a$; (186) gives $h_H \le C[(D^{2a-1-\eta} + D^{-2a-\eta})\ell^3 + D^{-2\eta}\ell^6]$. For $N > D^2$ and $a \ge 1/4$, $h_T \le C[D^{-1/2-\eta}\ell^3 + D^{-2\eta}\ell^6]$. For $N > D^2$ and $a < 1/4$, use $N^2 > D^{3+4a+\eta}$ to obtain $h_T \le C[D^{-2a-\eta}\ell^3 + D^{-\eta}\ell^6]$. Thus always $h \le C D^{-\eta}\ell^6$, so eventually $h \le \nu$ and $B_*^2 \ell h \le \nu$. The raw trace variance is at most $B_*^2 h \le \mu_0^2 \nu$. Lemma O.3 applies with analysis-only variance bound $\mu_0^2 \nu$, giving mean θU, $1 - \theta \le \nu$, and all hypotheses of Lemma K.2.

Since $B_0 \ge 1 + 2a$, also $\Delta/n \le C D^{-1-\eta}\ell^3$. The product and full-risk calculation in the proof of Theorem O.2 applies: Gaussian query chaoses are orthogonal even for shared training data, and $\sum_q c_q^2 = 1$. It bounds the risk by

$$e^{\epsilon/100} - 1 + 2Pt\nu + P2^P \Delta/n \le (1/99 + 1/500 + 1/2)\epsilon < \epsilon$$

eventually. Taking the largest of the finitely many thresholds makes all bounds uniform over the stated sample counts and noise laws. Each raw matrix or trace has label degree at most four and joint degree at most eight. The final degrees are therefore at most $8mPt + 1$ and $16mPt + 2P + 1$, respectively, for every D, N. All denominators are deterministic. A cubic pilot needs only the sixth noise moment; tensor blocks need only the second. Every scalar monomial uses each independent trace at most once. Independence of raw objects, matrix factors and the final block therefore gives all required second moments without higher noise moments as m, t grow. □

Q A LOWER BOUND FROM THE POLYNOMIAL MOMENT PROFILE

This appendix gives a direct proof of the general profile rule. The lower uses the Hermite shift and cumulant identities (19) and (20). The upper uses the Gaussian tensor identities in Lemma B.2 and the matrix-product bound in Lemma K.2. The intervening special-case results are not prerequisites.

Fix integers $3 \le p \le P$ and a centered polynomial

$$f(z) = \sum_{q \in I} c_q h_q(z), \qquad I \subseteq \{p, p+1, \dots, P\}, \qquad c_p \ne 0, \qquad \sum_{q \in I} c_q^2 = 1. \tag{187}$$

Here $h_q = \mathrm{He}_q / \sqrt{q!}$, and I indexes the nonzero coefficients. For $Z \sim N(0,1)$, define

$$m_j = \min\{b \in \mathbb{N}_{\ge 1} : \mathbb{E}[\mathrm{He}_j(Z) f(Z)^b] \ne 0\}, \qquad 1 \le j \le p, \tag{188}$$

where an empty minimum is $+\infty$, and set $\mathcal{J} = \{1 \le j \le p : m_j < \infty\}$. The quantity m_2 is the noise-sensitivity exponent of Defilippis et al. (2026, arXiv version 1, Definition 2.1). For fixed f and $j \in \mathcal{J}$, put $Y_s = f(Z) + sG$, where $s > 0$ and G is standard Gaussian independent of Z. Then

$$\lambda_j(s)^2 := \mathbb{E}[\mathbb{E}[h_j(Z) \mid Y_s]^2] \sim \frac{\mathbb{E}[h_j(Z) f(Z)^{m_j}]^2}{m_j!\, s^{2m_j}} \qquad (s \to \infty).$$

Apply Lemma F.2 with $V = h_j(Z)$, $r = m_j$, and $t = 1/s$: orthogonality and (188) give the vanishing moments, and $Y_s/s = G + f(Z)/s$. For square-integrable $T(Y_s)$, conditional expectation and Cauchy–Schwarz give $\mathbb{E}[h_j(Z) T(Y_s)]^2 \le \lambda_j(s)^2 \mathbb{E}[T(Y_s)^2]$, with equality at $T(Y_s) = \mathbb{E}[h_j(Z) \mid Y_s]$, which is square-integrable by conditional Jensen. Thus m_j measures the high-noise decay of the optimal scalar signal at order j.

Lemma Q.1. *If f is even, $\mathcal{J}$ consists exactly of the even orders in $\{1,\dots,p\}$. Otherwise $\mathcal{J} = \{1,\dots,p\}$. In particular $2, p \in \mathcal{J}$, $m_p = 1$, and $m_j \geq 2$ for $j < p$. If f is odd, m_j has the parity of j and $m_1 \geq 3$.*

Proof. All moments are finite. Orthogonality gives $m_p = 1$ and excludes power one for $j < p$. If f is even, every odd-order moment vanishes by symmetry. If f is odd, the moment vanishes unless the power has the parity of j.

For even j, use positive even powers b. Outside a fixed interval $[-R, R]$, the product $\mathrm{He}_j(z)f(z)^b$ is nonnegative. Put $M = \max\{1, \max_{|z|\leq R} |f(z)|\}$ and $C = \max_{|z|\leq R} |\mathrm{He}_j(z)|$. On a fixed positive-tail interval of Gaussian probability $v > 0$, arrange $|f| \geq 2M$ and $\mathrm{He}_j \geq 1$. The moment is at least $v(2M)^b - CM^b > 0$ for all sufficiently large even b. If f has odd leading degree and j is odd, change its global sign to make its leading coefficient positive and use odd powers. Then $\mathrm{He}_j(z)f(z)^b$ is again nonnegative on both tails after enlarging R. Define M, C for this enlarged interval and choose the positive-tail interval as above; the same bound proves finiteness.

It remains to treat odd j when f has even leading degree but is not even. A global sign change preserves every moment's nonvanishing, so suppose the leading coefficient is positive. Write $f = g + h$ with g even and h odd and nonzero. Choose R so that for $z > R$ we have $g(z) > |h(z)| > 0$, $\mathrm{He}_j(z) > 0$, and $h(z)$ has a fixed sign $\epsilon \in \{-1, 1\}$. Pairing the two tails, their integrand is

$$\mathrm{He}_j(z)\big((g(z) + h(z))^b - (g(z) - h(z))^b\big), \qquad z > R.$$

For every positive integer b it has sign ϵ. With M, C defined on $[-R, R]$ as above, choose a fixed compact interval $J \subset (R, \infty)$ on which $g + |h| \geq 2M$ and $\mathrm{He}_j \geq 1$. Continuity and $g > |h| > 0$ give $\sup_J (g - |h|)/(g + |h|) = r_0 < 1$. If v is the Gaussian probability of J, then

$$\epsilon\, \mathbb{E}[\mathrm{He}_j(Z)f(Z)^b] \geq v(2M)^b(1 - r_0^b) - CM^b > 0$$

for all sufficiently large integers b. This proves all remaining finiteness claims and, together with symmetry, the stated description of $\mathcal{J}$. □

Set

$$C_* = \begin{cases} 2m_2, & 1 \notin \mathcal{J}, \\ 2\min\left\{ m_2, \min\limits_{\substack{j\in\mathcal{J}\\ j\geq 3}} \dfrac{(j-2)m_1 + m_j}{j-1} \right\}, & 1 \in \mathcal{J}. \end{cases} \tag{189}$$

For $x \geq 1$, define the finite-dimensional minimum

$$\phi(x) = \min\left\{ \sum_{j\in\mathcal{J}} \frac{j}{2} w_j : w_j \geq 0,\ \sum_{j\in\mathcal{J}} w_j = 1, \sum_{j\in\mathcal{J}} m_j w_j \leq x \right\}. \tag{190}$$

Lemma Q.2. *The minimum (190) exists for every $x \geq 1$. We have $C_* \geq 3$, and $C_* \geq 4$ if f is odd or even. If $1 \in \mathcal{J}$, then $C_* < 2m_1$. For every $a > 0$,*

$$\inf_{x\geq 1} [2ax + \max\{1, \phi(x)\}] = B(a) := \min\left\{ 1 + C_* a, \min_{\substack{j\in\mathcal{J}\\ j\geq 3}} \left(\frac{j}{2} + 2m_j a \right) \right\}. \tag{191}$$

If $m_2 = 2$ and $m_1 \geq 3$, allowing $m_1 = \infty$, then $B(a) = \min\{1 + 4a, p/2 + 2a\}$. If $m_1 = 2$, then $C_* = 4 - 2/(p-1)$ and $B(a) = \min\{1 + C_* a, p/2 + 2a\}$. For $p = 3$, put $\kappa = C_*$. Then

$$\kappa = \min\{2m_2, m_1 + 1\}, \qquad \phi(x) \geq \frac{3}{2} + \frac{1 - x}{\kappa - 2}. \tag{192}$$

Proof. The feasible set is compact and nonempty, since $m_p = 1$. If $1 \notin \mathcal{J}$, then $C_* = 2m_2 \geq 4$. Otherwise $m_1 \geq 2$ and $m_j \geq 1$, so every fraction in (189) is at least $2 - 1/(j-1) \geq 3/2$. For odd f, $m_1 \geq 3$ raises this bound to $3 - 2/(j-1) \geq 2$. Also $m_2 \geq 2$. The choice $j = p$ with $m_p = 1 < m_1$ proves $C_* < 2m_1$.

For weights on the simplex indexed by $\mathcal{J}$, write $M(w) = \sum_j m_j w_j$ and $Y(w) = \sum_j (j/2) w_j$. A minimizing vector for $\phi(x)$ has $M(w) \le x$; every vector is feasible at $x = M(w) \ge 1$. Since $a > 0$, these facts identify the infimum in (191) with

$$\min_w \big[2aM(w) + \max\{1, Y(w)\}\big].$$

If $1 \notin \mathcal{J}$, then $Y \ge 1$, and this linear objective is minimized at a vertex. Its values are $1 + 2m_2 a$ and $j/2 + 2m_j a$ for $j \in \mathcal{J}, j \ge 3$, which proves the formula in this case.

If $1 \in \mathcal{J}$, on each compact simplex region $Y \ge 1$ and $Y \le 1$ the objective is linear. A minimizer with at least three positive coordinates has a nonzero perturbation on its support preserving both the sum and Y. Small multiples with either sign are feasible, so optimality forces zero objective change. Extend one sign until a coordinate vanishes and repeat. With two positive coordinates and $Y \ne 1$, varying them oppositely is feasible in both directions locally and again preserves the objective; continue until a coordinate vanishes or $Y = 1$. Thus a minimum occurs at a vertex or at $Y = 1$ on an edge from index 1 to some $j \ge 3$. The edge weights are $(j-2)/(j-1)$ and $1/(j-1)$, giving value $1 + 2a((j-2)m_1 + m_j)/(j-1)$. The index-1 value $1 + 2m_1 a$ is larger than its edge value with $j = p$. The remaining vertex values and these edges give (191).

If $m_1 \ge 3$ and $m_2 = 2$, the same bounds give $C_* = 4$, including the case $1 \notin \mathcal{J}$. For $3 \le j < p$, $m_j \ge 2$ makes every retained Gram line larger than $1+4a$. If $m_1 = 2$, the fraction at p is $2 - 1/(p-1)$; every fraction with $j < p$ is at least two. Thus $C_* = 4 - 2/(p-1) < 4$, and again all $j < p$ Gram lines are dominated by $1 + C_* a$. This proves both specializations. For $p = 3$, f is not even, so all three indices are finite and $m_3 = 1$. The formula for κ follows. Each point $(m_j, j/2)$ lies on or above $y = 3/2 + (1-x)/(\kappa - 2)$: at $j = 1, 2$ this uses $m_1 \ge \kappa - 1$ and $m_2 \ge \kappa/2$, and at $j = 3$ there is equality. Feasible weights have $M(w) \le x$. Averaging the pointwise inequality and using its negative slope proves (192). □

Let $D \ge 1$ and $N, L \ge 0$ be integers and let $s > 0$. Take $U = \sigma/\sqrt{D}$ with uniform independent signs, independent Gaussian design rows $A_i \sim N(0, I_D)$, and independent $G_i \sim N(0, 1)$. Observe

$$V_i = T_i + G_i, \qquad T_i = f(A_i^{\mathsf{T}} U)/s, \qquad F_U(x) = f(U^{\mathsf{T}} x), \quad x \sim N(0, I_D).$$

The query and learner seed ξ are independent of the other random variables. The predictor class is

$$\widehat{F}(A, V, \xi; x) = \sum_{|\alpha| \le L} a_\alpha(A, \xi, x) V^\alpha, \tag{193}$$

with arbitrary jointly measurable real coefficients and no teacher-dependent side information. The coefficients may depend on f, s, the entire design, and the query. Rescaling to the raw labels $Y = sV$ preserves this class.

Use the finite-dimensional target space $\mathcal{H} = \bigoplus_{q \in I} \mathcal{H}_q$ and define

$$\begin{aligned} m_\alpha &= \mathbb{E}_U T^\alpha, & v_\alpha &= \mathbb{E}_U[F_U T^\alpha], \\ \kappa_\alpha &= v_\alpha - \sum_{\substack{\zeta \le \alpha \\ \zeta \ne \alpha}} \binom{\alpha}{\zeta} \kappa_\zeta m_{\alpha - \zeta}, & E_k &= \mathbb{E}_A \sum_{|\alpha| = k} \frac{\|\kappa_\alpha(A)\|_{\mathcal{H}}^2}{\alpha!}. \end{aligned} \tag{194}$$

All target-space expectations are coordinatewise. Set

$$\begin{aligned} A_P &= 3\sqrt{P\,P!}, \\ C_P &= \left\lceil \max\left\{4P, 2, 4A_P\sqrt{P+1}(2P)^{P/2}\right\}\right\rceil, & \Gamma_P &= 10P + 20, \\ \delta_f(k, r) &= \max\left\{r, \frac{p}{2} + r\phi(k/r)\right\}. \end{aligned} \tag{195}$$

Theorem Q.3. *Every predictor* (193) *satisfies*

$$\mathbb{E}[(\widehat{F} - F_U(x))^2] \ge 1 - \sum_{k=0}^{L} E_k.$$

All energies are finite. With $C_P^0 = \max\{1, (2\lfloor P/2\rfloor - 1)!!\}$ *and* $\rho_D = D^{-1}\sum_{j=1}^{D}\epsilon_j$ *for independent fair signs,*

$$E_0 = \sum_{\substack{q\in I\\ q \text{ even}}} c_q^2 \mathbb{E}\rho_D^q \le C_P^0 D^{-\lceil p/2\rceil}. \tag{196}$$

If f *is odd,* $E_k = 0$ *for every even* k*, including zero. For every* $k \ge 1$*,*

$$E_k \le (C_P k)^{\Gamma_P k} s^{-2k} \sum_{r=1}^{\min\{N,k\}} N^r D^{-\delta_f(k,r)}. \tag{197}$$

An empty sum is zero. If $N = 0$ *or* $L = 0$*, sign-prior risk is at least* $1 - E_0$*, attained by the prior-mean predictor* $\mathbb{E}_U F_U$*. The bounds hold for every link* (187)*, with its own profile* (188)*, and all the displayed finite parameters. The constants* C_P *and* Γ_P *do not depend on the coefficients or the profile indices.*

Proof. **The conditional certificate and its constant term.** We justify the certificate for all orders rather than applying the same-parity statement of Theorem I.1. Gaussian chaos orthogonality gives

$$\|F_u\|_2^2 = 1, \qquad \langle F_u, F_v\rangle = \sum_{q\in I} c_q^2 (u^\top v)^q.$$

A finite-risk predictor has finite expected squared query norm. At each fixed design its label law is a finite mixture of identity-covariance Gaussian laws. The monomial Gram matrix is positive definite: a nonzero polynomial cannot vanish Lebesgue almost everywhere. Its positive least eigenvalue and Tonelli's theorem imply that every coefficient is in $L^2(\gamma_D)$ for almost every design and seed. Project each onto $\mathcal{H}$; this preserves label degree and cannot increase risk. Integration against the finite Hermite basis makes the projection measurable; assign zero on the exceptional null set.

At a fixed design and seed, write the projected polynomial as $\sum_{|\alpha|\le L} b_\alpha q_\alpha(V)$ with $q_\alpha(y) = \prod_i \mathrm{He}_{\alpha_i}(y_i)/\sqrt{\alpha!}$. The finite shift identity (19), followed by Jensen's inequality over U and Gaussian orthogonality, gives

$$\mathbb{E}\langle F_U, \widehat{F}(T+G)\rangle = \sum_\alpha \langle z_\alpha, b_\alpha\rangle,$$
$$\mathbb{E}\|\widehat{F}(T+G)\|^2 \ge \sum_\beta \|d_\beta\|^2.$$

where $z_\alpha = v_\alpha/\sqrt{\alpha!}$, $d = Mb$, and

$$M_{\beta,\alpha} = \mathbf{1}_{\{\beta\le\alpha\}} m_{\alpha-\beta}\sqrt{\beta!/\alpha!}\binom{\alpha}{\beta}.$$

The recursion (194) gives $z = M^\top w$ for $w_\beta = \kappa_\beta/\sqrt{\beta!}$, including $\beta = 0$. Completing squares bounds conditional risk below by $1 + \sum_\beta \|d_\beta - w_\beta\|^2 - \sum_\beta \|w_\beta\|^2$. Averaging proves the certificate. Every coordinate of $\kappa_\alpha(A)$ is a polynomial in the design by the finite sign prior and recursion, so all energies are finite. The bound also holds for infinite-risk predictors.

The displayed target inner product gives the identity in (196). Odd Rademacher moments vanish. For even q, each nonzero coordinate tuple in the expanded moment admits a pairing of its q positions. There are $(q-1)!!$ pairings and at most $D^{q/2}$ assignments per pairing, so $\mathbb{E}\rho_D^q \le (q-1)!!D^{-q/2}$. The least possible even $q \ge p$ is $2\lceil p/2\rceil$; summing c_q^2 proves the bound. If $N = 0$ or $L = 0$, only E_0 remains; the prior-mean predictor attains risk $1 - E_0$. These arguments are the finite Hilbert calculation underlying (176) and (177); none uses the nonzero-signal hypothesis of the surrounding mixed-parity theorem. For odd f, replacing U by $-U$ negates the target and every signal. Multilinearity of the order-$(k+1)$ cumulant therefore makes $\kappa_\alpha = 0$ when k is even.

We prove the improved finite energy estimate. The deletion below applies the independence property of cumulants, as in Schramm and Wein (2022, arXiv version 2, Proposition 2.11); we include its finite-moment justification. No part of the argument assumes analyticity of the moment-generating function of $f(Z)$.

Centered powers and their coefficients. For an integer $a \geq 1$, put

$$\mu_a = \mathbb{E} f(Z)^a, \qquad g_a(z) = f(z)^a - \mu_a = \sum_{\ell=1}^{Pa} b_{a,\ell} h_\ell(z).$$

Centering removes the constant coefficient; no parity of g_a is assumed. For $1 \leq j \leq p$, a nonzero coefficient $b_{a,j}$ implies $j \in \mathcal{J}$ and $a \geq m_j$ by (188), since $b_{a,j} = \mathbb{E}[\mathrm{He}_j(Z) f(Z)^a]/\sqrt{j!}$.

We record a uniform coefficient bound. The explicit formula for He_q gives

$$|h_q(z)| \leq 3\sqrt{q!}(1 + |z|^P) \quad (q \leq P), \qquad |f(z)| \leq A_P(1 + |z|^P).$$

For the first inequality, take absolute values in $\mathrm{He}_q(z) = q! \sum_j (-1)^j z^{q-2j}/(2^j j!(q-2j)!)$ and use $\sum_{j \geq 0} 1/j! < 3$; the second uses $\sum_q |c_q| \leq \sqrt{P}$. Minkowski's inequality and $\mathbb{E} Z^{2Pa} = (2Pa - 1)!! \leq (2Pa)^{Pa}$ yield

$$\|f^a\|_{L^2(\gamma_1)} \leq \left[2A_P(2Pa)^{P/2}\right]^a.$$

Centering is an orthogonal projection, so $\|g_a\|_2 \leq \|f^a\|_2$. Cauchy–Schwarz therefore gives

$$\begin{aligned} |\mu_a| + \sum_{\ell \geq 1} |b_{a,\ell}| &\leq (1 + \sqrt{Pa})\|f^a\|_2 \\ &\leq \left[4A_P\sqrt{P+1}(2P)^{P/2}\right]^a a^{(P+1)a/2} \leq (C_P a)^{(P+1)a/2}. \end{aligned} \tag{198}$$

Here $1 + \sqrt{Pa} \leq [2\sqrt{(P+1)a}]^a$ for $a \geq 1$, and the last inequality follows from (195).

Deleting all-constant selections. Fix a multiindex α of order $k \geq 1$ and support size r. Treat its k label occurrences as distinct positions and add one target position. The joint-cumulant partition identity, applied coordinatewise to the target and label occurrences, represents $s^k \kappa_\alpha$ as

$$\sum_\pi (-1)^{|\pi|-1}(|\pi| - 1)! \prod_{B \in \pi} \mathbb{E}_U \left[F_U^{\mathbf{1}\{0 \in B\}} \prod_{i \in \mathrm{supp}\, \alpha} f(A_i^\top U)^{a_{iB}} \right].$$

Its target-weighted moment recursion is exactly (194), as in (20); this identity requires no parity restriction. The product is interpreted in $\mathcal{H}$ through its one vector-valued factor, and a_{iB} counts the row-i occurrences in block B. Each expectation has its own independent teacher variable. In every occupied pair (i, B) expand $f^{a_{iB}} = \mu_{a_{iB}} + g_{a_{iB}}$. Call a row all-constant in a selection if every one of its occupied pairs uses $\mu_{a_{iB}}$.

For any nonempty subset J of rows, the sum of all expanded terms making every row in J all-constant is zero, pointwise in the design. To see this, introduce independent variables $W_i = f(Z_i)$ for $i \in J$, independent of the teacher and all other variables, and replace every occurrence of row i by the same W_i. Each block expectation then has the factor $\mathbb{E} W_i^{a_{iB}} = \mu_{a_{iB}}$, so its cumulant partition formula is exactly the asserted sum after expanding the other rows back into their powers. This cumulant has an independent nonempty group of W_i arguments and a nonempty complement containing the target, and hence is zero.

This last independence property uses only finite moments. One can work in the commutative ring with one variable for each labeled argument and with its square set to zero. The finite moment series has constant term one, so its logarithm is the finite sum of powers of its nilpotent nonconstant part. The coefficient using all variables is the cumulant partition sum. Independence factors the moment series into the two groups; the formal logarithm is their sum and its mixed coefficient is zero. Thus no analytic generating function is required.

Inclusion–exclusion over all-constant row events now shows that the full cumulant equals the sum of selections having at least one g factor in every row. This is an identity among the entire partition sums. It does not say that the zero Gaussian-chaos projection of a row is zero; products belonging to distinct teacher blocks can still have a nonzero Gaussian mean.

A common order bound for the retained rows. Fix any nonzero retained term after the Hermite expansions. In each row and each replica, select one of its positive-order Hermite factors. Suppose its power is a and its order is j. If $j \le p$, then $j \in \mathcal{J}$; associate the profile point $(m_j, j/2)$ to it, with $m_j \le a$ by the preceding coefficient identity. If $j > p$, associate the point $(m_p, p/2) = (1, p/2)$ instead. In both cases the point's first coordinate is at most the row's multiplicity, and its second coordinate is at most half the total selected Hermite order in that row and replica. Let R be the sum of all selected Hermite orders over the two replicas. There are $2r$ chosen profile points. Their mean first coordinate is at most k/r, because the row multiplicities sum to k in each replica, and their mean second coordinate is at most $R/(4r)$. Their empirical weights are feasible in (190); hence

$$\frac{R}{4} \ge r\phi(k/r). \tag{199}$$

This bound holds for every retained term, including rows whose powers are split among several teacher blocks. Other positive-order factors can only increase the total order.

Loop-free overlap products. Project the target onto a fixed order $q \in I$ and square its norm, using independent teacher variables for the two partition replicas. Distinct target orders are orthogonal, so they will be summed with weights c_q^2. The inner product of the order-q target tensors is $(U^\mathsf{T}U')^q$, with the teachers belonging to the target blocks in the two replicas.

In any design row, there is at most one Hermite factor for a given teacher block: all occurrences of that row and block were grouped before centering. The Gaussian normal-ordering identity therefore expands the expectation of its factors into pairings with no pair inside a single factor. Every pair joins distinct teacher blocks. The two replicas use disjoint sets of teacher variables, so a pair crossing replicas also joins distinct blocks. All target bridges have distinct endpoints for the same reason. This proves that every resulting edge is loop-free, even when multiple edges have the same two endpoints.

A surviving pairing has even total selected order R and contributes a product of overlaps of degree $e = q + R/2$. Equation (199) gives

$$\frac{e}{2} \ge \frac{q}{2} + r\phi(k/r), \qquad e \le P(k+1) \le 2Pk.$$

There is a factor $1/\sqrt{\ell!}$ for a factor of order ℓ; we bound these denominators by one. Since the teachers at the vertices are independent signs, the overlap on every loop-free edge has the marginal law $\rho_D = D^{-1}\sum_{j=1}^{D} \epsilon_j$. Treat parallel edges as separate factors. Hölder's inequality on all e factors gives

$$\mathbb{E}\prod_{b=1}^{e} |U_{v_b}^\mathsf{T} U_{w_b}| \le \mathbb{E}|\rho_D|^e \le (2e)^{e/2} D^{-e/2}.$$

For even e, the last bound follows by assigning a perfect pairing to every tuple with even coordinate multiplicities in the expanded Rademacher moment, giving $(e-1)!!D^{-e/2}$. For odd e, monotonicity of L^p norms and the even moment of order $e+1$ give at most $((e+1)/D)^{e/2}$. Since $q \ge p$, each term is consequently bounded by

$$(4Pk)^{Pk} D^{-p/2 - r\phi(k/r)}. \tag{200}$$

No independence between edges is needed for this use of Hölder.

Finite counting and the other loss branch. For partitions of $n = k+1$ positions, the sum of absolute cumulant coefficients is at most n^n. Indeed, for b blocks, $b!$ times the partition count is at most b^n, and hence the weighted sum is at most $\sum_{b=1}^{n} b^{n-1} \le n^n$. The squared coefficient sum is therefore at most $(k+1)^{2(k+1)} \le (2k)^{4k}$. For each fixed pair of partitions, summing the absolute power coefficients over all selections and Hermite orders is, by (198), at most $(C_P k)^{(P+1)k}$, since the powers in each replica sum to k. This already counts both constant and centered choices; restricting to the retained choices only decreases that absolute sum. There are at most $(2Pk)^{Pk}$ pairings of their at most $2Pk$ Gaussian positions. Combining these counts with (200), using $C_P \ge 4P$, and summing $\sum_q c_q^2 = 1$ gives

$$\mathbb{E}_A \|\kappa_\alpha(A)\|^2 \le (C_P k)^{(3P+5)k} s^{-2k} D^{-p/2 - r\phi(k/r)}. \tag{201}$$

For the other branch, we give the parity-free fixed-multiindex coordinate count underlying (179). Expand the two tensor cumulants into label orders $j_1, \dots, j_{2k} \in I$ and a common target order $q \in I$. A row with an odd number of Gaussian positions contributes zero. Otherwise normal ordering supplies $\frac{1}{2}\sum_v j_v$ pairs inside rows and the target supplies q bridges. Write $e = \frac{1}{2}\sum_v j_v + q \le 2Pk$. The exact scalar factor is

$$s^{-2k} D^{-e} c_q^2 \prod_{v=1}^{2k} \frac{c_{j_v}}{\sqrt{j_v!}}.$$

Reduce teacher coordinate products to sign characters. At positive cumulant order a constant argument gives zero, and independent groups of characters give zero, by the finite cumulant partition identity. Thus nonzero terms have nonempty reduced supports and a connected character-intersection graph in each replica.

Give one initial coordinate variable to every pair and bridge, and let h be their number of equality classes. Form a bipartite multigraph with the r rows and target on one side, these h classes on the other, and one edge per initial variable. Shared coordinates in the character graph give paths in this graph. Identifying factors within a row preserves connections; coordinates absent after sign reduction still attach to a row or target. The graph is therefore connected, so $e-(r+1+h)+1 = e-h-r \ge 0$. There are at most $(2Pk)^{Pk}$ Gaussian pairings, $(2Pk)^{2Pk}$ equality partitions, and D^h coordinate assignments. The two character cumulants have product at most $(k+1)^{2(k+1)} \le (2Pk)^{4k}$ by the partition-coefficient bound above. Dropping factorial denominators, summing the label coefficients costs at most $(\sum_q |c_q|)^{2k} \le P^k$; target coefficients sum to one. Consequently

$$\mathbb{E}_A \|\kappa_\alpha(A)\|^2 \le (2Pk)^{(3P+5)k} s^{-2k} D^{-r} \le (C_P k)^{(3P+5)k} s^{-2k} D^{-r}.$$

No part of this count assumes a transformed signal is nonzero or that the link has one parity. Combining it with (201) at each fixed multiindex gives the maximum loss $\delta_f(k,r)$. There are at most $N^r 2^{k-1}$ multiindices of order k and support size r. Drop $1/\alpha! \le 1$, absorb $2^{k-1} \le (C_P k)^k$, and use $\Gamma_P \ge 3P+6$ to obtain (197). All sums are finite and include repeated rows, arbitrary coordinate coincidences, and $k > D$. For $N = 0$, the sum over r is empty, as required. □

Corollary Q.4. *Fix a link (187), its profile (188), and $a, \eta > 0$. Let $B = B(a)$ be as in (191), and put*

$$\tau = \min\{a, \eta/C_*\}, \qquad L_D = \left\lfloor D^{\tau/(4\Gamma_P)} \right\rfloor.$$

For integers

$$D \ge \left\lceil \max\left\{2, (2C_P)^{4\Gamma_P/\tau}, 2^{2/\tau}\right\} \right\rceil, \qquad 0 \le N \le D^{B-\eta}, \qquad s = D^a,$$

every predictor (193) of label degree at most L_D has sign-prior risk at least

$$1 - E_0 - 2D^{-\tau/2} \ge 1 - C_P^0 D^{-\lceil p/2 \rceil} - 2D^{-\tau/2}.$$

The same lower bound holds for its worst-teacher risk and for minimax risk in this class. It also applies to joint polynomials of total degree at most L_D in the training inputs, labels and query. These are Gaussian-noise lower bounds; no other noise law is asserted here.

Proof. If $N = 0$, Theorem Q.3 gives risk at least $1 - E_0$. This also covers the case $B - \eta < 0$, since $D \ge 2$ then forces $N = 0$. Suppose $N \ge 1$. For $1 \le r \le k$, first consider

$$e^0_{k,r} = Br - 2ak - \delta_f(k,r).$$

Since $\delta_f(k,r) \ge r \max\{1, \phi(k/r)\}$, Lemma Q.2, applied at $x = k/r \ge 1$, gives

$$e^0_{k,r} \le r\big[B - 2a(k/r) - \max\{1, \phi(k/r)\}\big] \le 0.$$

If $r \ge k/C_*$, the sample margin gives $e^0_{k,r} - \eta r \le -\eta k/C_*$. If $r < k/C_*$, use $B \le 1 + C_* a$ and the first loss branch:

$$(B-\eta)r - 2ak - \delta_f(k,r) \le C_* a r - 2ak - \eta r \le -ak.$$

Consequently every summand in (197) has its dimension factor at most $D^{-\tau k}$. There are at most k values of r, so

$$E_k \le k(C_P k)^{\Gamma_P k} D^{-\tau k} \le (2C_P k)^{\Gamma_P k} D^{-\tau k}.$$

For $k \le L_D$, the stated dimension threshold gives $(2C_P k)^{\Gamma_P} \le (2C_P)^{\Gamma_P} D^{\tau/4} \le D^{\tau/2}$. Thus $E_k \le D^{-\tau k/2}$, and $D^{-\tau/2} \le 1/2$ gives $\sum_{k=1}^{L_D} E_k \le 2D^{-\tau/2}$. The risk certificate, retaining E_0, proves the sign-prior claim. For each learner, worst-teacher risk is at least its sign-prior average, so taking the infimum over learners preserves the bound. A joint polynomial of total degree at most L_D has label degree at most L_D and is contained in (193), completing the proof. □

For $p = 3$, put $\kappa = C_*$. Equation (192) gives

$$\delta_f(k, r) \ge \max\left\{r, \left(\frac{3}{2} + \frac{1}{\kappa - 2}\right) r - \frac{k}{\kappa - 2} + \frac{3}{2}\right\}.$$

Thus (197) also holds with the displayed loss in place of $\delta_f(k, r)$. Corollary Q.4 gives $B = \min\{1 + \kappa a, 3/2 + 2a\}$ and $\tau = \min\{a, \eta/\kappa\}$, with constants C_P, Γ_P from (195), the stated dimension threshold and degree cutoff, and risk $1 - E_0 - 2D^{-\tau/2}$. For odd links $E_0 = 0$, giving the previous rank-three bound exactly. In particular, $\kappa = 6$ gives $B = \min\{1 + 6a, 3/2 + 2a\}$ and $\tau = \min\{a, \eta/6\}$. For an odd or even link with $m_2 = 2$, Lemma Q.2 gives $B = \min\{1 + 4a, p/2 + 2a\}$.

The recovery bounds in Damian et al. (2024, arXiv version 2, Theorem 4.1 and Lemma F.4) supply the ancestry for transformed tensors, quadratic matrices and contractions with linear pilots. The proof below controls full expected prediction risk with logarithmic joint degree.

Q.1 Prediction from the transformed-moment profile

Proposition Q.5. *Fix integers $3 \le p \le P$ and a supplied fixed centered link $f = \sum_{q=p}^{P} c_q h_q$, with $\sum_q c_q^2 = 1$ and $c_p \ne 0$. For standard Gaussian Z, put*

$$m_j = \min\{m \in \mathbb{N}_{\ge 1} : \mathbb{E}[h_j(Z) f(Z)^m] \ne 0\} \quad (1 \le j \le p),$$
$$J = \{1 \le j \le p : m_j < \infty\}.$$

where an empty minimum is $+\infty$. The set J consists of the positive even orders at most p if f is even, and all orders $1, \dots, p$ otherwise. In particular $2, p \in J$. Define

$$C_* = \begin{cases} 2m_2, & 1 \notin J, \\ 2\min\left\{m_2, \min\limits_{\substack{j \in J \\ j \ge 3}} \dfrac{(j-2)m_1 + m_j}{j-1}\right\}, & 1 \in J, \end{cases}$$
$$H_* = \max\{C_* - 2, C_*/2\}.$$

One has $C_* \ge 3$. Set $r_* = m_2$ if $C_* = 2m_2$, including ties, and $r_* = m_1$ otherwise. Write $\mathbf{m} = (m_j)_{j \in J}$. For every $\eta > 0$ there is a predictor collection indexed by integers $D \ge 2, N \ge 0$, whose coefficients depend only on f, p, P, D, N, η and whose joint degree is $O_{p,P,\mathbf{m},\eta}(\log(eD))$ in the training inputs, labels and query. Observe iid $X_i \sim N(0, I_D)$ and $Y_i = f(u^\top X_i) + s\xi_i$, and an independent query $x \sim N(0, I_D)$, where $\|u\| = 1$. For every fixed $a > 0$ and $M \ge 1$, put

$$s = D^a, \qquad B = \min\left\{1 + C_* a, \min_{\substack{j \in J \\ j \ge 3}} (j/2 + 2m_j a)\right\}.$$

Eventually in D, for every integer $N \ge \lceil D^{B+\eta} \rceil$,

$$\sup_{\|u\|=1} \mathbb{E}\big(\widehat{f}(x) - f(u^\top x)\big)^2 \le D^{-\eta/H_*},$$

uniformly over iid noise laws independent of the Gaussian inputs, with mean zero, variance one and $\mathbb{E}|\xi|^{2r_*} \le M$. The learner receives no a, s, M or noise law. Dimension thresholds and resource constants may depend on the fixed link and its moment indices; thresholds may also depend on a, η, M.

Proof. Lemma Q.1 identifies J and gives $m_p = 1$ and $m_j \ge 2$ for $j < p$. Thus $C_* \ge 3$: if $1 \notin J$, then $C_* = 2m_2 \ge 4$; otherwise $m_1 \ge 2$, and every hybrid average is at least $(2(j-2)+1)/(j-1) \ge$

3/2. When $1 \in J$, its order-p hybrid also gives $C_* < 2m_1$. The supplied link determines J by whether its odd part vanishes; finite searches over integer powers then determine its indices and nonzero moments. Only finite indices in J are used below.

Keep one branch of intercept one: use the quadratic matrix if $C_* = 2m_2$; otherwise choose the smallest minimizing order $h \in J \cap \{3, \dots, p\}$ and use its hybrid below. When $1 \notin J$ the quadratic branch is the only such branch; no linear pilot or moment m_1 is used. In the hybrid case $m_h < m_1$ and $C_* = 2((h-2)m_1 + m_h)/(h-1) > 2m_h$. Keep Gram orders only in $\mathcal{G} = \{j \in J : j \geq 3,\ 2m_j < C_*\}$. The discarded Gram lines are dominated by $1 + C_* a$ for $a > 0$, so the retained lines still have envelope B. The set $\mathcal{G}$ contains p, and contains h in the hybrid case. All transformed powers used are at most r_*: in the quadratic case $m_j < m_2$ for $j \in \mathcal{G}$; in the hybrid case $m_h < m_1$ and $m_j < C_*/2 < m_1$ for $j \in \mathcal{G}$.

Write $U = uu^\top, Q = I - U, d = D - 1$. All primitive blocks below have common size $n \geq 1$ and are disjoint. Let $R_j(X)$ denote the order-j normal-ordered Gaussian tensor, and let Π_j be symmetric projection. Write $J_j(F)(x) = \langle F, R_j(x)\rangle/\sqrt{j!}$; the Gaussian tensor isometry gives $\|J_j(F)\|_2 = \|\Pi_j F\|_F$ and $J_j(u^{\otimes j})(x) = h_j(u^\top x)$. For each used order define

$$b_j = \mathbb{E}[h_j(Z) f(Z)^{m_j}] \neq 0, \qquad S_{j,n} = \frac{1}{n b_j} \sum_i Y_i^{m_j} \frac{R_j(X_i)}{\sqrt{j!}}.$$

The conditional identity $\mathbb{E}[R_j(X)/\sqrt{j!} \mid Z] = h_j(Z) u^{\otimes j}$ and the expansion

$$\mathbb{E}[Y^{m_j} \mid Z] = \sum_{v=0}^{m_j} \binom{m_j}{v} f(Z)^v s^{m_j - v} \mathbb{E}\xi^{m_j - v}$$

give $\mathbb{E} S_{j,n} = u^{\otimes j}$: all powers below m_j have zero h_j moment, including the constant. No noise symmetry is used.

Here and below constants C may depend on $f, P, \mathbf{m}, M$. For every used power $m \leq r_*$, Lyapunov's inequality gives $\mathbb{E}|\xi|^{2m} \leq M$, and for $s \geq 1$, $\mathbb{E}[|Y|^{2m} \mid Z] \leq 2^{2m-1} s^{2m} (f(Z)^{2m} + M)$. For a deterministic tensor F, write its order-j chaos as $g(Z, W) = \langle R_j(X)/\sqrt{j!}, F\rangle = \sum_{v=0}^{j} h_v(Z) g_{j-v}(W)$, where W is the independent Gaussian orthogonal component. Orthogonality of different orders in W gives

$$\mathbb{E}[(f(Z)^{2m} + M) g(Z, W)^2] = \sum_{v=0}^{j} \mathbb{E}[(f(Z)^{2m} + M) h_v(Z)^2] \|g_{j-v}\|_2^2 \leq K_{j,m} \|\Pi_j F\|_F^2,$$

where $K_{j,m} = \max_{0 \leq v \leq j} \mathbb{E}[(f^{2m} + M) h_v^2] < \infty$. The tensor isometry and independence of rows therefore prove

$$\mathrm{Cov}(S_{j,n}) \preceq \omega_j \Pi_j, \qquad \omega_j = A_j s^{2m_j}/n, \qquad A_j = 2^{2m_j - 1} K_{j,m_j}/b_j^2. \tag{202}$$

These are analysis constants, not learner inputs. For every integer $p \leq q \leq P$, the untransformed tensor $C_{q,n} = n^{-1} \sum_i Y_i R_q(X_i)/\sqrt{q!}$ has mean $c_q u^{\otimes q}$ and covariance at most $(\Delta/n)\Pi_q$, where $\Delta = \max_{0 \leq v \leq P} \mathbb{E}[f(Z)^2 h_v(Z)^2] + s^2 \leq C_f s^2$. Indeed centered independent unit-variance noise gives $\mathbb{E}[Y^2 \mid Z] = f(Z)^2 + s^2$, and the same chaos decomposition proves the bound. This is the parity-free interface (180); no nonzero quadratic or linear transformed-signal condition enters its proof.

We verify the three matrix bounds

$$\mathbb{E} M = U, \qquad \mathbb{E}(u^\top M u)^2 \leq 1 + \zeta, \quad \mathbb{E}\|QMu\|^2 \leq d\gamma, \quad d^{-1}\mathbb{E}\|QMQ\|_F^2 \leq D\gamma. \tag{203}$$

For the quadratic branch, take $M_E = S_{2,n}$. Equation (202), tested against the parallel, column and block entries, gives (203) with $\zeta = \gamma = \omega_2$, hence $H_E := \zeta + D\gamma \leq C D s^{C_*}/n$.

For a hybrid, take $h - 2$ independent pilots $L_v = S_{1,n}$ and a fresh $S_{h,n}$, and put $M_H = S_{h,n}[\cdot, \cdot, L_1, \dots, L_{h-2}]$. With $\rho = \omega_1$, the pilots have mean u, $\mathrm{Var}(u^\top L_v) \leq \rho$ and $\mathbb{E}\|L_v\|^2 \leq 1 + D\rho$. Conditionally on them, the matrix mean is $\mathcal{T} U$, where $\mathcal{T} = \prod_v u^\top L_v$, and its tested variance is at most $\omega_h \|F\|_F^2 \prod_v \|L_v\|^2$. Thus (203) holds with

$$\gamma = \omega_h \mathbb{E} \prod_v \|L_v\|^2, \qquad \zeta = \mathrm{Var}\, \mathcal{T} + \gamma.$$

Indeed $\mathbb{E}\mathcal{T} = 1$, and the perpendicular entries have conditional mean zero. Independence bounds $\operatorname{Var}\mathcal{T} \le (1+\rho)^{h-2} - 1$ and $\gamma \le \omega_h(1+D\rho)^{h-2}$. For $\rho \le 1$, using $(1+x)^{h-2} \le C_h(1+x^{h-2})$ gives

$$H_H := \zeta + D\gamma \le C\left[\frac{s^{2m_1}}{n} + \frac{Ds^{2m_h}}{n} + \frac{D^{h-1}s^{C_*(h-1)}}{n^{h-1}}\right]. \tag{204}$$

For a Gram order $j \in \mathcal{G}$, flatten two fresh copies of $S_{j,n}$ as $A_v = uw^\top + E_v$, with $w = u^{\otimes(j-1)}$, and set $M_j = (A_1A_2^\top + A_2A_1^\top)/2$. The centered errors are independent and have covariance at most $\omega_j I$ on ordered entries. Expand $A_1A_2^\top = U + u(E_2w)^\top + (E_1w)u^\top + E_1E_2^\top$. Conditioning on a centered error makes all linear-bilinear cross expectations zero. The parallel linear variances sum to at most $2\omega_j$, and the bilinear variance is at most $\omega_j^2 D^{j-1}$. The perpendicular column has second moment at most $d\omega_j + d\omega_j^2 D^{j-1}$. The perpendicular block has squared norm at most $\omega_j d\mathbb{E}\|E_2^\top Q\|_F^2 \le \omega_j^2 d^2 D^{j-1}$. Convexity under symmetrization proves (203) with $\zeta = \gamma = 2\omega_j + D^{j-1}\omega_j^2$, and hence

$$H_j := \zeta + D\gamma \le C\left[\frac{Ds^{2m_j}}{n} + \frac{D^j s^{4m_j}}{n^2}\right]. \tag{205}$$

All these matrices have finite second moments. Rotating Gaussian rows by a map fixing u preserves their laws and rotates their tensors, pilots and free matrix indices. Their laws are therefore invariant under the stabilizer conjugations required by Lemma K.2.

We next define the selector and allocation at every D, N. Put $\ell = \log(eD)$,

$$e = D^{-\eta/H_*}, \qquad r = e/P, \qquad t = \lceil \log_2(100D/r) \rceil, \qquad n = \left\lfloor \frac{N}{(p-1)Pt+1} \right\rfloor.$$

If $n = 0$, output zero. Otherwise set $R = \log_D N$ and select a largest inverse value among

$$x_0 = (R-1)/C_*, \qquad x_j = (R - j/2)/(2m_j) \quad (j \in \mathcal{G}). \tag{206}$$

Break ties in favor of branch zero, then in increasing order j. Branch zero uses its fixed quadratic or hybrid matrix; branch j uses M_j. The line coefficients are fixed rationals. Comparisons can be implemented by clearing denominators, exponentiating to comparisons of integer powers of N, D, and moving negative exponents to the other side. Selection uses no noise parameter. Each matrix uses at most $p-1$ primitive blocks: one for a quadratic, $h-1$ for a hybrid and two for a Gram matrix.

Fix a, M for the analysis and suppose $N \ge \lceil D^{B+\eta} \rceil$. Eventually $n \ge N/(C_{p,P,\mathbf{m},\eta}\ell)$, uniformly in these N. All retained slopes are at most C_*. Choosing a line attaining B shows that the selected inverse x satisfies

$$x \ge a + \delta, \qquad \delta = \eta/C_*. \tag{207}$$

If branch zero is selected, $R = 1 + C_*x \ge 1 + C_*a + \eta$. Consequently $H_E \le CD^{-\eta}\ell$ in the quadratic case. In the hybrid case its own Gram order h belongs to $\mathcal{G}$. The inequality $x = x_0 \ge x_h$ gives

$$(C_* - 2m_h)x \le h/2 - 1, \qquad x \le \frac{h-1}{4(m_1 - m_h)}.$$

Since $C_* - 2m_1 = -2(m_1 - m_h)/(h-1)$, it follows that

$$R - 2m_1a = 1 + (C_* - 2m_1)x + 2m_1(x-a) \ge \frac{1}{2} + \frac{2m_1\eta}{C_*}.$$

In particular $\rho \to 0$. The three terms in (204) are respectively at most

$$CD^{-1/2-2m_1\eta/C_*}\ell, \qquad CD^{-\eta}\ell, \qquad CD^{-(h-1)\eta}\ell^{h-1}.$$

For the middle term we used $C_* > 2m_h$. If a Gram order j is selected, $R = j/2 + 2m_jx$, and (205) instead gives

$$H_j \le C\left[D^{-(j/2-1)-2m_j\eta/C_*}\ell + D^{-4m_j\eta/C_*}\ell^2\right].$$

Each displayed exponent margin is strictly larger than η/H_*: for the Gram linear term use $j/2 - 1 \ge 1/2, m_j \ge 1$ and $1/H_* \le 2/C_*$; for the quadratic term use $4m_j/C_* \ge 4/C_* > 1/H_*$. For branch

zero, $H_* \geq 3/2$ makes $\eta > \eta/H_*$; a hybrid additionally has $2m_1/C_* > 1$. Thus every matrix choice, including all ties, satisfies

$$H \leq e/(100Pt) \tag{208}$$

eventually, since all powers of ℓ and the finite set of profile constants are fixed independently of D, N.

Apply Lemma K.2 with mean parameter one and $r = e/P$. Take P independent length-t products $\mathcal{B}^{(v)}$ of the selected matrices and one fresh final block shared by all target orders. The lemma gives

$$\mathbb{E}\mathcal{B} = U, \quad K = \mathbb{E}\mathcal{B}^\top\mathcal{B} = a_t U + b_t Q, \quad a_t \leq \exp(e/(100P)), \quad \operatorname{tr} K < 2.$$

For every integer $p \leq q \leq P$ set $T_q = (\mathcal{B}^{(1)} \otimes \cdots \otimes \mathcal{B}^{(q)})C_{q,n}$ and output $\widehat{f}(x) = \sum_q J_q(\Pi_q T_q)(x)$. Independence and the final tensor covariance yield

$$\mathbb{E}T_q = c_q u^{\otimes q}, \qquad \mathbb{E}\|T_q\|_F^2 \leq c_q^2 a_t^q + (\Delta/n)\operatorname{tr}(K^{\otimes q}\Pi_q) \leq c_q^2 a_t^q + 2^q\Delta/n.$$

Symmetric projection contracts norm and fixes the target tensor. The tensor isometry and orthogonality of different query chaoses, conditionally on training, bound the full expected risk by

$$\exp(e/100) - 1 + P2^P\Delta/n \leq e/99 + e/2 < e$$

eventually. Indeed every candidate line gives $B \geq 1 + 2a$, so $\Delta/n \leq CD^{-1-\eta}\ell = o(e)$, since $H_* \geq 3/2$. This includes both parities and zero target coefficients; the full centered target has squared norm $\sum_{q=p}^P c_q^2 = 1$. Every primitive block, including pilots within each matrix, is disjoint from the others, and the allocation is at most $((p-1)Pt+1)n \leq N$. The finite set of bounds above supplies a common threshold for all asserted N, u and noise laws with the stated moment bound.

For explicit degree bounds, assign the quadratic matrix label/joint degrees $(m_2, m_2 + 2)$, a hybrid of order h the degrees

$$\big(m_h + (h-2)m_1, \quad m_h + (h-2)m_1 + 2h - 2\big),$$

and a Gram matrix of order j the degrees $(2m_j, 2m_j + 2j)$. Let d_Y, d_X be the maxima over the retained matrices of these respective degrees. The predictor has label degree at most $d_Y Pt + 1$ and joint degree at most $d_X Pt + 2P + 1$ at every D, N, including the zero fallback. Each primitive block enters each tensor output linearly in its transformed tensor. Squaring requires at most twice that block's label power, bounded by $2r_*$; the final block needs only two noise moments. Products use independent blocks, and shared outputs have different query orders. Thus growing t requires no additional noise moments and risk is finite even below the eventual sample threshold. Divisions use only supplied nonzero moments and deterministic positive quantities. Fixed-order tensor operations, label powers and matrix products have polynomial real-arithmetic cost for each fixed link and profile; no finite-bit claim is made. □

Q.2 Uniformity over noise laws with a variance bound

Heavy-tailed integrability can restrict the degree in each observation (Kunisky, 2021, Section 2.4 and Appendix B.4). Multilinear tensor estimators can use variance bounds alone (Ding et al., 2020). We combine these principles with the present cumulant count.

Let $\mathcal{N}_2$ be all probability laws with mean zero and variance one. For iid $X_i, x \sim N(0, I_D)$ and iid $\xi_i \sim \nu \in \mathcal{N}_2$ independent of them, observe $Y_i = f(u^\top X_i) + s\xi_i$. For a predictor $\widehat{F}$ whose coefficients do not depend on ν, put

$$\mathcal{R}_2(\widehat{F}) = \sup_{\nu \in \mathcal{N}_2} \sup_{\|u\|=1} \mathbb{E}_{\nu,u}(\widehat{F}(x) - f(u^\top x))^2.$$

Coefficients may depend measurably on the entire design, query and independent seed, as in (3); the lower also permits knowledge of s.

Proposition Q.6. *Fix integers $3 \leq p \leq P$, a real $\eta > 0$, and a supplied fixed centered link $f = \sum_{q=p}^P c_q h_q$, with $c_p \neq 0$ and $\sum_q c_q^2 = 1$. Use C_P, Γ_P from (195) and put*

$$D_0 = \left\lceil \max\{2, C_P^{4\Gamma_P/\eta}, 2^{2/\eta}\} \right\rceil, \qquad L_- = \lfloor D^{\eta/(4\Gamma_P)} \rfloor.$$

Let E_0 be (196); thus $E_0 \le C_P^0 D^{-\lceil p/2 \rceil}$. For every fixed $a > 0$, set $s = D^a$ and $b = p/2 + 2a$. For integers $D \ge D_0$ and $0 \le N \le D^{b-\eta}$, every law-independent $\widehat{F} \in \mathcal{P}_{L_-}$ satisfies

$$\mathcal{R}_2(\widehat{F}) \ge 1 - E_0 - 2D^{-\eta/2}.$$

For $w(z) = 2/[\pi(1+z^2)^2]$, the same lower bound holds for $\sup_{\|u\|=1} \mathbb{E}_{w,u}(\widehat{F}(x) - f(u^\top x))^2$, even when w is supplied to the predictor. There is a predictor collection indexed by integers $D \ge 2$, $N \ge 0$ and depending only on f, p, P, D, N, η, with joint degree $O_{p,P,\eta}(\log(eD))$, such that eventually in D, for every integer $N \ge \lceil D^{b+\eta} \rceil$,

$$\mathcal{R}_2(\widehat{f}) \le D^{-\eta/4}.$$

The dimension threshold may depend on f, p, P, a, η. No noise scale, law or higher-moment bound is supplied to this collection.

Proof. The density w defines a law in $\mathcal{N}_2$: substitution $z = \tan\theta$ gives equal normalizing and second-moment integrals $\pi/2$, and symmetry gives mean zero. Its fourth moment is infinite. Assume the worst-teacher risk under w is finite; otherwise the lower is immediate.

Fix a teacher u_0. By Tonelli, for almost every design, query and seed, the squared error has finite integral over the product label law with noise density w. Collect the polynomial in one label Y_i. If its degree in Y_i were at least two, its nonzero leading coefficient polynomial would vanish on a Lebesgue null set of the other labels; the null-set fact follows by induction in the variables. Outside that set the squared error has a positive leading even power of degree at least four. Its integral against any shifted, positively scaled w is infinite. Every other label has a strictly positive density, contradicting Tonelli. Thus every label has degree at most one. The coefficient null set lies in design, query and seed and works for every teacher. Multilinear squared loss uses only the first two noise moments, so its conditional risk equals the Gaussian-noise risk.

Scaling labels by s preserves multilinearity. Take the sign teacher prior and normalized Gaussian-label model of (193). The finite-risk projection onto $\bigoplus_{q=p}^{P} \mathcal{H}_q$ in Theorem Q.3 applies: the finite Gaussian monomial Gram matrix is positive definite. Projection preserves multilinearity. Let $\mathcal{A} = \{\alpha \in \{0,1\}^N : |\alpha| \le L_-\}$ and write $\widehat{F}(y) = \sum_{\alpha\in\mathcal{A}} b_\alpha y^\alpha$ after projection. Use $m_\alpha, v_\alpha, \kappa_\alpha$ from (194), conditionally on design and seed. The set $\mathcal{A}$ is downward closed. Expanding $(T+G)^\alpha$, Jensen in the teacher and Gaussian orthogonality give

$$\mathbb{E}\|\widehat{F}(T+G)\|^2 \ge \sum_{\beta\in\mathcal{A}} \|d_\beta\|^2, \qquad d_\beta = \sum_{\substack{\alpha\in\mathcal{A}\\ \alpha\ge\beta}} m_{\alpha-\beta} b_\alpha.$$

All binary-index factorials and binomial coefficients are one. The recursion gives $v_\alpha = \sum_{\beta\le\alpha} \kappa_\beta m_{\alpha-\beta}$, so the target cross term is $\sum_\beta \langle \kappa_\beta, d_\beta\rangle$. Completing squares and averaging proves the restricted certificate

$$\mathbb{E}\|\widehat{F} - F_U\|^2 \ge 1 - \sum_{\alpha\in\mathcal{A}} \mathbb{E}\|\kappa_\alpha\|^2.$$

Each retained degree-k index has support k. Since $m_p = 1$ and $m_j \ge 2$ for $j < p$, (190) has only the order-p point feasible at $x = 1$, giving $\phi(1) = p/2$. The fixed-index bound (201) and $\binom{N}{k} \le N^k$ give, for every $k \ge 1$,

$$\sum_{\substack{\alpha\in\mathcal{A}\\ |\alpha|=k}} \mathbb{E}\|\kappa_\alpha\|^2 \le (C_P k)^{\Gamma_P k} (N/s^2)^k D^{-p(k+1)/2} \le (C_P k)^{\Gamma_P k} D^{-\eta k - p/2}.$$

For $k \le L_-$, $(C_P k)^{\Gamma_P} \le D^{\eta/2}$ by $D \ge D_0$, so all positive energies sum to at most $\sum_{k\ge1} D^{-\eta k/2} \le 2D^{-\eta/2}$. The zeroth energy is E_0. Worst-teacher w-risk therefore dominates this Gaussian-prior average. For $N = 0$ or degree zero the certificate gives $1 - E_0$, also covering $b < \eta$, when $N = 0$ is forced.

For the upper set

$$\epsilon = D^{-\eta/4}, \quad r = \epsilon/P, \quad t = \lceil \log_2(100D/r) \rceil, \quad L_0 = 2Pt + 1, \quad n = \lfloor N/L_0 \rfloor.$$

Return zero if $n = 0$. For each primitive block put $C_{q,n} = n^{-1}\sum_i Y_i R_q(X_i)/\sqrt{q!}$. Form Pt Gram matrices (49), each from two independent normalized tensors $C_{p,n}/c_p$, and P independent length-t products $\mathcal{B}^{(v)}$. A single fresh final block supplies $C_{q,n}$ for all $p \le q \le P$. Ignore unused observations. These choices involve no noise parameter.

For every $\nu \in \mathcal{N}_2$, the conditional label moments are $f(u^\top X)$ and $f(u^\top X)^2 + s^2$. The parity-free chaos calculation (180) gives mean $c_q u^{\otimes q}$ and covariance at most $(\Delta/n)\Pi_q$ for $C_{q,n}$, and mean $u^{\otimes p}$ with the same covariance bound $(\Delta/n)\Pi_p$ for $C_{p,n}/c_p$, where

$$M_f = \max_{0\le v\le P} \mathbb{E}[f(Z)^2 h_v(Z)^2], \qquad K_f = (M_f+1)/c_p^2,$$
$$\Delta = (M_f + s^2)/c_p^2 \le K_f s^2 \quad (s \ge 1).$$

Independence and this covariance are precisely the hypotheses used in (51)–(53), which use no parity; they give (41) with $\gamma = 2\Delta/n + \Delta^2 D^{p-1}/n^2$ whenever

$$n \ge 400\Delta D t/r + 20\Delta D^{p/2}\sqrt{t/r}.$$

Indeed the two terms of $D\gamma$ are at most $r/(200t)$ and $r/(400t)$. Gaussian design and independent scalar noise preserve conjugation invariance under rotations fixing u. All matrix second moments are finite. Lemma B.3 gives $\mathbb{E}\mathcal{B} = uu^\top$, $K = \mathbb{E}\mathcal{B}^\top\mathcal{B}$, $Ku = a_t u$, $a_t \le e^{\epsilon/(100P)}$, and $\operatorname{tr} K < 2$. Set

$$T_q = (\mathcal{B}^{(1)} \otimes \cdots \otimes \mathcal{B}^{(q)})C_{q,n}, \qquad \widehat{f} = \sum_{q=p}^{P} J_q(\Pi_q T_q).$$

Here J_q is the Gaussian tensor isometry. Independence gives $\mathbb{E}T_q = c_q u^{\otimes q}$ and $\mathbb{E}\|T_q\|^2 \le c_q^2 a_t^q + 2^q\Delta/n$. Projection contracts norm. Orthogonality of distinct query chaoses and $\sum_q c_q^2 = 1$ give

$$\mathcal{R}_2(\widehat{f}) \le e^{\epsilon/100} - 1 + P2^P\Delta/n < \epsilon \quad \text{if } n \ge 2^{P+1}P\Delta/\epsilon,$$

since $e^{\epsilon/100} - 1 \le \epsilon/99$. Only the first two noise moments enter these calculations.

Write $\ell = \log(eD)$. The two required sample quantities divided by any $N \ge D^{p/2+2a+\eta}$ are bounded, respectively, by

$$C_{f,p,P,\eta}\big(D^{1-p/2-3\eta/4}\ell + D^{-7\eta/8}\sqrt{\ell}\big), \qquad C_{f,P}D^{-p/2-3\eta/4},$$

since $\Delta \le K_f D^{2a}$ and $t = O_{P,\eta}(\ell)$. Each is eventually at most $1/(2L_0)$; also $N \ge 2L_0$, so $n \ge N/(2L_0)$ meets both requirements, uniformly over N and ν. Every tensor is linear in each of its labels, and all primitive blocks are disjoint. The output is multilinear, has label degree at most $2Pt+1$, and joint degree at most $2Pt(p+1) + 2P + 1 = O_{p,P,\eta}(\log(eD))$. This proves both bounds. □

Corollary Q.7. *Fix integers $3 \le p \le P$ and $\eta > 0$. There are predictors indexed by integers $D \ge 2$, $N \ge 0$, depending only on p, P, D, N, η, of joint degree $O_{p,P,\eta}(\log^3(eD))$, such that for every fixed $f = \sum_{q=p}^{P} c_q h_q$ with $c_p \ne 0$, $\sum_q c_q^2 = 1$, and $a > 0$, eventually in D, with $s = D^a$,*

$$N \ge \lceil D^{p/2+2a+\eta}\rceil \quad \Longrightarrow \quad \mathcal{R}_2(\widehat{f}) \le D^{-\eta/4}.$$

The dimension threshold may depend on f, p, P, a, η. No coefficients, noise scale, noise law or higher-moment bound are supplied.

Proof. Put $I = \{p, \dots, P\}$, $\ell = \log(eD)$, $\epsilon = D^{-\eta/4}$, $\mu_0 = \ell^{-1/2}$ and $B = 4P + 3$. Set

$$r = \epsilon/P, \quad t = \lceil \log_2(100D/r)\rceil, \quad \nu = r/(1000t),$$
$$m = \left\lceil 4B^2\mu_0^{-2}\log(1000t/r)\right\rceil, \quad L_0 = 4mPt + 1, \quad n = \lfloor N/L_0\rfloor.$$

Return zero if $n = 0$. Otherwise form Pt matrices $R_m A_T$ using (124) and $2m-1$ independent tensor traces per matrix. Their P independent length-t products give (144), summed over I. Use $4mPt$ disjoint primitive blocks and one final block, all of size n.

For analysis put

$$C = \max_{0\le j\le P} \mathbb{E}\left[h_j(Z)^2 \sum_{q\in I} h_q(Z)^2\right], \qquad \Delta = C + s^2, \qquad \mu = c_p^2.$$

Cauchy–Schwarz and (180) give (140) for every $q \in I$ using only noise variance. The tensor calculation in Lemma K.5 gives

$$\gamma_{\text{raw}} = 2\Delta/n + \Delta^2 D^{p-1}/n^2, \qquad \mathrm{Var}(T) \le 2\Delta/n + \Delta^2 D^p/n^2.$$

Eventually $\mu \ge \mu_0$. With $v = \mu_0^2\nu$ and $\gamma_0 = \gamma_{\text{raw}}/\mu_0^2$, the bounds

$$n \ge n_* := \left\lceil 4\Delta D/(\mu_0^2\nu) + 2\Delta D^{p/2}/(\mu_0\sqrt{\nu})\right\rceil,$$
$$n \ge n_{\text{out}} := \lceil 2^{P+1}|I|\Delta/\epsilon\rceil$$

give $\mathrm{Var}(T) \le v/2$ and $D\gamma_0 \le 3\nu/4$. Lemmas K.3 and K.2 therefore give (143). Orthogonality of all query orders in (144) yields

$$\mathcal{R}_2(\widehat{f}) \le e^{\epsilon/100} - 1 + \epsilon/500 + |I|2^P\Delta/n < \epsilon.$$

Each primitive label occurs multilinearly; disjoint blocks make second moments sufficient throughout, including scalar calibration.

For $S_T = s^2 D^{p/2}$, the definitions give

$$m = O_{P,\eta}(\ell^2), \quad L_0 = O_{P,\eta}(\ell^3), \quad \max\{n_*, n_{\text{out}}\} \le C_{p,P,\eta} S_T D^{\eta/4}\ell^2.$$

Thus for every $N \ge S_T D^\eta$, eventually $\max\{n_*, n_{\text{out}}\}/N \le 1/(2L_0)$ and $N \ge 2L_0$; hence $n \ge N/(2L_0)$ suffices. The tensor degree count (139) gives joint degree at most $4mP(p+1)t + 2P + 1$. All construction parameters use only the stated indices. □

Q.3 A FIXED LINK WITH TWO VANISHING SIGNALS

The following exact arithmetic construction separates the quadratic, cubic-label and fifth-label signals.

Proposition Q.8. *There is a fixed odd polynomial $f = \sum_{q=3,5,\dots,21} c_q h_q$ with $\sum_q c_q^2 = 1$, $c_3 > 0$ and $c_{21} > 0$ such that, for $Z \sim N(0,1)$,*

$$\mathbb{E}[(Z^2-1)f(Z)^2] = 0, \qquad \mathbb{E}[Zf(Z)^3] = 0, \qquad \mathbb{E}[Zf(Z)^5] > 0.$$

Proof. Write $\mathrm{He}_q = \sqrt{q!}\, h_q$ and $R = \sum_{q=3,5,\dots,17} a_q \mathrm{He}_q$. Define its rational coefficients by $a_1 = 0$, $a_3 = 247005158400/6781$ and

$$a_{q+2} = -\frac{2qa_q + a_{q-2}}{(q+1)(q+2)} \quad (q = 3,5,\dots,15). \tag{209}$$

This gives $a_{15} = 3864848/101715$ and $a_{17} = -1$. Put $\nu = 267778621561651200/6781$ and $v = 203269/306817194600$. The identity $\mathrm{He}_2\mathrm{He}_q = \mathrm{He}_{q+2} + 2q\mathrm{He}_q + q(q-1)\mathrm{He}_{q-2}$, Gaussian orthogonality and (209) give

$$\mathbb{E}[\mathrm{He}_2 R^2] = 17!(34 - a_{15}) = -36\nu, \qquad \mathbb{E}[\mathrm{He}_2 R\mathrm{He}_3] = 0,$$
$$\mathbb{E}[\mathrm{He}_2 R\mathrm{He}_{21}] = 0, \qquad \mathbb{E}[\mathrm{He}_2\mathrm{He}_3\mathrm{He}_{21}] = 0,$$
$$\mathbb{E}[\mathrm{He}_2\mathrm{He}_3^2] = 36, \qquad \mathbb{E}[\mathrm{He}_2\mathrm{He}_{21}^2] = 42\cdot 21!.$$

Also $42\cdot 21!\, v = 36\nu$. For $0 < t < 1$, define

$$F_t = R + \sqrt{\nu}\frac{1-t^2}{1+t^2}\mathrm{He}_3 + \sqrt{v}\frac{2t}{1+t^2}\mathrm{He}_{21}, \qquad A_t = (1+t^2)F_t. \tag{210}$$

Since $(1-t^2)^2 + 4t^2 = (1+t^2)^2$, these identities imply $\mathbb{E}[\mathrm{He}_2 F_t^2] = 0$ throughout this interval.

Let $G_j(t) = \mathbb{E}[ZA_t(Z)^j]$ for $j = 3, 5$, and set

$$a = \frac{182325627249912264 5509}{10^{22}}, \qquad b = \frac{182325627249912264 5510}{10^{22}}.$$

The exact interval computation specified below gives the inclusive bounds

$$\begin{aligned} G_3(a) &\in [-245596, -245595], \\ G_3(b) &\in [3897302, 3897303], \\ G_5(t) &\in [2 \cdot 10^{61}, 3 \cdot 10^{61}] \quad (a \le t \le b). \end{aligned} \tag{211}$$

Here are all operations needed to verify (211). Generate the monomial coefficients of He_q from $\mathrm{He}_0 = 1, \mathrm{He}_1 = z$ and $\mathrm{He}_{q+1} = z\mathrm{He}_q - q\mathrm{He}_{q-1}$. For each positive rational $q = N/Q$ in $\{\nu, v\}$, set $B = 10^{80}$ and $m = \lfloor\sqrt{\lfloor NB^2/Q \rfloor}\rfloor$. The integer inequalities $m^2Q \le NB^2 < (m+1)^2Q$ certify $\sqrt{q} \in [m/B, (m+1)/B]$. For rational intervals $I = [i_-, i_+]$ and $J = [j_-, j_+]$, use

$$I + J = [i_- + j_-, i_+ + j_+], \qquad IJ = [\min_{\epsilon,\eta\in\{-,+\}} i_\epsilon j_\eta, \max_{\epsilon,\eta\in\{-,+\}} i_\epsilon j_\eta].$$

Using these radical intervals, form the coefficient intervals of $A_t = (1+t^2)R + (1-t^2)\sqrt{\nu}\,\mathrm{He}_3 + 2t\sqrt{v}\,\mathrm{He}_{21}$, first with $t \in [a, a]$, then $[b, b]$, and then $[a, b]$. Repeated interval convolution gives its third or fifth power. Replace each coefficient interval of z^r in that power by its product with $\mathbb{E}Z^{r+1}$ and sum, using

$$\mathbb{E}Z^{2\ell} = \frac{(2\ell)!}{2^\ell \ell!}, \qquad \mathbb{E}Z^{2\ell+1} = 0.$$

These finite rational operations yield (211). The self-contained verifier `verify_double_cancellation.py` performs them using integer arithmetic, rational fractions and integer square roots, and checks the displayed endpoint inequalities. Each interval operation contains every possible value of its arguments; induction through convolution therefore proves the bounds simultaneously for every $t \in [a, b]$, despite repeated occurrences of t and the radicals. No floating-point comparison enters this verification.

The polynomial G_3 is continuous, so its opposite endpoint signs give $t_* \in (a, b)$ with $G_3(t_*) = 0$. The third bound in (211) gives $G_5(t_*) > 0$. The He_{17} coefficient of F_{t_*} is -1, so its Gaussian norm is positive. Its He_3 coefficient is $a_3 + \sqrt{\nu}(1-t_*^2)/(1+t_*^2) > 0$, and its He_{21} coefficient is positive. Thus $f = F_{t_*}/\sqrt{\mathbb{E}F_{t_*}(Z)^2}$ has exactly the stated rank, degree and normalization. The positive normalization and $G_j(t) = (1+t^2)^j\mathbb{E}[ZF_t^j]$ preserve both zero moments and the positive fifth moment. This proves the proposition. □

Q.4 The possible rank-three exponents

For fixed normalized odd polynomials of Hermite rank three, (192) gives $\kappa = \min\{2m_2, m_1 + 1\}$. We prove that its possible values are exactly the even integers at least four. The degree of the realizing polynomial may depend on κ. The bounded smooth links with prescribed generative exponent in Damian et al. (2024, arXiv version 2, Theorem 5.1) concern a different index: odd polynomials have generative exponent one (Damian et al., 2024, arXiv version 2, Example 2.7). Here we prescribe transformed moments so that κ has any allowed value.

Write γ_1 for standard Gaussian measure and $\varphi(z) = (2\pi)^{-1/2}e^{-z^2/2}$ for its density. The next lemma supplies the polynomial approximation used in the construction.

Lemma Q.9. *For every $1 < q < \infty$, odd real polynomials are dense in the closed subspace of odd functions in $L^q(\gamma_1)$.*

Proof. If their closure were proper, the Hahn–Banach theorem and L^q duality would give a nonzero continuous linear functional on the odd subspace that vanishes on all odd polynomials. Represent it by $v \in L^{q'}(\gamma_1)$, where $q' = q/(q-1)$, and replace $v(z)$ by $(v(z) - v(-z))/2$. This replacement preserves the functional on odd functions and gives a nonzero odd v. Every odd moment of $v\,d\gamma_1$ vanishes by the assumed annihilation, and every even moment vanishes by symmetry. For every complex t, Hölder's inequality gives $\mathbb{E}[|v(Z)|e^{|t||Z|}] < \infty$. Absolute summability therefore permits termwise integration of the exponential series and gives

$$\mathbb{E}[v(Z)e^{tZ}] = \sum_{j=0}^{\infty} \frac{t^j}{j!}\mathbb{E}[v(Z)Z^j] = 0.$$

Taking t purely imaginary shows that the finite signed measure $v\,d\gamma_1$ has zero Fourier transform. Uniqueness of the Fourier transform for finite signed measures implies $v = 0$ almost everywhere, a contradiction. □

Proposition Q.10. *For every integer $R \geq 1$, there is an odd real polynomial f such that*

$$\begin{aligned} \mathbb{E}[f(Z)^2] &= 1, & \mathbb{E}[h_3(Z)f(Z)] &> 0, \\ \mathbb{E}[Zf(Z)^{2j+1}] &= 0 & &(0 \leq j < R), \\ \mathbb{E}[Zf(Z)^{2R+1}] &\neq 0, & & \\ \mathbb{E}[(Z^2-1)f(Z)^{2j}] &= 0 & &(1 \leq j \leq R). \end{aligned}$$

In particular, f has Hermite rank three, $m_1 = 2R+1$, $m_2 \geq 2R+2$, and $\kappa = 2R+2$.

Proof. Put $L = R + 1$. The measures with densities $(1 - z^2)\varphi(z)$ on $(0, 1)$ and $(z^2 - 1)\varphi(z)$ on $(1, \infty)$ both have total mass $\varphi(1) > 0$, by integration of $(z\varphi(z))'$. Partition each interval into L consecutive intervals of equal mass for its respective measure. Such partitions exist because the cumulative masses are continuous and strictly increasing in the interior of each interval. Pair the lth interval from each partition to form B_l. Up to endpoints, the sets B_l partition $(0, \infty)$, and

$$\int_{B_l} (z^2 - 1)\varphi(z)\,dz = 0. \tag{212}$$

The finite measure $z\varphi(z)\mathbf{1}_{B_l}(z)\,dz$ has positive mass and no atoms. Choose a finite median $r_l > 0$ whose mass on each side is half its total mass. Set $\sigma_l(z) = \operatorname{sign}(z - r_l)$ on B_l; its value at r_l does not affect any integral. Then

$$\int_{B_l} z\sigma_l(z)\varphi(z)\,dz = 0, \qquad \int_{B_l} z^3\sigma_l(z)\varphi(z)\,dz = \int_{B_l} z|z^2 - r_l^2|\varphi(z)\,dz > 0. \tag{213}$$

For the equality, subtract r_l^2 times the first integral and use $z > 0$. Strict positivity holds because B_l has positive Lebesgue measure. Both sign sets have positive Gaussian measure, since they have equal positive $z\varphi(z)$ mass.

Choose distinct positive numbers $a_1, \dots, a_L$ and define $g(z) = a_l \sigma_l(z)$ for $z \in B_l$. Extend g oddly to the negative half-line, with arbitrary odd values at the finitely many partition endpoints. This gives a bounded odd function. Equations (212)–(213) imply

$$\mathbb{E}[Zg(Z)^{2j+1}] = 0 \quad (j \geq 0), \qquad \mathbb{E}[(Z^2-1)g(Z)^{2j}] = 0 \quad (j \geq 1), \qquad \mathbb{E}[h_3(Z)g(Z)] > 0. \tag{214}$$

Indeed, each odd-power integral is a positive constant times the first integral in (213); each even-power integral is a constant times (212). For the cubic coefficient, $\mathrm{He}_3(z) = z^3 - 3z$ and both integrals in (213) give a strictly positive contribution from every B_l.

Let $q = 2R + 2$, and let $\mathcal{X}$ be the odd subspace of $L^q(\gamma_1)$. Define a map $\Phi : \mathcal{X} \to \mathbb{R}^{2R+1}$ with coordinates

$$A_j(u) = \mathbb{E}[Zu(Z)^{2j+1}] \quad (0 \leq j \leq R), \qquad C_j(u) = \mathbb{E}[(Z^2-1)u(Z)^{2j}] \quad (1 \leq j \leq R).$$

This is a continuously differentiable map. More explicitly, each coordinate of degree $m < q$ comes from a continuous m-linear form: Hölder's inequality bounds $|\mathbb{E}[w(Z)u_1(Z)\cdots u_m(Z)]|$ by $\|w\|_{q/(q-m)} \prod_i \|u_i\|_q$, and both weights $w(z) = z, z^2 - 1$ have all finite Gaussian moments. Its derivatives at g are

$$DA_j(g)[v] = (2j+1)\mathbb{E}[Zg(Z)^{2j}v(Z)], \qquad DC_j(g)[v] = 2j\mathbb{E}[(Z^2-1)g(Z)^{2j-1}v(Z)].$$

These $2R + 1$ derivative functionals are linearly independent. A vanishing linear combination has an odd representing kernel of the form

$$zP(g(z)^2) + (z^2-1)g(z)Q(g(z)^2), \qquad \deg P \leq R, \quad \deg Q \leq R - 1.$$

It is zero almost everywhere because it annihilates every odd function in $\mathcal{X}$. On the positive-measure set where $z \in B_l$ and $g(z) = a_l$, it is the polynomial $zP(a_l^2) + (z^2-1)a_l Q(a_l^2)$ in z. A polynomial that is zero on a positive-measure set is identically zero, so $P(a_l^2) = Q(a_l^2) = 0$. There are $L = R + 1$ distinct such values; hence $P = Q = 0$. The nonzero derivative multipliers $2j + 1$ and $2j$ then show that every coefficient of the original combination was zero. Thus $D\Phi(g)$ maps $\mathcal{X}$ onto $\mathbb{R}^{2R+1}$. By Lemma Q.9, its image on odd polynomials is a dense linear subspace of this finite-dimensional space and is therefore the whole space. We may choose odd polynomials $v_1, \dots, v_{2R+1}$ with $D\Phi(g)[v_i] = e_i$.

We now correct the moments exactly. Write $Vt = \sum_i t_i v_i$, and choose odd polynomials $p_n \to g$ in $L^q(\gamma_1)$ using Lemma Q.9. Let $y_n \in \mathbb{R}^{2R+1}$ have coordinate A_R equal to $1/n$ and all other coordinates zero. Since $D_t\Phi(g + Vt)|_{t=0} = I$ and the derivative is continuous, there is a radius $\rho > 0$ such that, for all sufficiently large n and all $\|t\|_2 \leq \rho$,

$$\|I - D_t\Phi(p_n + Vt)\|_{\mathrm{op}} \leq \tfrac{1}{2}, \qquad \|y_n - \Phi(p_n)\|_2 \leq \rho/2.$$

The second inequality uses $\Phi(g) = 0$ and $y_n \to 0$. Consequently $T_n(t) = t - \Phi(p_n + Vt) + y_n$ maps the closed Euclidean ball of radius ρ into itself and has Lipschitz constant at most $1/2$. The contraction mapping theorem gives a fixed point t_n, for which

$$\Phi(p_n + Vt_n) = y_n, \qquad \|t_n\|_2 \leq 2\|y_n - \Phi(p_n)\|_2 \longrightarrow 0.$$

The odd polynomials $u_n = p_n + Vt_n$ therefore converge to g in L^q. Since $q \geq 4$, they also converge in L^2, and their cubic Hermite coefficients converge to the positive coefficient of g by Hölder's inequality. Choose one sufficiently large n and set $f = u_n/\|u_n\|_2$. This is a fixed odd polynomial with positive cubic coefficient and the exact prescribed zero moments. Its $(2R+1)$st odd-power moment against Z is nonzero because that coordinate of y_n is $1/n$.

Normalization preserves each zero and nonzero homogeneous moment. Oddness removes all even Hermite coefficients, and $A_0(f) = 0$ removes the first one, so the positive cubic coefficient gives Hermite rank three. Parity removes the intervening transformed moments, yielding $m_1 = 2R + 1$ and $m_2 \geq 2R + 2$. Lemma Q.1 makes m_2 finite. Finally, (192) gives $\kappa = \min\{2m_2, m_1 + 1\} = 2R + 2$. □

Corollary Q.11. *Among fixed normalized odd polynomial links of Hermite rank three, the possible values of κ are exactly $4, 6, 8, \dots$. For every fixed $a > 0$, there is such a fixed link whose Gaussian-noise polynomial prediction exponent is $3/2+2a$, while its unrestricted statistical exponent is $1+2a$. The link and its degree may depend on a and remain fixed as $D \to \infty$.*

Proof. Lemma Q.1 gives finite odd $m_1 \geq 3$ and even $m_2 \geq 2$. Thus $\kappa = \min\{2m_2, m_1 + 1\}$ is even and at least four. Proposition Q.10 realizes every such value. For fixed $a > 0$, choose an even $\kappa \geq \max\{4, 2 + 1/(2a)\}$ and a realizing link. The matching profile bounds in Corollary Q.4 and Proposition Q.5 give

$$B_f(a) = \min\{1 + \kappa a, 3/2 + 2a\} = 3/2 + 2a.$$

The unrestricted exponent is $1 + 2a$ by Corollary 2.5, since this link has a fixed finite degree. All choices were made with a fixed and before allowing D to grow. □