The Conditional Geometry of Statistical Verification

Thomas Y. Chen

Department of Computer Science
Columbia University

October 2026

Abstract

When can an untrusted prover reduce the number of private examples needed to certify near-optimal prediction? We study protocols that are complete under Massart noise and sound for arbitrary labels within a promised feature family. Both parties use independent samples.

Our main result identifies the conditional geometry that exactly classifies the dimension cost of verification for every known totally bounded family of Borel rules on a fixed number of Gaussian projections, with both rule and projection unknown. Two conditional error-allocation programs, comparing ordered rule pairs, give necessary and sufficient criteria for dimension-independent, $\Theta(\sqrt{d})$, or $\Theta(d)$ private sample complexity at sufficiently small fixed accuracy and constant confidence. The same programs produce both certificates and matching statistical obstructions. Upper bounds use one message from a sample-only prover; lower bounds allow arbitrary interaction and improper outputs.

For all unknown Boolean combinations of at most K orthogonal homogeneous halfspaces, we obtain verifier complexity $2^{O(K)}\varepsilon^{-1}\log(e/(\varepsilon\delta))$ at an explicit noise allowance $2^{-3K}/(2^{14}K^4)$. Dimension-independent verification at every accuracy requires exponential noise decay in rank. The upper bound gives subpolynomial verifier samples for $K = o(\log d)$ at fixed accuracy and confidence. Shifting one threshold in a balanced rank-two parity nevertheless changes dimension-independent verification to $\Theta(d)$, while preserving balance, rank, threshold degree, and VC-dimension order.

For fixed thin slabs in a specified positive-noise range, we determine the joint cost $\widetilde{\Theta}(d + 1/\varepsilon)$ at sufficiently small accuracy: local scalar certificates and distant hidden improvements contribute additive costs. For halfspaces above the first-moment signal threshold, localized certificates also yield $\widetilde{O}(\sqrt{d}/\varepsilon)$ verification under unknown centered isotropic smooth feature laws whose potential Hessians lie in a fixed neighborhood of I. This neighborhood is independent of dimension and accuracy and includes laws asymptotically at total variation distance one from Gaussian. The results concern sample complexity, without a general runtime claim.

Contents

1 Introduction **1**
- 1.1 An exact classification and its consequences . . . 1
- 1.2 Relation to prior work . . . 2

2 Model and Preliminaries **3**

3 The Conditional-Moment Classification **4**
- 3.1 A worked example: rotating a halfspace . . . 5
- 3.2 Bounded conditional duals and lifting . . . 6
- 3.3 Certificate construction and matching obstructions . . . 7

4 Quantitative Families and Geometric Separations **8**
- 4.1 Explicit geometry at every overlap . . . 8
- 4.2 Moving one threshold changes the dimension exponent . . . 9
- 4.3 Dimension and accuracy costs can add . . . 10

5 Verification under Unknown Smooth Feature Laws **11**
- 5.1 A certificate stable at every angular scale . . . 11
- 5.2 Compression without a conditional oracle . . . 12

6 Conclusion **12**

A Classification for Gaussian projection rules **13**
- A.1 Model and the two explicit conditional programs . . . 13
- A.2 Uniform bounded duals for arbitrary disagreement sets . . . 14
- A.3 Prediction nets for an arbitrary fixed Borel rule . . . 15
- A.4 Conditional lifting and the residual certificate . . . 16
- A.5 Finite-sample norm tests . . . 17
- A.6 The square-root-dimensional protocol . . . 18
  - A.6.1 An independent sample-only honest prover . . . 19
- A.7 The dimension-independent and linear upper bounds . . . 20
- A.8 Noise thresholds . . . 21
- A.9 Unknown intrinsic rules from a totally bounded family . . . 21

B Lower bounds from conditional moment obstructions **24**
- B.1 Measurable conditional allocations . . . 25
- B.2 A labeled random-frame construction . . . 25
- B.3 Multivariate Gaussian contraction . . . 26
- B.4 Haar-frame moments and unrestricted testing . . . 26
- B.5 Interactive and improper verification reduction . . . 27

C Orthogonal Boolean Compositions and the Quadratic Separation **28**
- C.1 Geometry modulo the finite symmetry group . . . 28
- C.2 A uniform negative conditional allocation margin . . . 31
- C.3 Finite protocol and sample accounting . . . 32
- C.4 The Boolean composition can also be unknown . . . 32
- C.5 Growing rank and necessary noise decay . . . 34

C.6 Optimal inverse-accuracy dependence . . . 35
C.7 An exact balanced quadratic obstruction . . . 36

D Halfspace geometry and a localized dual certificate **37**
D.1 The signed conditional moment problem . . . 38
D.2 An exact threshold and a single crossing . . . 39
D.3 A uniform angular margin . . . 40
D.4 Revealed coordinates and the certificate . . . 41
D.5 The complete halfspace dimension classification . . . 42

E Dimension-independent halfspace verification at low noise **43**
E.1 Model and constants . . . 43
E.2 An orientation-free lower bound from a band error count . . . 44
E.3 The finite-angle protocol . . . 45
E.4 Global soundness . . . 46
E.5 Sample-only Massart completeness . . . 46

F A nonlinear geometric separation: centered slabs **47**
F.1 An explicit linear-dimensional obstruction . . . 47
F.2 Every fixed slab also has a positive low-noise regime . . . 48
F.3 Additive dimension and accuracy costs . . . 49
F.4 Geometry and an exact local scalar margin . . . 49
F.5 A finite-resolution check for distant competitors . . . 51
F.6 The complete upper protocol . . . 52
F.7 Unrestricted matching lower bounds . . . 54
F.7.1 A linear dimension obstruction without validation cost . . . 54
F.7.2 The accuracy and confidence obstruction . . . 56

G Verification under an unknown smooth feature law **57**
G.1 Distribution family and guarantee . . . 57
G.2 Low-dimensional densities and conditional drift . . . 57
G.3 The observable certificate remains valid at every scale . . . 59
G.4 Localization and compression without Gaussian independence . . . 60
G.5 Finite-sample verifier and honest construction . . . 61
G.6 Nonparametric scope . . . 63

H A localized square-root lower bound for halfspaces **64**
H.1 Conditional allocation and its small-angle gain . . . 64
H.2 Local alternatives for every amplitude below the threshold . . . 65
H.3 A rank-one mixture bound, without an SQ restriction . . . 66
H.4 Implication for interactive verification with local soundness . . . 67

I Further comparison with prior work **68**
I.1 PAC verification and statistical certificates . . . 68
I.2 Low-dimensional Gaussian functions and halfspace testing . . . 69
I.3 Testable learning, Massart noise, and polynomial thresholds . . . 70
I.4 Geometric and lower-bound ingredients . . . 71

1 Introduction

An untrusted prover proposes a predictor learned from a large dataset. How many private examples suffice to certify that its risk is close to the best in a comparator class? An error estimate alone cannot rule out a better predictor hidden in directions that the proposal does not reveal. We characterize exactly how the geometry of hidden improvement determines the dimension dependence of verification.

The distinction is already visible for parity of two orthogonal Gaussian halfspaces. At sufficiently low positive noise, zero thresholds permit dimension-independent verification. Moving one threshold to a suitable fixed offset forces $\Theta(d)$ samples at the same noise level. Balance, intrinsic rank, minimum threshold degree, and VC-dimension order are unchanged. Our main theorem identifies the conditional geometric invariant behind this separation.

Completeness covers every ambient-dependent Massart noise function bounded by the specified rate. Under the promised feature law, soundness certifies excess risk for arbitrary stochastic labels, even when this noise assumption fails. Both parties receive independent ordinary examples.

1.1 An exact classification and its consequences

Theorem 1.1 (Main classification, informal)**.** *Fix a known totally bounded family $\mathcal{F}$ of Borel Boolean rules on $\mathbb{R}^k$ and a positive Massart allowance below $1/2$. For $x \mapsto f(W^\top x)$ with both $f \in \mathcal{F}$ and the orthonormal frame W unknown under Gaussian features, two explicit conditional programs J_0, J_1 characterize private verifier samples at sufficiently small fixed accuracy and constant confidence:*

$J_0 \le 0$ *for every rule pair and overlap*	*dimension independent*
$\sup J_0 > 0,\ J_1 \le 0$ *for every rule pair and overlap*	$\Theta(\sqrt{d})$
$\sup J_1 > 0$	$\Theta(d)$.

Upper bounds use one message from a sample-only honest prover. Lower bounds allow arbitrary interaction and improper outputs.

Fixing $\mathcal{F}, k$, noise, and accuracy isolates a structural question: which ambient-dimension costs are possible, and what makes each unavoidable? Theorem 3.1 answers it exhaustively, without smoothness or boundary assumptions on individual rules. The programs optimize improvement while preserving the candidate's conditional error rate (J_0), or both that rate and its transverse first moments (J_1). Their criteria hold uniformly over unknown rules and frames. Even homogeneous halfspaces realize all three regimes as noise increases (Corollary 3.2). The consequences below evaluate this geometry, quantify rank and accuracy dependence, and establish stability under feature uncertainty.

Quantitative rank dependence and sharp separations. For all unknown Boolean combinations of at most K orthogonal homogeneous halfspaces, we prove $m_V \le 2^{CK}\varepsilon^{-1}\log(e/(\varepsilon\delta))$ at $\eta_K = 2^{-3K}/(2^{14}K^4)$, with universal C (Theorem 4.1). Soundness compares against the whole class. Quantitative facet matching controls both frame distance and inequivalent truth tables, giving a bound valid for growing rank: at fixed accuracy and confidence, $K = o(\log d)$ yields $m_V = d^{o(1)}$ with $\eta_K = d^{-o(1)}$. A complementary obstruction shows that dimension-independent verification at every accuracy requires exponential noise decay in rank. For the balanced shifted parity $\operatorname{sign}(s)\operatorname{sign}(u-t)$, choosing $\mathbb{P}(G > t) = \eta/2$ gives $J_1 = J_0 = \eta^2$ and hence $\Theta(d)$ samples at fixed $\varepsilon \le \eta^2/8$ (Theorem 4.3).

Dimension and accuracy can contribute additive costs. For fixed thin slabs and a nonempty range of positive noise allowances, we prove the joint rate $\widetilde{\Theta}(d + 1/\varepsilon)$ at sufficiently small accuracy (Theorem 4.5). Local competitors admit scalar certificates with inverse-linear accuracy cost, while distant first-moment-hidden improvements impose the linear dimension cost. This separates the geometric sources of dimension and accuracy costs: one class can require linear dimension dependence while retaining an additive inverse-accuracy cost.

The certificate survives feature uncertainty. For halfspaces above the Gaussian first-moment signal threshold, we obtain $\widetilde{O}_a(\sqrt{d}/\varepsilon)$ verification under an unknown centered isotropic density e^{-V}/Z with $\nabla^2 V$ in a fixed neighborhood of I (Theorem 5.1). This neighborhood is independent of dimension and accuracy, and contains laws whose total variation distance from Gaussian tends to one. The extension preserves a localized certificate at every accuracy using ordinary samples and no density or conditional oracle.

1.2 Relation to prior work

PAC verification separates learning and certification costs [GRSY21, MS26, BDFP26]. Our contribution is an exact criterion for when each dimension regime occurs, with matching bounds even against unrestricted protocols. The table compares the closest access models and guarantees.

Work and access	Completeness	Soundness or certified goal	Statistical cost
[DMN19, DMN21]: point queries	Smooth low-dimensional Gaussian classes	Membership/tolerant testing; correlation estimation	Dimension-independent queries
[NK26]: verifier samples, SQ algorithm	Underlying algorithm's promise	That algorithm's learning guarantee	$O(\tau^{-2}\log(eB))$ at fixed confidence
[GKSV26]: ordinary samples	Gaussian halfspaces with full Massart noise	Population excess risk; also arbitrary accepted-dataset guarantees	Polynomial in $d, 1/\varepsilon$
This work: independent samples for both parties	Full Massart noise around an unknown Gaussian projection rule	Arbitrary-label population excess risk; unrestricted dishonest prover	Exact $O(1)$, $\Theta(\sqrt{d})$, or $\Theta(d)$ classification

Here B is the SQ query count and τ its tolerance. Our classification fixes rank, intrinsic family, noise, and sufficiently small accuracy. The Boolean and slab theorems additionally quantify rank and joint dimension–accuracy dependence. The noise-testable row fixes a Massart bound below $1/2$. Appendix I gives precise theorem comparisons.

Unknown-rule query testing [DMN19, DMN21], testable learning of Boolean combinations and polynomial thresholds [GKK23, STW24], and Massart-complete arbitrary-label certification [GKSV26] are established. Our advance is the exact conditional classification of ordinary-example verification and the quantitative unknown-composition verifier. Distribution-free realizable learning of Boolean functions of a fixed number of halfspaces admits subexponential time [APS26]. SQ certification retains its underlying algorithm's guarantee [NK26], so a Massart-only learner does not yield arbitrary-label optimality. Generic circuit verification achieves inverse-linear labeled-sample

cost with deterministic labels and a query prover [GJK+24]; noisy query conversions [AGR26] also use different access and noise assumptions.

The proofs build on testing–verification reductions [MS26, BDFP26], mean-norm statistics [BBBY12, Har19, KV18], and Gaussian moment matching [DKS17]. Our technical advance is to prove that the same conditional criteria govern certificates and unrestricted obstructions. Nonnegative scalar dual multipliers cover full ambient Massart noise; lifting preserves the criterion after additional coordinates are revealed; and compression bounds the directions a sample-only prover must reveal. An allocation with positive gain gives matching lower bounds. Beyond classical boundary estimates [Kan11], quantitative facet matching controls every relative frame. Computational biased-halfspace results [DIKL26], multi-index learning [DLB25], relative-error testing [CDH+26], and universal tester-learners [GKSV23] are compared in Appendix I.

2 Model and Preliminaries

For a law P on $(X, Y) \in \mathbb{R}^d \times \{-1, 1\}$, define $R_P(h) = \mathbb{P}_P(h(X) \neq Y)$ and $\mathrm{OPT}_P(\mathcal{H}) = \inf_{h \in \mathcal{H}} R_P(h)$. We omit the subscript when the law is understood. Unless stated otherwise, $0 < \varepsilon < 1$ and $0 < \delta \leq 1/4$. The verifier outputs a predictor or rejects. Its two requirements concern different families of distributions.

Definition 2.1 (Statistical verification)**.** An (ε, δ) verifier is *sound* if, for every law in the soundness family and every prover strategy,

$$\mathbb{P}(\text{output } h \text{ with } R_P(h) > \mathrm{OPT}_P(\mathcal{H}) + \varepsilon) \leq \delta.$$

It is *complete* if, for each law in the completeness family, an honest prover causes a correct nonreject output with probability at least $1 - \delta$. The probabilities include the samples and all protocol randomness. A proper verifier outputs a member of $\mathcal{H}$; an improper verifier may output any measurable Boolean predictor.

Completeness always includes *full Massart noise*: for an unknown target $h_* \in \mathcal{H}$ and a specified $0 \leq \eta < 1/2$,

$$\mathbb{P}(Y \neq h_*(X) \mid X) \leq \eta \quad \text{almost surely.} \tag{2.1}$$

The error probability may depend on every coordinate of X. We also use the signal floor $a = 1 - 2\eta$. Soundness permits every conditional label law under the promised feature marginal or family. It never assumes (2.1). The special case of constant error probability η is random classification noise.

Samples, messages, and computation. Write m_V, m_P for worst-case numbers of independent observations used by the verifier and honest prover. Their samples are independent. All upper bounds use one message, sent before the verifier draws its private sample; a malicious prover may know the whole law. All lower bounds permit arbitrary interaction and improper outputs. Every observed example counts, including observations whose labels are ignored. This measures the statistical power of verification: computation and communication are unrestricted, so the lower bounds are intrinsic sample barriers. The upper bounds concern sample complexity; protocols for general Borel rules may use exact Gaussian integrals and real-valued messages. Reading one sample still requires a d-dimensional vector.

Projection classes and basic identities. Fix $k \geq 1$ and a known Borel $f : \mathbb{R}^k \to \{-1, 1\}$. For $d \geq k$,

$$\mathcal{H}_{f,d} = \{h_W : x \mapsto f(W^\top x) : W \in \mathbb{R}^{d\times k},\ W^\top W = I_k\}.$$

More generally, for a known nonempty family $\mathcal{F}$ of such rules, put $\mathcal{H}_{\mathcal{F},d} = \bigcup_{f\in\mathcal{F}} \mathcal{H}_{f,d}$. The target rule is unknown. We assume $\mathcal{F}$ is totally bounded in Gaussian disagreement: for every $b > 0$, a finite subset $\mathcal{F}[b] \subset \mathcal{F}$ approximates every rule to disagreement at most b. Equivalently, $\mathcal{F}$ is totally bounded in $L^2(N(0, I_k))$. A singleton recovers the fixed-rule model. Sections 3 and 4 assume $X \sim N(0, I_d)$. Section 5 replaces this known marginal by an unknown centered isotropic smooth density. The symbols Φ, φ denote the standard Gaussian distribution function and density, and $\overline{\Phi} = 1 - \Phi$; $\|\cdot\|_F, \|\cdot\|_{\mathrm{op}}$ denote Frobenius and operator norm. With $E_v = \mathbf{1}\{Y \neq h_v(X)\}$ and $D_{v,w} = \mathbf{1}\{h_v(X) \neq h_w(X)\}$, arbitrary labels satisfy

$$R(h_v) - R(h_w) = \mathbb{E}[(2E_v - 1)D_{v,w}]. \tag{2.2}$$

This paired-loss identity underlies both the upper certificates and the improving alternatives in the lower bounds.

3 The Conditional-Moment Classification

Why exactly three regimes? If improvement is exposed by scalar tests on the proposed coordinates, verification needs no dimension-dependent observations. Otherwise, a transverse mean may expose it; testing the norm of an unknown d-dimensional mean with bounded covariance costs $\Theta(\sqrt{d})$ samples at fixed signal. If improving labels can conceal both the scalar information and that mean, random hidden directions force $\Omega(d)$ samples. Ordinary learning under the known Gaussian marginal supplies the matching $O(d)$ ceiling. Thus higher-order concealment cannot create a fourth dimension regime in this model. The programs below determine exactly which information suffices.

Fix a totally bounded intrinsic family $\mathcal{F}$ of rank k and noise bound $0 < \eta < 1/2$ from the model. For an ordered pair $f, g \in \mathcal{F}$, the first rule is the anchor and the second is the competitor. The geometry of two candidate frames is specified by their overlap $A \in \mathbb{R}^{k\times k}$, with $\|A\|_{\mathrm{op}} \leq 1$. For independent $U, G \sim N(0, I_k)$, define

$$B_A = (I_k - A^\top A)^{1/2}, \qquad D_{f,g,A}(u, z) = \mathbf{1}\{g(A^\top u + B_A z) \neq f(u)\}.$$

For any Borel disagreement indicator D on Gaussian space, set

$$\begin{aligned} M_{0,\eta}(D) &= \sup_{0\leq e\leq 1,\ \mathbb{E}e=\eta} \mathbb{E}[eD], \\ M_{1,\eta}(D) &= \sup_{0\leq e\leq 1,\ \mathbb{E}e=\eta,\ \mathbb{E}[Ge]=0} \mathbb{E}[eD], \\ J_j(f, g, A; \eta) &= \mathbb{E}_U \left[2M_{j,\eta}(D_{f,g,A}(U, \cdot)) - \mathbb{E}_G D_{f,g,A}(U, G)\right]. \end{aligned}$$

The paired-risk identity explains these programs: an error rule e gives improvement $2\mathbb{E}[eD] - \mathbb{P}(D)$ to the competitor. The first program holds the error rate fixed conditional on the anchor; the second also cancels its first transverse moments. Thus J_j is a maximal concealed improvement, rather than a statistic of the unknown honest noise function. Always $J_1 \leq J_0$ and $J_j(f, f, I_k; \eta) = 0$. We suppress the rule arguments for a singleton family.

The scalar program has a direct interpretation. If $q_A(u) = \mathbb{P}_G(D_{f,g,A}(u, G) = 1)$, then

$$M_{0,\eta}(D_{f,g,A}(u, \cdot)) = \min\{\eta, q_A(u)\}, \qquad J_0(f, g, A; \eta) = \mathbb{E}_U[2\min\{\eta, q_A(U)\} - q_A(U)].$$

An allocation spends its error budget on disagreement first. The first-moment constraint asks whether it can do so without leaving a detectable transverse mean. Averaging over the anchor coordinates is essential: negative gain on some slices can offset positive gain on others. A global disagreement probability alone loses this information.

The equalities $\mathbb{E}e = \eta$ in the programs describe the alternatives hidden around a constant-noise anchor; they do not restrict honest Massart errors to be constant. For the upper bounds, nonnegative scalar dual multipliers accommodate every pointwise error bound $e_*(X) \le \eta$. For the lower bounds, alternatives need only obey the arbitrary-label soundness promise, and may violate Massart noise. This asymmetry lets the same conditional programs govern both sides of the characterization.

Theorem 3.1 (Sharp dimension classification)**.** *Fix a nonempty totally bounded Borel-rule family $\mathcal{F}$, rank k, noise bound η, and accuracy $0 < \varepsilon < 1$. For $d \ge k$ and $0 < \delta \le 1/4$, the following private verifier sample bounds hold, with constants depending on $\mathcal{F}, k, \eta, \varepsilon$. Conditions range over every ordered $f, g \in \mathcal{F}$ and contraction A:*

condition	*upper bound*
$J_0(f,g,A;\eta) \le 0$ *everywhere*	$O(\log(1/\delta))$
$J_1(f,g,A;\eta) \le 0$ *everywhere*	$O(\sqrt{d}\log(1/\delta))$
no additional condition	$O(d + \log(1/\delta))$.

Every upper protocol is proper and uses one prover message, with full Massart completeness and arbitrary-label Gaussian soundness. An honest independent sample-only prover uses $O(d + \log(1/\delta))$ observations. At constant confidence and sufficiently small fixed accuracy, the exhaustive dimension classification is

$J_0 \le 0$ *everywhere*	*dimension independent*
$\sup_{f,g,A} J_0(f,g,A;\eta) > 0, \quad J_1 \le 0$ *everywhere*	$\Theta(\sqrt{d})$
$\sup_{f,g,A} J_1(f,g,A;\eta) > 0$	$\Theta(d)$.

The lower bounds allow arbitrary interaction and improper outputs; completeness is needed at only one known random-noise anchor.

Upper bounds hold at every fixed accuracy; a positive witness $J_j(f,g,A;\eta) = \Gamma$ gives the lower bound for fixed $\varepsilon \le \Gamma/8$, with an additive validation cost depending on Γ. Parameters are fixed before d grows; constant rules may need no samples. Proper outputs use a fixed finite rule subset. Theorem A.5 proves the result from the fixed-rule foundations in Theorems A.1 and B.1, under the computational conventions of Section 2.

3.1 A worked example: rotating a halfspace

For independent standard normals T, G and $0 < \theta < \pi/2$, compare $\operatorname{sign}(T)$ with $\operatorname{sign}(T\cos\theta + G\sin\theta)$. Disagreement requires $\operatorname{sign}(T)G < -|T|\cot\theta$, giving

$$q_\theta(T) = \overline{\Phi}(|T|\cot\theta), \qquad J_0(\cos\theta;\eta) = \mathbb{E}[2\min\{\eta, q_\theta(T)\} - q_\theta(T)].$$

The dual weight $\alpha_\theta(T) = \mathbf{1}\{|T| < z_\eta \tan\theta\}$, where $z_\eta = \Phi^{-1}(1-\eta)$, makes the scalar certificate an error count in a boundary band. Substituting $|T| = u\tan\theta$ gives

$$\lim_{\theta \downarrow 0} \frac{J_0(\cos\theta;\eta)}{\theta} = 2\varphi(0)\,[2\varphi(z_\eta) - \varphi(0)].$$

The integral identity behind this formula is $\int_0^\infty \min\{\eta, \overline{\Phi}(u)\}\, du = \varphi(z_\eta)$. The slope changes sign at $\eta_c = \overline{\Phi}(\sqrt{2\log 2})$.

For J_1, put $M_\eta(u) = M_{1,\eta}(\mathbf{1}\{G < -u\})$. The corresponding expression replaces $\min\{\eta, q_\theta(T)\}$ by $M_\eta(|T|\cot\theta)$. Error allocated to the negative disagreement tail now needs positive-side compensation to cancel the first moment. This reduces the concealed gain and delays its sign change. Appendix D identifies $\eta_\dagger > \eta_c$ as the unique zero of $\int_0^\infty [2M_\eta(u) - \overline{\Phi}(u)]\, du$.

Corollary 3.2 (Halfspace noise phases)**.** *For homogeneous Gaussian halfspaces with full Massart completeness and arbitrary-label soundness, fix $0 < \eta < 1/2$, constant confidence, and sufficiently small fixed accuracy depending on η. The optimal verifier complexity as $d \to \infty$ is*

noise allowance	*verifier complexity in d*
$0 < \eta \le \eta_c$	*dimension independent*
$\eta_c < \eta \le \eta_\dagger$	$\Theta(\sqrt{d})$
$\eta_\dagger < \eta < 1/2$	$\Theta(d)$,

where $\eta_c = \overline{\Phi}(\sqrt{2\log 2}) \approx 0.119516$ and $\eta_\dagger \in (\eta_c, 2\eta_c)$ is the integral root above. Lower bounds allow arbitrary interaction and improper outputs.

The global bounds are Lemmas E.3 and D.4; positive obstructions feed Theorem B.1. Section A.8 supplies endpoint continuity; Appendix D.5 assembles the proof. Endpoint inclusion here does not extend the sharper inverse-accuracy upper bounds to η_c or $\eta_\dagger$.

3.2 Bounded conditional duals and lifting

Lemma 3.3 (Uniform dual certificate)**.** *For any Borel $D : \mathbb{R}^r \to \{0,1\}$ and standard Gaussian G,*

$$M_{1,\eta}(D) = \min_{\alpha,\gamma}\{\eta\alpha + \mathbb{E}(D - \alpha - \gamma\cdot G)_+\}, \qquad 0 \le \alpha \le 1, \quad \|\gamma\| < 4,$$

where the bounds can be imposed without changing the minimum. Optimal duals are measurable for jointly Borel families of events.

Proof. The feasible box is weak-star compact, and $(\eta, 0)$ is interior to its attainable moment set, giving attainment and strong duality. Complementarity forces error one where $D > \alpha + \gamma\cdot G$. If $\alpha < 0$, this region has mass greater than $1/2$, a contradiction. If $\alpha > 1$, positive error requires $\alpha + \gamma\cdot G \le 1$; integration against the optimizer gives $\alpha\eta \le \eta$. The dual value is at most η and at least $\|\gamma\|/\sqrt{2\pi} - 1$, yielding the norm bound. Compact minimization supplies a measurable selector; the details, including measurable primal near-optimizers, appear in Lemma A.2. □

Let a revealed orthonormal frame P begin with the candidate frame V. Write $U = P^\top X = (T, S)$, with $T = V^\top X$, and $Z = (I - PP^\top)X$. Every competing frame is $PC + NR$, where N is an orthonormal residual frame of rank $r \le k$ and $C^\top C + R^\top R = I_k$. Apply the dual to disagreement between $f(T)$ and $g(C^\top U + R^\top G)$, obtaining $\alpha(U), \gamma(U)$. For $E_V = \mathbf{1}\{Y \ne h_V(X)\}$ define

$$A_{C,R} = \mathbb{E}[\alpha(U)(E_V - \eta)], \qquad M_{C,R} = \mathbb{E}[Z\gamma(U)^\top E_V].$$

Lemma 3.4 (Lifting preserves the geometric criterion)**.** *If every J_1 is nonpositive, then for every labeling and every residual orientation N,*

$$R(h_{f,V}) - R(h_{g,PC+NR}) \le 2A_{C,R} + 2\sqrt{k}\|M_{C,R}\|_F.$$

Proof. Assemble conditional feasible allocations with $\mathbb{E}[e \mid T, S] = \eta$ and $\mathbb{E}[Ge \mid T, S] = 0$. Independence gives $\mathbb{E}[e \mid T] = \eta$ and $\mathbb{E}[(S, G)e \mid T] = 0$. After rotation and independent padding, their gain is bounded by $J_1(f, g, A; \eta)$ at the candidate–competitor overlap. Hence the integrated conditional dual baseline is nonpositive. The remaining dual terms are $2A_{C,R}$ and $2\,\mathrm{tr}(N^\top M_{C,R})$, bounded as stated. This includes singular overlaps; see Lemma A.4 and Theorem A.5. □

Only finitely many rules and coefficient pairs are tested. Total boundedness gives finitely many Lipschitz approximants with a common Lipschitz bound L: for each $f \in \mathcal{F}$ choose q with $\mathbb{E}|f - q|^2 \le b$. Then $4\mathbb{P}(h_{f,W} \ne h_{f,W'}) \le 6b + 3L^2\|W - W'\|_F^2$, uniformly over rules and ambient dimension. A finite rule net and a finite frame net therefore control every comparator's risk. Rounding predictions avoids any continuity assumption on the dual coefficients (Lemma A.6).

3.3 Certificate construction and matching obstructions

Lemma 3.5 (Testing and compressing residual moments)**.** *For a fixed revealed span and coefficient pair, vectorize $W = Z\gamma(U)^\top E_V$. Its covariance is at most $16I$ under arbitrary labels. A tolerant norm test distinguishes $\|\mathbb{E}W\| \le t/4$ from $\|\mathbb{E}W\| > t$ using $O(\sqrt{kd}/t^2)$ observations at constant error. An honest prover can reveal at most $O_k(t^{-2})$ additional directions so that all final coefficient-net matrices have norm at most $t/4$.*

Proof. For every unit Frobenius matrix Q, discard $E_V \le 1$ and use independence of U, Z to obtain $\mathbb{E}\langle Q, W\rangle^2 \le \mathbb{E}\|Q\gamma(U)\|^2 \le 16$. For $\mu = \mathbb{E}W$, $\Sigma = \mathrm{Cov}(W)$, the off-diagonal statistic obeys

$$U_n = \frac{\sum_{i \ne j}\langle W_i, W_j\rangle}{n(n-1)}, \qquad \mathbb{E}U_n = \|\mu\|^2,$$
$$\mathrm{Var}(U_n) = \frac{4\mu^\top\Sigma\mu}{n} + \frac{2\,\mathrm{tr}(\Sigma^2)}{n(n-1)}.$$

Chebyshev proves the norm test; independent blocks amplify confidence. For compression, estimate the residual matrices for all comparator rule-net points on fresh prover batches. A large column supplies a new direction n_j and witness $F_j = \gamma_{j,\ell}(U_j)(n_j \cdot X)$, correlated by $\Omega_k(t)$ with $e_V(X) = \mathbb{P}(Y \ne h_V(X) \mid X)$. Each F_j has squared norm at most 16 and conditional mean zero given the preceding span. Earlier witnesses are measurable in that span, so they are orthogonal. Bessel gives

$$\sum_j \frac{|\mathbb{E}[e_V F_j]|^2}{\mathbb{E}F_j^2} \le \mathbb{E}e_V^2 \le 1,$$

bounding the additions even when the comparator rule changes at every stage. Fresh-batch concentration and a deterministic stage cap give the honest budget in Theorem A.5. □

The honest prover first learns a candidate of small disagreement with the target. Then $e_V \le \eta + \mathbf{1}\{h_{f,V} \ne h_*\}$, so all nonnegative scalar duals pass. Taking $t = \Theta_k(\varepsilon)$, the compressed coefficient net and Lemma 3.4 give the square-root upper bound. The J_0 regime uses only scalar duals; ordinary finite-net risk comparison gives the unconditional linear upper bound. The final message precedes all private verification.

Matching lower bounds. Suppose $J_j(f, g, A; \eta) = \Gamma > 0$. A measurable conditional allocation $e(U, G)$ with the required moments achieves gain at least $g_0 = \Gamma/2$. Against the fixed null regression $(1 - 2\eta)f(U)$, hide an orthonormal frame V in the remaining $d - k$ coordinates Z, and use regression $f(U)(1 - 2e(U, V^\top Z))$. The classifier with intrinsic rule g and overlap A improves anchor risk by

at least g_0. For two hidden frames, normalized conditional Hermite expansion gives the likelihood kernel bound

$$|K(V,W)-1| \le \|V^\top W\|_{\mathrm{op}}^{j+1}.$$

Random-frame overlap moments make the mixtures indistinguishable below $c\sqrt{d}$ samples for $j=0$ and cd for $j=1$ (Lemmas B.3–B.4). Simulating the null's honest prover converts any interactive verifier into a test. Fresh relative-risk validation handles improper outputs because the anchor is Bayes under the null. Its cost is $O(\Gamma^{-2})$; for $\varepsilon \le \Gamma/8$, Theorem A.5 transfers Theorem B.1 to cross-rule witnesses with the same constants.

4 Quantitative Families and Geometric Separations

For $K \ge 1$, let $\mathcal{C}_{K,d}$ contain all predictors $b(\mathrm{sign}(W^\top x))$ with an arbitrary Boolean rule b and an orthonormal frame of rank at most $\min(K,d)$, including constants. The rule, rank, and directions are unknown. This class includes all orthogonal parities, majorities, and intersections of homogeneous halfspaces. We quantify both the verification saving and the noise cost of allowing every Boolean rule.

Theorem 4.1 (Quantitative verification of unknown Boolean compositions)**.** *There is a universal $C>0$ such that, for every $K,d \ge 1$, $0<\varepsilon<1$, and $0<\delta\le 1/4$, the class $\mathcal{C}_{K,d}$ has a proper one-message verifier at noise allowance $\eta_K = 2^{-3K}/(2^{14}K^4)$ using*

$$m_V \le \frac{2^{CK}}{\varepsilon}\log\frac{e}{\varepsilon\delta}, \qquad m_P \le \frac{2^{CK}}{\varepsilon^2}\left(d\log\frac{e}{\varepsilon}+\log\frac{1}{\delta}\right).$$

Completeness allows full ambient Massart noise bounded by η_K; soundness allows arbitrary Gaussian-marginal labels and compares against the entire class. The message is a truth table and a frame.

The explicit rank dependence permits growing rank: at fixed accuracy and confidence, $K = o(\log d)$ gives $m_V = d^{o(1)}$, $m_P = d^{1+o(1)}$, and $\eta_K = d^{-o(1)}$. Rare Boolean regions force a rank-dependent noise scale. A single-orthant rule, of minority probability $p = 2^{-K}$, has

$$J_0(0;\eta) = 2[(1-p)\min\{\eta,p\} + p\eta - p(1-p)] > 0 \quad\Longleftrightarrow\quad \eta > p(1-p).$$

Thus any allowance giving dimension-independent verification of the whole family at every accuracy is at most $2^{-K}(1-2^{-K})$. For $K \ge 2$, the centrally symmetric union of two opposite orthants has $J_1 = J_0$; it forces linear dimension complexity once $\eta > 2^{1-K}(1-2^{1-K})$ (Corollary C.9). Thus exponential noise decay reflects a statistical obstruction, with different upper and lower exponents. These lower bounds fix rank and require sufficiently small accuracy; the growing-rank statement is an upper bound.

The $1/\varepsilon$ dependence is optimal up to logarithms: already noiseless halfspaces in $d \ge 2$ require $\Omega(\varepsilon^{-1}\log(1/\delta))$ samples, even with interaction and improper outputs (Proposition C.10).

4.1 Explicit geometry at every overlap

For an essential rank-k rule b, let $\mathcal{S}_b$ be its signed permutation symmetries and put $r = \min_{Q\in\mathcal{S}_b}\|W - VQ\|_F$. For $k \le K$, define $c_K = 2^{-3K}/(64K^{5/2})$ and $\gamma_K = 2^{-3K}/(32K^2)$.

Lemma 4.2 (Quantitative identifiability and allocation margin)**.** *Every same-rule frame pair satisfies*

$$c_K r \le \mathbb{P}(h_V \ne h_W) \le \tfrac{1}{2}\sqrt{k}\,r, \qquad J_0(V,W;\eta_K) \le -c_K r/2.$$

Predictors from different essential-rule orbits disagree with probability at least γ_K, uniformly in dimension.

The key estimate detects each unmatched normal through a pivotal facet. Fix an anchor normal v_i and put $s_i = \min_j \sqrt{1-(v_i \cdot w_j)^2}$. Its pivotal orthant has Gaussian mass $2^{-(k-1)}$. Remove narrow bands around the competing hyperplanes; a positive fraction remains. At each remaining point, a short segment crossing the anchor facet changes the anchor's sign while the competitor stays constant. Integrating paired points gives

$$\mathbb{P}(h_{b,V} \neq h_{c,W}) \geq \frac{4^{-k}}{8\ell} s_i,$$

where ℓ is the competitor's rank. Applying this in both directions forces a bijective matching of normals when disagreement is below γ_K. Rounding the matched signs changes less than one orthant's mass, so the Boolean tables coincide. This proves both the metric and the cross-rule gap (Lemma C.2–Corollary C.5).

For $r \leq 1/4$, take the anchor-measurable boundary band $B = \{\min_i |v_i \cdot X| \leq Lr\}$ with $L = \sqrt{8\log(32K/c_K)} \leq 9\sqrt{K}$. Gaussian tails give $\mathbb{P}(B) \leq 2K\varphi(0)Lr$ and $\mathbb{P}(D \setminus B) \leq c_K r/16$. Hence $J_0 \leq 2\eta_K \mathbb{P}(B) + 2\mathbb{P}(D \setminus B) - \mathbb{P}(D) \leq -c_K r/2$. Larger distances use $J_0 \leq 2\eta_K - \mathbb{P}(D)$. The scalar dual $\alpha = \mathbf{1}\{\mathbb{P}(D \mid V^\top X) > \eta_K\}$ has variance at most $\sqrt{K}r/\eta_K$.

At each dyadic distance scale, coefficient nets need only $\exp(O(K^3))$ points, independent of ambient dimension. The negative margin pays for rounding; Bernstein estimation costs $2^{O(K)}/r$ observations. Summing scales down to $r = \Theta(\varepsilon/\sqrt{K})$ yields Theorem 4.1. A single risk test handles every different-rule orbit because its gain is at most $2R(h_V) - \gamma_K$. The honest learner searches finite rule/frame nets: although there are doubly exponentially many truth tables, their logarithmic count is $O(2^K)$, preserving the single exponential sample factor. Theorem C.7 gives the complete protocol and accounting; no runtime guarantee is asserted.

4.2 Moving one threshold changes the dimension exponent

Consider the two projection rules

$$f_0(s,u) = \operatorname{sign}(s)\operatorname{sign}(u), \qquad f_t(s,u) = \operatorname{sign}(s)\operatorname{sign}(u-t).$$

Each is parity of two orthogonal halfspaces. For $d \geq 2$, write $\mathcal{Q}_{t,d} = \mathcal{H}_{f_t,d}$.

Theorem 4.3 (Matched balanced quadratic separation)**.** *There is $\eta_* > 0$ such that, for every fixed $0 < \eta \leq \eta_*$, choosing $t = \Phi^{-1}(1-\eta/2)$ gives, at constant confidence and every fixed $0 < \varepsilon \leq \eta^2/8$,*

class	*private verifier samples*
$\mathcal{Q}_{0,d}$	$O(\varepsilon^{-1}\log(e/\varepsilon))$
$\mathcal{Q}_{t,d}$	$\Theta_{\eta,\varepsilon}(d)$.

Both classes have Gaussian positive prevalence $1/2$, minimum projection rank two, minimum polynomial threshold degree two, and VC dimension $\Theta(d)$. Both guarantees use the same feature law and noise bound, with full Massart completeness and arbitrary-label soundness. The lower bound allows interaction and improper outputs.

The first row applies the within-rule protocol to two-bit parity (Theorem C.1), so its output remains in $\mathcal{Q}_{0,d}$. For the second, use overlap $A = \operatorname{diag}(1,0)$: the gate is shared and the other coordinate is replaced by an independent $G \sim N(0,1)$. Put $p = \mathbb{P}(G > t) = \eta/2$. When the anchor

has $u \le t$, disagreement is $G > t$; the allocation $e = \mathbf{1}\{|G| > t\}$ has mass η, zero first moment, and captures all disagreement. When $u > t$, disagreement is $G \le t$; use $e = \eta\mathbf{1}\{|G| \le t\}/(1-2p)$, again of mass η and zero first moment. Both attain the mass-only optimum, giving exactly

$$J_1(A;\eta) = J_0(A;\eta) = 2\{(1-p)p + p\eta\} - 2p(1-p) = \eta^2.$$

Theorem 3.1 and ordinary learning give matching linear bounds. The two active boundary hyperplanes force rank and threshold degree two; Lemma C.12 proves the VC claim. Even the observed labels in the hard null and alternatives are balanced. A single offset therefore changes the dimension exponent while preserving these measures of class complexity.

Corollary 4.4 (Allowing one offset)**.** *Let $\mathcal{Q}_d = \bigcup_{t\in\mathbb{R}} \mathcal{Q}_{t,d}$. For every fixed $0 < \eta < 1/2$ and $0 < \varepsilon \le \eta^2/8$, this balanced quadratic class has private verifier complexity $\Theta_{\eta,\varepsilon}(d)$ at constant confidence. Its homogeneous subclass admits dimension-independent verification at a positive noise allowance.*

The same allocation remains feasible for every $\eta < 1/2$. The lower bound uses a fixed-offset null and fixed-offset improving comparators, all contained in $\mathcal{Q}_d$; soundness against the larger class preserves the obstruction. The entire offset class still has VC dimension $O(d)$, giving the matching upper bound (Corollary C.13).

4.3 Dimension and accuracy costs can add

Centered slabs resolve a joint dimension–accuracy question beyond the fixed-accuracy classification: local and distant competitors contribute additive costs. Put $\mathcal{S}_{t,d} = \{x \mapsto 2\mathbf{1}\{|w\cdot x| \le t\} - 1 : \|w\| = 1\}$ and $p = 2\Phi(t) - 1$.

Theorem 4.5 (Additive dimension–accuracy complexity)**.** *Fix $t > 0$ and $p(1-p) < \eta < \eta_c$. There is $\varepsilon_{t,\eta} > 0$ such that, for all $d \ge 2$, $0 < \varepsilon \le \varepsilon_{t,\eta}$ and $0 < \delta \le 1/4$,*

$$\Omega\left(d + \frac{1}{\varepsilon}\log\frac{1}{\delta}\right) \le m_V^*(\mathcal{S}_{t,d}) \le O_{t,\eta}\left(d + \frac{1}{\varepsilon}\log\frac{e}{\varepsilon\delta}\right).$$

The upper verifier is proper and uses one message from an independent prover with $O_{t,\eta}((d\log(e/\varepsilon) + \log(1/\delta))/\varepsilon^2)$ samples. Completeness is full Massart and soundness permits arbitrary labels. The lower bound allows interaction and improper outputs.

Here m_V^* is the minimum worst-case private sample count in this model. The parameter range is nonempty: for any $\eta < \eta_c$ choose $p = \eta/2$. Width and noise are fixed while both d and ε vary. The new feature is the additive cost $d + 1/\varepsilon$.

For slabs at angle θ, direct Gaussian integration gives

$$\lim_{\theta\downarrow 0} \frac{J_0(\theta;\eta)}{\theta} = 4\varphi(t)[2\varphi(\Phi^{-1}(1-\eta)) - \varphi(0)] < 0.$$

Nearby competitors therefore have scalar certificates with a negative linear margin and variance $O_{t,\eta}(\theta)$, costing $\widetilde{O}_{t,\eta}(1/\varepsilon)$. Competitors outside this fixed angular neighborhood have disagreement bounded below. A fixed-resolution ambient net and paired-risk tests with a negative acceptance threshold handle them using $O_{t,\eta}(d + \log(1/\delta))$ samples; Massart completeness supplies that strict negative margin. Orthogonal slabs have $J_1(0) > 0$, forcing $\Omega(d)$, and two nearby noiseless targets force $\Omega(\varepsilon^{-1}\log(1/\delta))$. These bounds combine by taking their maximum. Appendix F.3 proves the local limit, the finite protocol, and both unrestricted lower bounds.

5 Verification under Unknown Smooth Feature Laws

The conditional certificate tolerates feature uncertainty at every accuracy: the permitted curvature perturbation stays fixed as dimension grows and accuracy improves. Its effect scales with boundary width. Use the halfspace threshold $\eta_\dagger$ from Section 3.1, and put $a_\dagger = 1 - 2\eta_\dagger$, $a_c = 1 - 2\eta_c$. Let $\mathcal{D}_\tau$ be the centered isotropic laws with unknown density proportional to $e^{-V(x)}$, where $V \in C^2$ and

$$(1-\tau)I_d \preceq \nabla^2 V(x) \preceq (1+\tau)I_d \quad \text{for every } x \in \mathbb{R}^d. \tag{5.1}$$

Both completeness and soundness promise a marginal in $\mathcal{D}_\tau$.

Theorem 5.1 (Verification under curvature uncertainty)**.** *For each fixed $a \in (a_\dagger, 1)$, there is $\tau_a > 0$, independent of dimension and accuracy, such that homogeneous halfspaces admit a proper one-message verifier over $\mathcal{D}_{\tau_a}$ with full Massart completeness of signal floor a and arbitrary-label soundness, using*

$$m_V = O_a\left(\frac{\sqrt{d}}{\varepsilon}\log\frac{e}{\varepsilon\delta}\right), \qquad m_P = O_a\left(\frac{d\log(e/\varepsilon) + \log(e/(\varepsilon\delta))}{\varepsilon^2}\right).$$

The honest prover uses independent samples. Its message contains a candidate and $O(\log(e/\varepsilon))$ frames of dimension at most K_a, a constant depending only on a.

The verifier uses only samples, without a density, score, or conditional-mean oracle. Curvature and isotropy are promises; the signal range excludes the critical endpoint.

Lemma 5.2 (Projected densities)**.** *For $\tau \le 1/16$, any orthonormal $d \times q$ matrix P, $q \le 3$, has projected density p_P satisfying*

$$p_P(z) \le Ce^{-c\|z\|^2}, \qquad |p_P(z) - \phi_q(z)| \le C\tau(1+\|z\|^2)e^{-c\|z\|^2},$$

where ϕ_q is the standard Gaussian density and $C, c > 0$ are universal.

Writing $X = (U, Z)$ and $p_P = e^{-W}$ gives $\nabla^2 W = \mathbb{E}[\nabla^2_{UU} V \mid U] - \mathrm{Cov}(\nabla_U V \mid U)$. Conditional Brascamp–Lieb bounds the covariance term by $\tau^2/(1-\tau)$ [CCEL13, Equation (1.3)]. Thus $\|\nabla^2 W - I_q\|_{\mathrm{op}} \le 2\tau$; centering and isotropy bound $\|\nabla W(0)\| \le 2\tau\sqrt{q}$. Integration and normalization prove the lemma. Only projections of dimension at most three are needed.

5.1 A certificate stable at every angular scale

Put $\eta = (1-a)/2$. The Gaussian dual supplies $0 \le \alpha \le 1$, $|\gamma| \le B = (2\log 2)^{-1/2}$ with $\{\alpha \ne 0\} \subset \{u < k_\eta\}$ and $\{\gamma \ne 0\} \subset \{|u| < k_\eta\}$, where $k_\eta = \Phi^{-1}(1-\eta/2)$ (Theorem D.5). For a candidate h_v, let $P^\top X = (T, S)$, $T = v \cdot X$, and $Z = (I - PP^\top)X$. A competitor is $h_w = \mathrm{sign}(cT + \beta \cdot S + \sigma N)$, with $N = n \cdot X$, $n \perp P$, $c^2 + \|\beta\|^2 + \sigma^2 = 1$, and $\sigma \ge 0$. Its angle is $\theta = \arccos c$. For $\sigma > 0$ evaluate the dual at $u = \mathrm{sign}(T)(cT + \beta \cdot S)/\sigma$ and define observable means

$$A_\lambda = \mathbb{E}[\alpha(E_v - \eta)], \qquad V_\lambda = \mathbb{E}[\gamma\,\mathrm{sign}(T) Z E_v], \quad E_v = \mathbf{1}\{Y \ne h_v(X)\}.$$

For $\sigma = 0$ use $\alpha = \mathbf{1}\{h_v \ne h_w\}$ and $\gamma = 0$.

Lemma 5.3 (Localized stability)**.** *There are $g_a, C_a > 0$, independent of frame dimension, such that all labels and competitors satisfy*

$$R(h_v) - R(h_w) \le -(g_a - C_a\tau)\theta + 2A_\lambda + 2\|V_\lambda\|.$$

The pointwise dual baseline is $H_\lambda = 2(D - \alpha - \gamma \operatorname{sign}(T)N)_+ + 2\eta\alpha - D$, where $D = \mathbf{1}\{h_v \neq h_w\}$. Its Gaussian mean is at most $-g_a\theta$. It depends only on the orthogonal coordinates $T, Q = (\beta/\|\beta\|) \cdot S, N$, omitting Q when $\beta = 0$. For $\theta \leq 1/4$, the dual support and disagreement wedge imply

$$|H_\lambda| \leq C_a(1 + |N|)\mathbf{1}\{|T| \leq C_a\theta(1 + |Q| + |N|)\}.$$

Integrating the projected-density error first over this T interval gives $|\mathbb{E}H_\lambda - \mathbb{E}_G H_\lambda| \leq C_a\tau\theta$. For larger angles the global $O_a(\tau)$ bound is already $O_a(\tau\theta)$; $\sigma = 0$ uses $H_\lambda = -aD$ and the same wedge. Taking $\tau \leq g_a/(2C_a)$ preserves the margin at every accuracy. A fixed additive perturbation would not suffice.

5.2 Compression without a conditional oracle

On an annulus $r \leq \theta \leq 2r$, three-dimensional density estimates give $\operatorname{Cov}(\gamma \operatorname{sign}(T)ZE_v) \preceq L_a r I_d$ and scalar variance at most $L_a r$. A witness $F = \gamma \operatorname{sign}(T)n \cdot X$ has $\mathbb{E}F^2 \leq L_a r$ and $\mathbb{E}[|F|\mathbf{1}\{|T| > Mr\}] \leq r\omega_a(M)$, with $\omega_a(M) \to 0$. For the current k-dimensional revealed coordinates U, introduce only in the proof $m(U) = \mathbb{E}[Z \mid U]$. Conditional differentiation and Poincare give $\|Dm\|_{\mathrm{op}} \leq \tau/(1 - \tau)$ and $\|m(U)\| \leq \tau(\|U\| + \sqrt{k})/(1 - \tau)$. Localization yields

$$\mathbb{E}[|\gamma|\|m(U)\|] \leq C'_a\tau\sqrt{k}\, r. \tag{5.2}$$

The prover appends directions with raw mean correlation exceeding λr, for fixed $\lambda = \lambda_a > 0$. For successive witnesses put $W_j = F_j - \mathbb{E}[F_j \mid U_j]$. These proof-only innovations are orthogonal, have squared norm at most $L_a r$, and tails at most $2r\omega_a(M)$. Choose M with $2\omega_a(M) \leq \lambda/4$, then $K > 4L_a C_T M/\lambda^2$, where $\mathbb{P}(|T| \leq Mr) \leq C_T Mr$. Finally choose τ_a with $C'_a\tau_a\sqrt{K} \leq \lambda/4$. For the first K additions, the drift and tail bounds give correlation above $\lambda r/2$ with $e_v\mathbf{1}\{|T| \leq Mr\}$, where $e_v(X) = \mathbb{P}(Y \neq h_v(X) \mid X)$. Bessel would imply

$$K\frac{\lambda^2 r}{4L_a} < \mathbb{E}[e_v^2\mathbf{1}\{|T| \leq Mr\}] \leq C_T Mr,$$

a contradiction. Choosing K before τ_a bounds the revealed span without evaluating conditional means.

Over finite coefficient nets, scalar Bernstein tests and mean-norm tests at tolerance $\Theta_a(r)$ cost respectively $O_a(1/r)$ and $O_a(\sqrt{d}/r)$ observations per confidence block. Summing scales to $r = \Theta_a(\varepsilon)$ proves the verifier bound. A proper learner and fresh estimates over bounded compression stages give the prover bound (Theorem G.1). For $a_\dagger < a < a_c$, the Gaussian subfamily gives an $\Omega_a(\sqrt{d}/\varepsilon)$ lower bound at sufficiently small accuracy and sufficiently large dimension (Theorem H.4); non-Gaussian product laws in the family have total variation distance tending to one from Gaussian (Lemma G.7).

6 Conclusion

Conditional allocation geometry exactly classifies the dimension cost of Gaussian projection verification through matching certificates and unrestricted lower bounds. Halfspace noise phases and balanced quadratic separations demonstrate its explanatory power. Quantitative rank bounds, additive slab costs, and smooth-law stability extend its reach. The governing statistical obstacle is the conditional concealment of improvement, captured by the same programs that yield the certificates.

A Classification for Gaussian projection rules

We first establish the fixed-rule foundations, then prove the unknown-family classification in Theorem A.5. Two conditional programs describe whether an improvement over a proposed classifier can be hidden while preserving its conditional error rate, or both that rate and its first transverse moments. Their signs give an exhaustive dimension classification.

The result is statistical. The rule is a fixed, known Borel function; the protocol may depend on its exact Gaussian integrals. Constants may depend on the rule, its projection rank, the noise bound, and accuracy. We assert neither computational efficiency nor effective integration for an arbitrary Borel description. Gaussian approximation in L^2 supplies the required finite prediction nets without imposing smoothness on the rule or on the actual Massart error function.

A.1 Model and the two explicit conditional programs

Fix $k \geq 1$, a Borel $f : \mathbb{R}^k \to \{-1, 1\}$, and $0 < \eta < 1/2$. For $d \geq k$, write

$$\mathcal{H}_{f,d} = \{h_W(x) = f(W^\top x) : W \in \mathbb{R}^{d\times k},\ W^\top W = I_k\}.$$

The marginal is $X \sim N(0, I_d)$ and $Y \in \{-1, 1\}$. Risk is $R(h) = \mathbb{P}(h(X) \neq Y)$ and OPT is its infimum over $\mathcal{H}_{f,d}$. Soundness requires that, under every conditional law of Y given X and every prover strategy, the probability of outputting a classifier of risk greater than $\mathrm{OPT} + \varepsilon$ is at most δ; rejection is allowed. Completeness is required for every law satisfying

$$\mathbb{P}(Y \neq h_{W_*}(X) \mid X) \leq \eta \quad \text{almost surely}$$

for some $W_*^\top W_* = I_k$: an honest prover must cause a correct nonreject output with probability at least $1 - \delta$. That prover has its own labeled observations, independent of the verifier's. The upper bounds use one message sent before the verifier's independent observations are drawn.

For a matrix $A \in \mathbb{R}^{k\times k}$ with $\|A\|_{\mathrm{op}} \leq 1$, put $B_A = (I_k - A^\top A)^{1/2}$ and, for independent standard Gaussian $U, G \in \mathbb{R}^k$, define

$$D_A(u, g) = \mathbf{1}\{f(A^\top u + B_A g) \neq f(u)\}.$$

For an arbitrary Borel $D : \mathbb{R}^r \to \{0, 1\}$, define

$$M_{0,\eta}(D) = \sup\{\mathbb{E}[e(G)D(G)] : 0 \leq e \leq 1,\ \mathbb{E}e = \eta\},$$
$$M_{1,\eta}(D) = \sup\{\mathbb{E}[e(G)D(G)] : 0 \leq e \leq 1,\ \mathbb{E}e = \eta,\ \mathbb{E}[Ge] = 0\}.$$

Here $G \sim N(0, I_r)$; the $r = 0$ convention is $M_{0,\eta}(D) = M_{1,\eta}(D) = \eta D$. Set

$$J_j(A; \eta) = \mathbb{E}_U\left[2M_{j,\eta}(D_A(U, \cdot)) - \mathbb{E}_G D_A(U, G)\right], \qquad j \in \{0, 1\}.$$

These are fixed-dimensional functionals of a known rule and its overlap geometry. They do not involve the unknown label law.

Theorem A.1 (Conditional-moment classification)**.** *For every fixed f, k, η, every fixed $0 < \varepsilon < 1$, and $0 < \delta \leq 1/4$, there are proper one-message verifiers with full Massart completeness and arbitrary-label Gaussian soundness satisfying the following.*

1. *If $J_0(A; \eta) \leq 0$ for every contraction A, the private sample count is $O_{f,k,\eta,\varepsilon}(\log(1/\delta))$.*
2. *If $J_1(A; \eta) \leq 0$ for every contraction A, the private sample count is $O_{f,k,\eta,\varepsilon}(\sqrt{d}\log(1/\delta))$.*

3. Without either condition, the private sample count is $O_{f,k,\eta,\varepsilon}(d+\log(1/\delta))$.

In all three cases an honest independent sample-only prover needs $O_{f,k,\eta,\varepsilon}(d+\log(1/\delta))$ observations. The second protocol's message contains at most $k+O_k(\varepsilon^{-2})$ orthonormal directions.

Together with Theorem B.1, these bounds give the following exhaustive classification, at sufficiently small fixed accuracy in each row and fixed success probability:

$$\begin{array}{c|c} J_0(A;\eta)\le 0 \text{ for every } A & O(1) \\ \sup_A J_0(A;\eta)>0,\quad J_1(A;\eta)\le 0 \text{ for every } A & \Theta(\sqrt{d}) \\ \sup_A J_1(A;\eta)>0 & \Theta(d). \end{array}$$

"Dimension independent" in the first row includes degenerate classes requiring zero observations; no universal positive lower bound is asserted for a constant rule.

The lower rows mean that there is an $\varepsilon_0>0$, depending on a positive envelope witness, such that the lower bound holds for every fixed $0<\varepsilon\le\varepsilon_0$ as d tends to infinity. They are not claims of a lower bound uniform in accuracy. Theorem B.1 gives explicit positive-gain constants and allows arbitrary interaction and improper outputs. We prove every upper bound below.

A.2 Uniform bounded duals for arbitrary disagreement sets

Lemma A.2 (Attainment, bounded duals, and measurability)**.** *For $r\ge 1$ and any Borel $D:\mathbb{R}^r\to\{0,1\}$, the first-moment program is attained and has the dual representation*

$$M_{1,\eta}(D)=\min_{\alpha\in\mathbb{R},\,\gamma\in\mathbb{R}^r}\left\{\eta\alpha+\mathbb{E}(D(G)-\alpha-\gamma\cdot G)_+\right\}.$$

Every dual minimizer satisfies

$$0\le\alpha\le 1,\qquad \|\gamma\|\le\sqrt{2\pi}(1+\eta)<4.$$

In particular one may restrict minimization to the fixed compact set $K_r=[0,1]\times\{\gamma:\|\gamma\|\le 4\}$. For a jointly Borel family $D_\theta(g)$ on a standard Borel parameter space, its value is Borel, and dual minimizers can be chosen jointly Borel in θ. For every $\zeta>0$ there is also a jointly Borel choice of primal feasible $e_\theta(g)$ with objective at least $M_{1,\eta}(D_\theta)-\zeta$.

The corresponding mass-only formula is $M_{0,\eta}(D)=\min\{\eta,\mathbb{E}D\}$, with a Borel dual $\alpha\in\{0,1\}$ and no γ.

Proof. The feasible set in $L^\infty(N(0,I_r))$ is weak-star compact: the box $0\le e\le 1$ is weak-star closed, and the finitely many constraints pair against integrable functions $1,G_1,\dots,G_r$. The objective also pairs against an integrable function. Hence a primal optimum exists.

For completeness, strong duality can be obtained by finite-dimensional supporting hyperplanes. Consider the attainable moment vectors $(\mathbb{E}e,\mathbb{E}[Ge])$ for $0\le e\le 1$. The point $(\eta,0)$ is an interior point of this set: the perturbations

$$e(g)=\eta+c_0+\sum_{i=1}^{r}c_i\tanh(g_i)$$

remain in $[0,1]$ for sufficiently small $\sum_i|c_i|+|c_0|$, and their moments vary as $(\eta+c_0,c_1b,\dots,c_rb)$, where $b=\mathbb{E}[G_1\tanh(G_1)]>0$. The maximum objective at each attainable moment vector is a finite concave function. At the stated interior point it has a supporting affine function. Its coefficients (α,γ) give, by maximizing pointwise over the box, exactly the displayed dual value. This proves

equality and dual attainment. Equivalently, the dual objective is continuous and coercive: it is at least

$$\eta\alpha + \mathbb{E}(-\alpha - \gamma \cdot G)_+,$$

a positively homogeneous function strictly positive on every nonzero (α, γ). Its minimum on the unit sphere is positive for each fixed r, η.

Let e_* be a primal optimizer and (α, γ) a dual optimizer. Equality in the pointwise inequality

$$eD \le (D - \alpha - \gamma \cdot G)_+ + \alpha e + \gamma \cdot Ge$$

forces $e_* = 1$ where $D > \alpha + \gamma \cdot G$, and $e_* = 0$ where $D < \alpha + \gamma \cdot G$. If $\alpha < 0$, the first of these regions contains $\{\alpha + \gamma \cdot G < 0\}$, of Gaussian mass greater than $1/2$ (or mass one if $\gamma = 0$). This contradicts $\mathbb{E}e_* = \eta < 1/2$. If $\alpha > 1$, on the set where $e_* > 0$ one has $\alpha + \gamma \cdot G \le D \le 1$. Integration gives $\alpha\eta = \mathbb{E}[(\alpha + \gamma \cdot G)e_*] \le \eta$, again a contradiction. Finally, the dual value is at most η by choosing $(1, 0)$. Since $D \ge 0$ and $\alpha \le 1$,

$$\eta \ge \eta\alpha + \mathbb{E}(D - \alpha - \gamma \cdot G)_+ \ge \mathbb{E}(-1 - \gamma \cdot G)_+ \ge \frac{\|\gamma\|}{\sqrt{2\pi}} - 1.$$

This proves the uniform bound.

The parameterized dual objective is jointly Borel and continuous in its compact dual variable: measurability follows from integration of a jointly Borel integrand, and continuity from the integrable Lipschitz bound $|\Delta\alpha| + \|\Delta\gamma\|\|G\|$. Its infimum on any fixed compact subset is Borel, by taking the infimum on a fixed countable dense subset. A measurable minimizer can be constructed without an uncountable choice: successively choose the first closed ball of radius 2^{-n} from a fixed finite cover of K_r whose intersection with the previously selected balls contains a minimizer. The test for each such intersection is equality of its minimum with the unrestricted minimum, hence is Borel; there are only countably many finite histories. The nested nonempty compact intersections have diameter tending to zero. Their unique limit is a Borel minimizer.

For the primal selection, take a countable L^2-dense sequence in the fixed feasible set, assigning a Borel representative to each member. This exists because Gaussian L^2 is separable. The supremum of its pairings with D_θ equals the optimum, by L^2 continuity of the pairing. Select the first sequence element whose pairing is at least the optimum minus ζ. This is jointly Borel and retains all moment constraints exactly. The same argument works for the mass-only program; its explicit value and dual follow by taking $\alpha = 1$ when $\mathbb{E}D \ge \eta$ and $\alpha = 0$ otherwise. □

Why the sign of the scalar dual matters. The nonnegativity of α is essential for full Massart completeness: a true error probability below η must not violate a scalar certificate. A bound on the absolute values alone would not establish the protocol. The preceding lemma proves the needed sign for every Borel disagreement set, using $\eta < 1/2$.

A.3 Prediction nets for an arbitrary fixed Borel rule

Lemma A.3 (Dimension-independent disagreement modulus)**.** *There is a nondecreasing function $\omega_f : [0, \infty) \to [0, 1]$ with $\omega_f(s) \to 0$ as $s \downarrow 0$ such that, in every ambient dimension, orthonormal k-frames W, W' satisfy*

$$\mathbb{P}\{f(W^\top X) \ne f(W'^\top X)\} \le \omega_f(\|W - W'\|_F).$$

Consequently, a Euclidean ρ-net of any compact set of orthonormal coefficient frames is a prediction-disagreement $\omega_f(\rho)$-net. No continuity of dual weights is required.

Proof. Given $b > 0$, choose a bounded Lipschitz function $q : \mathbb{R}^k \to \mathbb{R}$ with $\mathbb{E}|f(G) - q(G)|^2 \le b$; bounded Lipschitz functions are dense in Gaussian L^2, for example by truncation to a large box followed by continuous approximation and smoothing. Both projected feature vectors are standard Gaussian. If L_q is a Lipschitz constant, the inequality $(a+b+c)^2 \le 3(a^2+b^2+c^2)$ gives

$$4\mathbb{P}(f(W^\top X) \neq f(W'^\top X)) \le 6b + 3L_q^2 \mathbb{E}\|(W-W')^\top X\|^2 = 6b + 3L_q^2 \|W - W'\|_F^2.$$

First choose b small and then choose the frame distance small. Taking the supremum over all frame pairs within distance s defines the claimed modulus, uniformly in dimension. □

It follows also that $J_j(A;\eta)$ is continuous on the compact set of contractions. Indeed, both coefficient frames $\binom{A}{B_A}$ are orthonormal, and $A \mapsto B_A$ is continuous. For either program,

$$|M_{j,\eta}(D) - M_{j,\eta}(D')| \le \mathbb{E}|D - D'|.$$

Thus the change in J_j is bounded by three times the disagreement between the corresponding competitors. In particular each supremum in Theorem A.1 is attained. Also $J_1 \le J_0$ and $J_j(I_k;\eta) = 0$.

For later use, a compact set of k-frames in $\mathbb{R}^s$ has a Euclidean ρ-net contained in the set with cardinality at most

$$(1 + 2\sqrt{k}/\rho)^{sk}.$$

Take a maximal ρ-separated set and compare disjoint ambient balls of radius $\rho/2$ with a ball of radius $\sqrt{k} + \rho/2$. The net can be fixed in advance for every integer s.

A.4 Conditional lifting and the residual certificate

Let a proposed orthonormal basis $P \in \mathbb{R}^{d\times m}$, $m \ge k$, start with the candidate frame $V \in \mathbb{R}^{d\times k}$. Write

$$U = P^\top X \in \mathbb{R}^m, \quad T = (U_1, \dots, U_k), \quad Z = (I - PP^\top)X, \quad E_V = \mathbf{1}\{Y \neq f(T)\}.$$

Let $r = \min\{k, d-m\}$. Consider coefficient matrices $C \in \mathbb{R}^{m\times k}$ and $R \in \mathbb{R}^{r\times k}$ with

$$C^\top C + R^\top R = I_k.$$

For any orthonormal residual frame $N \in \mathbb{R}^{d\times r}$ with $P^\top N = 0$, this represents a competitor

$$h_{C,R,N}(X) = f(C^\top U + R^\top N^\top Z).$$

Every competitor has such a representation: its residual component has rank at most r, and its span can be completed to an orthonormal r-frame. The case $r = 0$ has no residual variable.

For each coefficient pair define the conditional disagreement

$$D_{C,R}(u,g) = \mathbf{1}\{f(C^\top u + R^\top g) \neq f(t)\}.$$

Apply Lemma A.2 to choose bounded measurable $\alpha_{C,R}(u)$ and $\gamma_{C,R}(u) \in \mathbb{R}^r$. Put

$$\begin{aligned} A_{C,R} &= \mathbb{E}[\alpha_{C,R}(U)(E_V - \eta)], \\ V_{C,R} &= \mathbb{E}[Z\gamma_{C,R}(U)^\top E_V] \in \mathbb{R}^{d\times r}, \\ L_{C,R} &= \mathbb{E}_U[2M_{1,\eta}(D_{C,R}(U,\cdot)) - \mathbb{E}_G D_{C,R}(U,G)]. \end{aligned}$$

When $r = 0$ put $\alpha_{C,0}(u) = D_{C,0}(u)$, $V_{C,0} = 0$.

Lemma A.4 (Lifting the conditional program)**.** *Let A be the first k rows of C. Then*

$$L_{C,R} \le J_1(A;\eta).$$

Thus if the first-moment envelopes are nonpositive, every labeling satisfies, simultaneously for every residual orientation N,

$$R(h_V) - R(h_{C,R,N}) \le 2A_{C,R} + 2\sqrt{k}\,\|V_{C,R}\|_F.$$

Proof. Fix $\zeta > 0$ and use the measurable near-optimal primal selectors from Lemma A.2 to assemble a rule $e(U,G)$ whose mean conditional on U is η and whose conditional G-moment is zero. Write $U = (T,S)$, with independent standard Gaussian blocks. Conditional on T,

$$\mathbb{E}[e \mid T] = \eta, \quad \mathbb{E}[Se \mid T] = \mathbb{E}[S\mathbb{E}(e \mid T,S) \mid T] = 0, \quad \mathbb{E}[Ge \mid T] = 0.$$

Any additional residual coordinates can be integrated out. Project the combined Gaussian transverse vector (S,G) onto the span used by the competitor. Its covariance in the competitor coordinates is $I_k - A^\top A$. After an orthogonal rotation and padding by independent unused standard coordinates if necessary, the competitor has the same representation as $f(A^\top T + B_A G')$. The assembled error rule, averaged over all unused coordinates, has conditional mass η and first moment zero in every coordinate of G'. It is therefore feasible for the program defining $J_1(A;\eta)$. Its improvement is at least $L_{C,R} - 2\zeta$, proving the first claim on letting ζ tend to zero. This reasoning also covers singular B_A and $r = 0$; unused Gaussian coordinates have zero error moment by independence.

For an arbitrary actual labeling, the dual pointwise inequality gives

$$R(h_V) - R(h_{C,R,N}) \le L_{C,R} + 2A_{C,R} + 2\,\mathrm{tr}(N^\top V_{C,R}).$$

This follows by conditioning on U and using $N^\top Z \sim N(0, I_r)$ independently of U; no independence of the label is needed. Since $\|N\|_F = \sqrt{r} \le \sqrt{k}$, Frobenius Cauchy–Schwarz and the nonpositive-envelope assumption prove the second claim. For $r = 0$ the exact gain is $2A_{C,0} - (1-2\eta)\mathbb{E}D_{C,0} \le 2A_{C,0}$. □

The rotation-and-padding step can also be expressed using matrices. If the competitor's transverse coefficient matrix is H, then $H^\top H = I_k - A^\top A = B_A^2$. A singular value decomposition identifies its nonzero Gaussian coordinates with those of $B_A G'$. The zero first moments are preserved by orthogonal transformations, conditional expectation, and independent padding. No inverse of a possibly singular covariance matrix is being assumed.

A.5 Finite-sample norm tests

Throughout this section let $B = 4$. Conditional on a fixed message and coefficient pair, vectorize the matrix observation

$$W = Z\gamma_{C,R}(U)^\top E_V \in \mathbb{R}^{d\times r}.$$

For any matrix Q with $\|Q\|_F = 1$, independence of U, Z gives

$$\mathbb{E}\langle Q, W\rangle_F^2 \le \mathbb{E}\{Z^\top Q\gamma_{C,R}(U)\}^2 \le B^2.$$

Consequently $\mathrm{Cov}(\mathrm{vec}\, W) \preceq B^2 I_{dr}$ for every actual labeling. The scalar observation in $A_{C,R}$ belongs to $[-\eta, 1-\eta]$.

For iid vectors $W_i \in \mathbb{R}^p$ with mean μ and covariance $\Sigma \preceq B^2 I$, use

$$U_n = \frac{1}{n(n-1)} \sum_{i \neq j} \langle W_i, W_j \rangle.$$

Direct expansion yields

$$\mathbb{E}U_n = \|\mu\|^2, \qquad \mathrm{Var}(U_n) = \frac{4\mu^\top \Sigma \mu}{n} + \frac{2\,\mathrm{tr}(\Sigma^2)}{n(n-1)}.$$

For $n \geq \max\{2, 512B^2\sqrt{p}/t^2\}$, comparison with $t^2/2$ accepts if $\|\mu\| \leq t/4$ and rejects if $\|\mu\| > t$, with error at most $1/16$ in either case. For the first case, use the gap $7t^2/16$ and substitute $\|\mu\|^2 \leq t^2/16$ in the variance bound. For the second use the gap at least $\|\mu\|^2/2$ and substitute $\|\mu\| > t$. Chebyshev gives the stated constant. The median of $O(\log(M/\delta))$ independent blocks makes the test simultaneously valid for any fixed family of M statistics, with failure at most δ. Blocks may be reused across the statistics in this family.

The honest prover needs a stronger, ordinary mean estimate. For a unit Frobenius matrix Q, the random variable $H = \langle Q, W \rangle_F$ is pointwise dominated in absolute value by $|Z^\top Q \gamma(U)|$. Conditional on U, the latter is the absolute value of a centered Gaussian of variance at most B^2. Thus all absolute moments of H are dominated by those of BG_1. For an independent copy H', symmetry and $|x-y|^{2j} \leq 2^{2j-1}(|x|^{2j} + |y|^{2j})$ show

$$\mathbb{E}(H - H')^{2j} \leq 4^j B^{2j} \frac{(2j)!}{2^j j!}.$$

Jensen and the exponential series imply

$$\mathbb{E}\exp\{s(H - \mathbb{E}H)\} \leq \mathbb{E}e^{s(H-H')} \leq e^{2B^2 s^2}.$$

A $1/2$-net of the unit sphere in $\mathbb{R}^{dr}$ therefore gives, with a conservative numerical constant,

$$\mathbb{P}\{\|\widehat{V}_{C,R} - V_{C,R}\|_F > \xi\} \leq 2 \cdot 5^{dr} \exp\{-n\xi^2/(64B^2)\}. \tag{A.1}$$

The argument does not assume the label independent of the Gaussian coordinate; the initial domination discards its indicator.

A.6 The square-root-dimensional protocol

It suffices to treat $0 < \varepsilon < 1$ and $0 < \delta \leq 1/4$. Choose $\rho > 0$ with $\omega_f(\rho) \leq \varepsilon/8$, and put

$$t = \frac{\varepsilon}{32(1+\sqrt{k})}, \quad K = \left\lceil \frac{256kB^2}{t^2} \right\rceil, \quad M_* = (1 + 2\sqrt{k}/\rho)^{(2k+K)k}.$$

For every $m \leq \min\{d, k+K\}$ use a fixed ρ-net contained in the coefficient-frame set $\{(C,R) : C^\top C + R^\top R = I_k\}$, where $r = \min\{k, d-m\}$. Its cardinality is at most M_*.

The prover sends an orthonormal basis P of at most $k+K$ directions, with the candidate frame its first k columns. Use fixed orderings to resolve every choice from a finite net and every tie in the constructions below. Invalid messages are rejected. For every coefficient pair in the appropriate net, the verifier estimates $A_{C,R}$ to accuracy t and requires its estimate to be at most $2t$. It also performs the preceding norm test, with acceptance complete for $\|V_{C,R}\|_F \leq t/4$ and sound against $\|V_{C,R}\|_F > t$. Set the total scalar and norm failure probability to $\delta/2$. On acceptance output h_V.

Conditional on any message, the scalar observations are bounded and the vector covariance bounds hold. Hoeffding and the norm test use a total of

$$O\left(B^2\sqrt{kd}\,t^{-2}\log\frac{M_*}{\delta}\right) = O_{f,k,\eta,\varepsilon}(\sqrt{d}\log(1/\delta))$$

private observations. When $r = 0$ there are no vector tests. At every accepted coefficient pair on the confidence event, $A_{C,R} \le 3t$ and $\|V_{C,R}\|_F \le t$. By Lemma A.4, its gain is at most $(6+2\sqrt{k})t \le \varepsilon/4$, uniformly over residual frames. To cover any actual competitor, represent its normal frame by (C, R, N) and replace (C, R) by a net pair, retaining N. Because $[P, N]$ is an isometry, the two resulting d-dimensional frames have Frobenius distance at most ρ. Their risk difference is at most $\omega_f(\rho) \le \varepsilon/8$. Thus every competitor improves risk by at most $3\varepsilon/8$. This proves soundness with room to spare, without any assumption on the prover's strategy or on the label law.

A.6.1 An independent sample-only honest prover

Candidate construction. Let $a = 1 - 2\eta > 0$ and $q_0 = t/2$. Choose a sufficiently fine Stiefel-frame net that its disagreement radius is at most $aq_0/4$. Its logarithmic cardinality is $O_{f,k,\eta,\varepsilon}(d)$ by Lemma A.3. Empirical risk minimization over this net, using independent observations and accuracy $aq_0/4$ for every net risk, returns a frame V with

$$R(h_V) - R(h_{W_*}) \le 3aq_0/4, \qquad \mathbb{P}(h_V(X) \ne h_{W_*}(X)) \le q_0,$$

except with probability $\delta/4$. The second assertion follows because Massart noise gives $R(h_V) - R(h_{W_*}) \ge a\mathbb{P}(h_V \ne h_{W_*})$. Hoeffding uses $O_{f,k,\eta,\varepsilon}(d + \log(1/\delta))$ observations.

Revealing a bounded nuisance span. Start with $P_0 = V$. At every non-full-dimensional stage, form its fixed coefficient net and use a fresh batch to estimate every matrix $V_{C,R}$ in Frobenius norm to accuracy

$$\xi = \frac{t}{32\sqrt{k}}.$$

By (A.1), a batch of

$$O\left(B^2\xi^{-2}\left[kd + \log\frac{(K+1)M_*}{\delta}\right]\right)$$

observations suffices with conditional failure at most $\delta/(4(K+1))$. If every estimated Frobenius norm is at most $t/8$, stop. Otherwise select one offending matrix, choose a column of largest estimated norm, and append that column after normalizing it. Each estimated column lies exactly in the current orthogonal complement, since every observation does. If the span becomes full-dimensional, stop. Allow at most K additions and abort if no stopping stage is reached by then.

On accurate batches, an added column has a direction n_j and a dual coordinate $\gamma_{j,\ell}$ satisfying

$$\mathbb{E}[e_V(X)\gamma_{j,\ell}(U_j)(n_j \cdot X)] > \frac{t}{8\sqrt{k}} - \xi > \frac{t}{16\sqrt{k}}, \qquad e_V(X) = \mathbb{P}(Y \ne h_V(X) \mid X).$$

Define $F_j(X) = \gamma_{j,\ell}(U_j)(n_j \cdot X)$. Its squared L^2 norm is at most B^2. For $i < j$, F_i is measurable in the coordinates already in P_j, whereas F_j has conditional mean zero given those coordinates. Thus the F_j are pairwise orthogonal. Bessel's inequality implies

$$\#\{\text{additions}\}\frac{t^2}{256kB^2} < \sum_j \frac{|\mathbb{E}[e_V F_j]|^2}{\mathbb{E}F_j^2} \le \mathbb{E}e_V^2 \le 1.$$

Consequently fewer than K additions occur on accurate batches, and the construction terminates. At a non-full-dimensional stopping stage, $\|V_{C,R}\|_F \le t/8 + \xi < t/4$ for every final coefficient pair. At a full-dimensional stage all residual tests vanish.

The fresh-batch conditional concentration bounds union bound over all stages, despite their adaptive choice, giving total failure at most $\delta/4$. Since K, M_*, ξ are constants at fixed f, k, η, ε, the total honest sample count remains $O_{f,k,\eta,\varepsilon}(d + \log(1/\delta))$.

Scalar completeness. On the candidate success event the Massart error rule gives

$$e_V(X) \le \eta + \mathbf{1}\{h_V(X) \ne h_{W_*}(X)\}.$$

Since every scalar dual is in $[0, 1]$, for every final coefficient pair one has $A_{C,R} \le q_0 \le t/2$. All scalar checks therefore accept on their confidence event; the construction guarantees the norm-completeness conditions. Prover failures cost at most $\delta/2$, and verifier failures another $\delta/2$. This proves the second upper bound and its full Massart completeness.

A.7 The dimension-independent and linear upper bounds

For the first upper bound use only the candidate frame V and its coordinates $T = V^\top X$. Put $r = \min\{k, d - k\}$ and use a fixed coefficient net in the $2k$-dimensional (or smaller) frame set. The mass-only conditional dual is

$$\alpha_{C,R}(T) = \mathbf{1}\{\mathbb{P}_G(D_{C,R}(T, G) = 1) \ge \eta\}.$$

Its value does not depend on residual orientation. Conditional Gaussian rotational invariance and independent padding show that the integrated envelope for (C, R) is $J_0(C; \eta)$, since the revealed span has exactly the anchor's k dimensions. Therefore

$$R(h_V) - R(h_{C,R,N}) \le J_0(C; \eta) + 2\mathbb{E}[\alpha_{C,R}(T)(E_V - \eta)].$$

Use a coefficient net with prediction radius at most $\varepsilon/8$ and put $t_0 = \varepsilon/32$. Estimate every displayed scalar expectation to accuracy t_0, and accept when all estimates are at most $2t_0$. If every J_0 is nonpositive, an accepted message on the confidence event has gain at most $6t_0$ against every net competitor, and at most $6t_0 + \varepsilon/8 < \varepsilon$ against every competitor. Hoeffding and a union bound require $O_{f,k,\eta,\varepsilon}(\log(1/\delta))$ private observations. For completeness, the preceding candidate learner is run with disagreement target $q_0 = t_0/2$. Nonnegative scalar duals make every population quantity at most q_0, so all estimates are below the acceptance threshold. Split the failure budget equally between candidate construction and verification. There is no nuisance-span construction in this case.

For the unconditional linear upper bound, fix a sufficiently fine net of all k-frames in $\mathbb{R}^d$ with prediction radius at most $\varepsilon/8$. The logarithm of its size is $O_{f,k,\varepsilon}(d)$. Given a candidate, estimate its risk and all net risks to accuracy $\varepsilon/8$ and accept if its estimated risk is at most the minimum net risk plus $\varepsilon/2$. On uniform accuracy, acceptance implies

$$R(h_V) \le \min_{W \text{ in the net}} R(h_W) + 3\varepsilon/4 \le \mathrm{OPT} + 7\varepsilon/8.$$

For completeness learn a candidate with risk at most the Massart target risk plus $\varepsilon/8$. The target is Bayes, so on uniform accuracy the candidate's estimated risk exceeds the smallest estimated net risk by at most $3\varepsilon/8$; it therefore passes. Uniform Hoeffding bounds and the candidate construction give both sample counts as $O_{f,k,\eta,\varepsilon}(d + \log(1/\delta))$.

A.8 Noise thresholds

The classification can be organized as two critical noise levels for each fixed rule, possibly coincident or degenerate. For each A, $J_j(A;\eta)$ is nondecreasing and 2-Lipschitz in η on $(0,1/2)$. Monotonicity follows by replacing a feasible error rule e by

$$e' = e + \frac{\eta' - \eta}{1-\eta}(1-e), \qquad \eta' > \eta,$$

which raises its mass to η', preserves zero first moment when present, and increases its disagreement objective. The Lipschitz bound follows from the dual formula and $0 \le \alpha \le 1$, by using an optimizer for either noise level in the other dual. Thus the sets of noise levels where every envelope is nonpositive are closed initial intervals (relative to $(0,1/2)$), with the mass-only interval contained in the first-moment interval. Any threshold lying in $(0,1/2)$ belongs to the lower-sample regime in this fixed-accuracy theorem. This does not determine sharp accuracy dependence at the thresholds.

A.9 Unknown intrinsic rules from a totally bounded family

The fixed-rule proof above extends to ordered comparisons between possibly different intrinsic rules. We give the uniform approximation, lifting, and finite implementation explicitly. Let $\mathcal{F}$ be a nonempty known family of Borel Boolean functions on $\mathbb{R}^K$, identified modulo Gaussian-null sets, totally bounded in Gaussian disagreement $d_\gamma(f,g) = \mathbb{P}(f(G) \ne g(G)) = \|f-g\|_2^2/4$. For every $b > 0$ fix a finite contained b-net $\mathcal{F}[b] \subset \mathcal{F}$. No measurable parametrization of the entire family is assumed: protocols use only these finite alphabets. For $d \ge K$, let $\mathcal{H}_{\mathcal{F},d}$ be the union of its orthonormal projection classes. Define

$$D_{f,g,A}(u,z) = \mathbf{1}\{g(A^\top u + B_A z) \ne f(u)\},$$

$$J_j(f,g,A;\eta) = \mathbb{E}_U[2M_{j,\eta}(D_{f,g,A}(U,\cdot)) - \mathbb{E}_G D_{f,g,A}(U,G)].$$

The ordered rules $f, g \in \mathcal{F}$ and contraction A range independently.

Theorem A.5 (Unknown-rule family classification)**.** *The three upper bounds and honest-prover budgets of Theorem A.1 hold for $\mathcal{H}_{\mathcal{F},d}$, with constants depending on $\mathcal{F}, K, \eta, \varepsilon$, when each envelope condition is imposed on every ordered triple (f,g,A). The protocols are proper, use one message, have full Massart completeness and arbitrary-label Gaussian soundness against the entire union, and name a rule from a fixed finite subset of $\mathcal{F}$. The square-root protocol reveals at most $K + O_K(\varepsilon^{-2})$ orthonormal directions.*

A positive witness $J_j(f,g,A;\eta) = \Gamma$ gives exactly the lower bounds of Theorem B.1, with rank K, for $0 < \varepsilon \le \Gamma/8$. Thus the exhaustive fixed-accuracy classification in Theorem 3.1 holds even when the intrinsic rule is unknown. Total boundedness suffices; neither closedness nor attainment of the comparator infimum is needed.

Lemma A.6 (Uniform family modulus)**.** *There is a nondecreasing $\omega_{\mathcal{F}}$ tending to zero at zero such that, uniformly over ambient dimensions and orthonormal frames,*

$$\mathbb{P}(f(W^\top X) \ne g(W'^\top X)) \le d_\gamma(f,g) + \omega_{\mathcal{F}}(\|W - W'\|_F).$$

Proof. For each $b > 0$, choose finitely many $f_i \in \mathcal{F}$ with every $f \in \mathcal{F}$ within squared L^2 distance $b/4$ of some f_i. Approximate each f_i within squared distance $b/4$ by a bounded Lipschitz q_i. Then

$\|f - q_i\|_2^2 \le b$ for an appropriate i, and the finite list has a common Lipschitz bound L. The proof of Lemma A.3 gives

$$4\mathbb{P}(f(W^\top X) \neq f(W'^\top X)) \le 6b + 3L^2\|W - W'\|_F^2.$$

First decrease b and then the frame distance. Taking a supremum defines a uniform modulus. Insert $g(W^\top X)$ and use the triangle inequality to prove the displayed joint bound. □

Proof of Theorem A.5. Write $a = 1 - 2\eta$. An honest candidate at target disagreement $q > 0$ is obtained by empirical risk minimization over $\mathcal{F}[aq/8]$ times an ambient frame net with $\omega_{\mathcal{F}}(\rho_q) \le aq/8$. This product contains a hypothesis with target disagreement at most $aq/4$. Uniform risk accuracy $aq/8$ gives excess risk at most $aq/2$, and full Massart noise converts this to target disagreement at most q. The number of observations is

$$O\left((aq)^{-2}\left[\log|\mathcal{F}[aq/8]| + dK\log(1 + 2\sqrt{K}/\rho_q) + \log(1/\beta)\right]\right) = O_{\mathcal{F},K,\eta,q}(d + \log(1/\beta)).$$

For each protocol the learner's finite rule net is its predeclared candidate alphabet. Every center is in $\mathcal{F}$, so the candidate is proper. Finite choices have fixed tie orders. No selection over the whole family is used.

For cross-rule lifting, fix a candidate $f(V^\top X)$, a revealed frame P beginning with V, and a comparator $g((PC+NR)^\top X)$. Write $U = P^\top X = (T, S)$ and $Z = (I - PP^\top)X$. Apply Lemma A.2 to $D(U, G) = \mathbf{1}\{g(C^\top U + R^\top G) \neq f(T)\}$, obtaining $\alpha(U), \gamma(U)$, and set

$$A_{g,C,R} = \mathbb{E}[\alpha(U)(E_V - \eta)], \qquad V_{g,C,R} = \mathbb{E}[Z\gamma(U)^\top E_V].$$

If A comprises the first K rows of C, the integrated conditional baseline is at most $J_1(f, g, A; \eta)$. Indeed the primal selections obey $\mathbb{E}[e \mid T, S] = \eta$ and $\mathbb{E}[Ge \mid T, S] = 0$; averaging gives $\mathbb{E}[e \mid T] = \eta$ and $\mathbb{E}[(S, G)e \mid T] = 0$. An orthogonal rotation, conditional averaging over unused coordinates, and independent padding identify the comparator with $g(A^\top T + B_A G')$. They preserve both moment constraints, including for singular B_A. Taking the primal selection error to zero proves the baseline bound, exactly as in Lemma A.4. The candidate and comparator rules need never coincide. Universal nonpositivity therefore gives

$$R(h_{f,V}) - R(h_{g,PC+NR}) \le 2A_{g,C,R} + 2\sqrt{K}\|V_{g,C,R}\|_F. \tag{A.2}$$

The scalar dual stays in $[0, 1]$, and the vector dual has norm at most $B = 4$. With no residual coordinates use $\alpha = D$ and $V_{g,C,0} = 0$.

For the square-root protocol choose

$$b = \varepsilon/16, \quad \mathcal{F}_C = \mathcal{F}[b], \quad \omega_{\mathcal{F}}(\rho) \le \varepsilon/16, \quad t = \frac{\varepsilon}{64(1 + \sqrt{K})}, \quad L = \left\lceil \frac{256KB^2}{t^2} \right\rceil.$$

The candidate alphabet is $\mathcal{F}[aq_0/8]$, where $q_0 = t/2$. For every possible revealed rank $m \le \min(d, K + L)$ take a contained ρ-net of the coefficient frames (C, R), with $r = \min(K, d - m)$. At any stage there are at most

$$M_* = |\mathcal{F}_C|(1 + 2\sqrt{K}/\rho)^{(2K+L)K}$$

rule/coefficient pairs. The message names a candidate rule and a frame P with at most $K + L$ columns, the first K being V. Reject invalid messages. Estimate every $A_{g,C,R}$ to tolerance t and require its estimate to be at most $2t$; apply the norm test from the fixed-rule proof with completeness

at $t/4$ and rejection above t. Give these tests total failure budget $\delta/2$. Conditional on any message, the same covariance bound $B^2 I$ holds, so their deterministic budget is $O(B^2\sqrt{Kd}\,t^{-2}\log(M_*/\delta))$.

On acceptance and simultaneous concentration, (A.2) bounds each tested gain by $(6+2\sqrt{K})t \le \varepsilon/8$, uniformly in residual orientation. To cover an arbitrary competitor, first round its intrinsic rule to $\mathcal{F}_C$, costing risk at most $\varepsilon/16$. Next round its coefficient frame while keeping its residual orientation, costing another $\varepsilon/16$ by the uniform modulus. The total gain is at most $\varepsilon/4$. This proves soundness against the entire union, even when its optimal risk is not attained.

For honest implementation, learn a candidate with target disagreement q_0 at failure $\delta/4$. At each revealed span estimate all M_* residual matrices on a fresh prover batch to tolerance $\xi = t/(32\sqrt{K})$. Equation (A.1) shows that

$$O\left(B^2\xi^{-2}[Kd + \log((L+1)M_*/\delta)]\right)$$

observations suffice per batch with conditional failure at most $\delta/(4(L+1))$. Stop when all estimated norms are at most $t/8$; otherwise append the normalized largest column of the first offending matrix. Stop also when the span is full, and cap additions at L, aborting if the stopping test has not succeeded. On accurate batches each added direction gives a witness $F_j = \gamma_{j,\ell}(P_j^\top X)(n_j \cdot X)$ with $|\mathbb{E}[e_V F_j]| > t/(16\sqrt{K})$ and $\mathbb{E}F_j^2 \le B^2$. Although the comparator rule may change at every step, F_j has conditional mean zero given the current span, and every earlier witness is measurable in that span. Bessel therefore gives fewer than $256KB^2/t^2 \le L$ additions. Aborting is impossible on accurate batches; the final residual norms are at most $t/8 + \xi < t/4$. The fresh-batch union bound costs $\delta/4$ and the deterministic cap keeps the total honest budget linear in d at fixed parameters. Finally, full ambient Massart noise gives $e_V \le \eta + \mathbf{1}\{h_{f,V} \ne h_*\}$. Nonnegative scalar duals imply $A_{g,C,R} \le q_0 = t/2$ for every final pair, so all honest scalar tests pass. Learning, compression, and verification failures total at most δ.

For universal $J_0 \le 0$, omit compression and vector tests. Use the candidate span, the same $\varepsilon/16$ rule and prediction meshes, and scalar tolerance $t_0 = \varepsilon/32$ with acceptance threshold $2t_0$. The scalar version of the cross-rule baseline bound gives gain at most $6t_0$ on accepted net points; the two rounding errors total $\varepsilon/8$, leaving total gain below ε. The honest learner uses target disagreement $t_0/2$, so all scalar tests pass. The finite net is independent of dimension, giving the first upper bound.

Without an envelope condition, use the product of $\mathcal{F}[\varepsilon/16]$ and an ambient frame net of disagreement radius $\varepsilon/16$. Its logarithmic size is $O_{\mathcal{F},K,\varepsilon}(d)$. Estimate the candidate and all net risks to tolerance $\varepsilon/8$, accepting when its estimated risk is at most the least net estimate plus $\varepsilon/2$. Acceptance gives risk at most $\mathrm{OPT} + 7\varepsilon/8$. An honest candidate within $\varepsilon/8$ of the target in risk passes, since the target is Bayes. This proves the linear budgets.

For the lower bounds fix an actual positive triple (f, g, A) and use Lemma A.2's measurable primal selection to obtain gain at least $\Gamma/2$. The null regression is $(1-2\eta)f(U)$, and the hidden-frame alternative is $f(U)(1 - 2e(U, V^\top Z))$. Its improving comparator is now $g(A^\top U + B_A V^\top Z)$, whose frame has Gram matrix I_K and belongs to the full union. The normalized likelihood kernel depends only on $e - \eta$, since $f(U)^2 = 1$; it is the same kernel bounded in Lemmas B.3–B.4. The known-null prover simulation and relative-risk validation in Theorem B.1 apply to this comparator without change: the anchor is Bayes at the null, while the alternative optimum is at most $\eta - \Gamma/2$. This proves its explicit interactive and improper lower bounds with the same constants. A positive supremum supplies an actual positive triple even if no maximum exists. Since $J_1 \le J_0$, the three cases are exhaustive. □

A concrete infinite family. For fixed degree s, Boolean polynomial threshold rules in K variables form a compact intrinsic family. Normalize each nonzero polynomial's coefficient vector to the finite-dimensional unit sphere. The sign map is continuous in Gaussian L^2: coefficients converge pointwise, and a nonzero limiting polynomial has a Gaussian-null zero set, so dominated convergence applies. Compactness of the sphere proves the assertion. The constant rules are included. Thus the theorem covers unknown polynomial shape and coefficients, without an entropy-independent bound.

Even affine thresholds illustrate why cross-rule comparisons matter. The one-dimensional family $f_t(u) = \mathrm{sign}(u - t)$ for $t \in \mathbb{R} \cup \{\pm\infty\}$ is compact in Gaussian disagreement. Take the constant anchor $f \equiv -1$, choose $g(u) = \mathrm{sign}(u - t)$ with $\overline{\Phi}(t) = \eta/2$, and set $A = 0$. The even allocation $e(G) = \mathbf{1}\{|G| > t\}$ has mass η, zero first moment, and captures all disagreement. Hence

$$J_0(f, g, 0; \eta) = J_1(f, g, 0; \eta) = \eta/2.$$

All affine Gaussian halfspaces, including constants, therefore have $\Theta_{\eta,\varepsilon}(d)$ verifier complexity for every fixed $0 < \eta < 1/2$ and $0 < \varepsilon \le \eta/16$, despite the low-noise dimension-independent homogeneous subclass. This also makes the full polynomial-threshold family linearly hard for every degree $s \ge 1$. The example evaluates the family criterion; it does not supply a computational lower bound or a sharp joint parameter law.

B Lower bounds from conditional moment obstructions

We prove the lower implications in Theorem A.1. Throughout this section, f, k, η, the class $\mathcal{H}_{f,d}$, and the programs $M_{j,\eta}$ and J_j are as in Section A. Fix a contraction $A \in \mathbb{R}^{k\times k}$, put $B = (I_k - A^\top A)^{1/2}$, and let $a = 1 - 2\eta$. The proof converts a positive conditional gain into a labeled random-subspace mixture that is indistinguishable from a fixed random-classification-noise null. Gaussian moment-correlation bounds are standard; see, for example, [DKS17, Lemma 3.4]. The last step is the null-prover testing reduction [MS26, Section 4], also developed in [BDFP26, Section 5]. We give the complete argument to retain unrestricted interaction and improper outputs.

Theorem B.1 (Unrestricted conditional-moment lower bounds)**.** *Suppose $d \ge 2k$ and $J_j(A; \eta) = \gamma > 0$ for some A and $j \in \{0, 1\}$. Put $g = \gamma/2$ and*

$$\ell = \left\lceil \frac{128 \log 24}{g^2} \right\rceil.$$

For every $0 < \varepsilon \le \gamma/8$, any verifier that is complete with probability at least $2/3$ at the single known null

$$Q: \qquad X \sim N(0, I_d), \qquad \mathbb{E}_Q[Y \mid X] = a f(X_1, \ldots, X_k),$$

and sound with probability at least $2/3$ against arbitrary Gaussian-marginal stochastic labels must satisfy

$$m_V + \ell > \begin{cases} \sqrt{d-k}/(100k), & j = 0, \\ (d-k)/(100k^2), & j = 1. \end{cases}$$

Soundness means that, for every prover, the probability of outputting a predictor of risk greater than $\inf_{h\in\mathcal{H}_{f,d}} R(h) + \varepsilon$ is at most $1/3$; rejection is allowed. Outputs may be arbitrary measurable Boolean predictors. Interaction, computation, communication, and the malicious prover's knowledge are unrestricted. An honest prover may use any number of samples independent of the verifier's samples.

In particular, positive J_0 implies $\Omega(\sqrt{d})$ and positive J_1 implies $\Omega(d)$ at fixed sufficiently small accuracy. The construction needs completeness only at exact random classification noise, hence applies to every protocol with full Massart completeness at noise bound η.

B.1 Measurable conditional allocations

For this proof, let γ_k denote standard Gaussian measure on $\mathbb{R}^k$, and write $\mathcal{E}_0$ for the feasible set $\{e \in L^2(\gamma_k) : 0 \le e \le 1,\ \mathbb{E}e = \eta\}$. Let $\mathcal{E}_1$ be its subset satisfying $\mathbb{E}[Ge(G)] = 0$. These are precisely the feasible sets in the definition of $M_{j,\eta}$. We use near-optimal allocations so that no measurable selection of exact primal optimizers is needed.

Lemma B.2. *For every $\xi > 0$ and fixed A, j, there is a jointly measurable $e : \mathbb{R}^k \times \mathbb{R}^k \to [0,1]$ such that, for every u, $e(u,\cdot) \in \mathcal{E}_j$ and*

$$\mathbb{E}_G[e(u,G)D_A(u,G)] \ge M_{j,\eta}(D_A(u,\cdot)) - \xi.$$

The functions $u \mapsto M_{j,\eta}(D_A(u,\cdot))$ are measurable.

Proof. The set $\mathcal{E}_j$ is nonempty (it contains the constant η) and is a subset of the separable metric space $L^2(\gamma_k)$. Choose a countable norm-dense subset $\{e_n : n \ge 1\}$ of it, with measurable representatives taking values in $[0,1]$. The moment conditions are statements about equivalence classes and thus hold for these representatives. Pairing with the bounded function $D_A(u,\cdot)$ is norm-continuous. Consequently

$$M_{j,\eta}(D_A(u,\cdot)) = \sup_{n \ge 1} \mathbb{E}_G[e_n(G)D_A(u,G)].$$

Each term on the right is measurable by Fubini, and so is the supremum. Select the least $n = n(u)$ with the displayed expectation at least $M_{j,\eta}(D_A(u,\cdot)) - \xi$. This is a measurable integer-valued function, and $e(u,g) = e_{n(u)}(g)$ has the required properties. □

Take $\xi = \gamma/4$. The resulting allocation has gain

$$\Gamma := \mathbb{E}[2e(U,G)D_A(U,G) - D_A(U,G)] \ge \gamma/2 = g. \tag{B.1}$$

Exact attainment of the conditional programs is unnecessary.

B.2 A labeled random-frame construction

Write $X = (U,Z)$, where $Z \sim N(0, I_D)$, $D = d - k$, independently of U. Let $V \in \mathbb{R}^{D \times k}$ have orthonormal columns. Define the alternative distribution P_V by the same feature marginal and the regression

$$r_V(U,Z) = f(U)(1 - 2e(U, V^\top Z)).$$

These are valid regressions since $0 \le e \le 1$. The anchor $h_0(U,Z) = f(U)$ satisfies

$$R_{P_V}(h_0) = \eta = R_Q(h_0), \qquad \mathbb{E}_{P_V}[Y \mid U] = af(U).$$

The competing function

$$h_V(U,Z) = f(A^\top U + BV^\top Z)$$

belongs to $\mathcal{H}_{f,d}$: its frame is the vertical block matrix with blocks A and VB, whose Gram matrix is $A^\top A + B^2 = I_k$. The paired-risk identity gives

$$R_{P_V}(h_0) - R_{P_V}(h_V) = \Gamma \ge g, \qquad \inf_{h \in \mathcal{H}_{f,d}} R_{P_V}(h) \le \eta - g. \tag{B.2}$$

This remains valid when B is singular. If $B = 0$, the disagreement does not depend on G and positive gain is impossible, consistently with the premise.

B.3 Multivariate Gaussian contraction

Lemma B.3 (A standard chaos contraction)**.** *Let (G, H) be jointly standard Gaussian vectors in $\mathbb{R}^k$, with cross-covariance $R = \mathbb{E}[GH^\top]$ and $\|R\|_{\mathrm{op}} \le 1$. If $b, c \in L^2(\gamma_k)$ are orthogonal to every polynomial of total degree less than s, then*

$$|\mathbb{E}[b(G)c(H)]| \le \|R\|_{\mathrm{op}}^s \|b\|_2 \|c\|_2.$$

The statement holds for $s \ge 1$, including singular joint Gaussian laws and $\|R\|_{\mathrm{op}} = 1$.

Proof. Take a singular value decomposition $R = L\,\mathrm{diag}(\rho_1, \dots, \rho_k)M^\top$ and rotate the two arguments. Orthogonal changes of coordinates preserve each homogeneous Gaussian chaos. In the rotated coordinates, the pairs (G_i, H_i) are independent across i, and their correlations are $\rho_i \in [0, 1]$. For normalized tensor Hermites H_α, the scalar generating function identity gives

$$\mathbb{E}[H_\alpha(G)H_\beta(H)] = \mathbf{1}\{\alpha = \beta\} \prod_i \rho_i^{\alpha_i}.$$

On total degree q, the resulting diagonal operator has norm at most $\max_i \rho_i^q = \|R\|_{\mathrm{op}}^q$. Expand b, c in orthogonal chaoses of degrees $q \ge s$, apply this bound degree by degree, and use Cauchy–Schwarz to sum their products of norms. Approximation by finite Hermite expansions proves the claim for L^2 functions; all cross-expectations are continuous under this approximation. □

Set $r_0 = af(U)$ and $\Delta_V = r_V - r_0$. Since $|r_0| = a < 1$,

$$\frac{dP_V}{dQ}(X, Y) = 1 + \frac{Y - r_0(X)}{1 - a^2}\Delta_V(X).$$

For two residual frames V, W, the one-observation likelihood cross-moment is

$$\begin{aligned} K(V, W) &:= \mathbb{E}_Q\left[\frac{dP_V}{dQ}\frac{dP_W}{dQ}\right] = 1 + \mathbb{E}\frac{\Delta_V(X)\Delta_W(X)}{1 - a^2} \\ &= 1 + \mathbb{E}_U\mathbb{E}_Z[b_U(V^\top Z)b_U(W^\top Z)], \quad b_u(g) = \frac{e(u, g) - \eta}{\sqrt{\eta(1 - \eta)}}. \end{aligned} \tag{B.3}$$

The conditional variance bound $\mathbb{E}(e - \eta)^2 \le \eta(1 - \eta)$ gives $\|b_u\|_2 \le 1$. For $j = 0$, b_u has zero mean; for $j = 1$, its first moments also vanish. With $s = j + 1$ and $R = V^\top W$, Lemma B.3 therefore gives the global bound

$$|K(V, W) - 1| \le \|V^\top W\|_{\mathrm{op}}^{j+1}. \tag{B.4}$$

The normalization is exact because $1 - a^2 = 4\eta(1 - \eta)$. No truncation in overlap or exceptional-frame argument is needed. Also $K(V, W) \ge 0$, directly from its definition as an expectation of nonnegative likelihood ratios.

B.4 Haar-frame moments and unrestricted testing

Let V, W be independent uniform elements of the Stiefel manifold of orthonormal k-frames in $\mathbb{R}^D$.

Lemma B.4. *For $0 \le t < D/(2k^2)$,*

$$\mathbb{E}\exp\{t\|V^\top W\|_F^2\} \le (1 - 2tk^2/D)^{-1/2}.$$

Proof. Each of the k^2 entries of $V^\top W$ has the distribution of a coordinate S of a uniform unit vector in $\mathbb{R}^D$. They need not be independent. The exact spherical moments yield

$$\mathbb{E}S^{2q} = \frac{(2q-1)!!}{D(D+2)\cdots(D+2q-2)} \le \frac{(2q-1)!!}{D^q}.$$

Expanding an exponential with nonnegative terms gives $\mathbb{E}e^{uS^2} \le (1-2u/D)^{-1/2}$ for $2u < D$. Hölder's inequality with k^2 equal exponents now gives

$$\mathbb{E}\prod_{i,l}\exp\{t(V^\top W)_{il}^2\} \le \prod_{i,l}(\mathbb{E}\exp\{tk^2(V^\top W)_{il}^2\})^{1/k^2},$$

which is the claimed bound. This proof is also valid for $k = 1$. □

Let $\overline{P}_n = \mathbb{E}_V P_V^{\otimes n}$, where the same random frame is shared across all observations. Independence of observations conditional on that frame gives exactly

$$1 + \chi^2(\overline{P}_n \| Q^{\otimes n}) = \mathbb{E}_{V,W} K(V,W)^n.$$

If $j = 1$, (B.4) and Lemma B.4 imply

$$1 + \chi^2 \le \mathbb{E}e^{n\|V^\top W\|_F^2} \le (1 - 2nk^2/D)^{-1/2}.$$

Thus $n \le D/(100k^2)$ makes this at most $\sqrt{50/49}$. If $j = 0$, put $t = D/(100k^2)$ and use $nx \le n^2/(4t) + tx^2$ to get

$$1 + \chi^2 \le \mathbb{E}e^{n\|V^\top W\|_F} \le \exp\{25n^2k^2/D\}\sqrt{50/49}.$$

For $n \le \sqrt{D}/(100k)$ the latter is at most $e^{1/400}\sqrt{50/49} < 10/9$. In either case

$$\mathrm{TV}(\overline{P}_n, Q^{\otimes n}) \le \tfrac{1}{2}\sqrt{\chi^2(\overline{P}_n \| Q^{\otimes n})} < 1/6. \tag{B.5}$$

These are bounds against all measurable tests, not only statistical queries, low-degree tests, or computationally efficient algorithms.

B.5 Interactive and improper verification reduction

Proof of Theorem B.1. Fix an honest prover strategy for the known Q. Run that strategy with the verifier on observations from the unknown distribution. All prover samples can be generated independently from Q. Under Q this is honest; under every P_V it is a permissible malicious strategy. Its replies may depend on the verifier's messages. The complete execution is nevertheless one fixed randomized mapping of the verifier's observations and independent coins.

If the verifier rejects, declare alternative. Otherwise, let h be its arbitrary output. On ℓ fresh observations estimate the relative loss

$$L = \mathbb{E}[\mathbf{1}\{h(X) \ne Y\} - \mathbf{1}\{h_0(X) \ne Y\}] = R(h) - R(h_0).$$

Each summand lies in $[-1, 1]$. Hoeffding and the specified ℓ give estimation error at most $g/8$ with failure probability at most $1/12$, conditionally on the entire preceding execution. Declare null if and only if the verifier accepted and the estimate is at least $-g/2$.

Under Q, the anchor is Bayes, so $L \ge 0$ for every improper output. Completeness gives null-declaration probability at least $2/3 - 1/12 = 7/12$. Under every alternative, soundness and (B.2)

give $L \le -g + \varepsilon \le -3g/4$ except with probability at most $1/3$, whenever the verifier outputs a predictor. Consequently null declaration has probability at most $1/3 + 1/12 = 5/12$. The same is true for the mixture.

The gap between null-declaration probabilities is at least $1/6$. Data processing contradicts (B.5) if $n = m_V + \ell$ obeys the relevant displayed upper bound. This proves the theorem. Worst-case bounded stopping rules can be padded with unused independent samples. Randomized predictors can be handled by conditioning on their randomness, or by including fresh prediction coins in validation. □

C Orthogonal Boolean Compositions and the Quadratic Separation

This appendix proves Theorems 4.1 and 4.3 and Corollary 4.4. Throughout, $X \sim N(0, I_d)$ and a sign on a zero hyperplane can be chosen arbitrarily. All constants in the family upper bounds may depend on the fixed rank bound, but not on dimension or accuracy.

Fix an essential Boolean function $b : \{-1, 1\}^k \to \{-1, 1\}$, meaning each coordinate is pivotal for some assignment of the others, and put

$$f(u) = b(\operatorname{sign} u_1, \dots, \operatorname{sign} u_k), \qquad \mathcal{H}_{b,d} = \{x \mapsto f(W^\top x) : W^\top W = I_k\}.$$

An inessential coordinate can be deleted; a constant rule is trivial. The feature marginal is $N(0, I_d)$. Completeness has arbitrary ambient Massart noise around an unknown member of $\mathcal{H}_{b,d}$; soundness has arbitrary stochastic labels and requires risk at most $\mathrm{OPT} + \varepsilon$. The honest prover has independent labeled samples. Computation and exact integration of fixed-dimensional Gaussian probabilities are unrestricted.

Theorem C.1. *For every fixed essential b there is $\eta_b > 0$ such that, for every $d \ge k$, $0 < \varepsilon < 1$, and $0 < \delta \le 1/4$, there is a proper one-message verifier with Massart completeness at noise bound η_b and arbitrary-label Gaussian soundness using*

$$m_V = O_b\left(\frac{1}{\varepsilon} \log \frac{e}{\varepsilon\delta}\right), \qquad m_P = O_b\left(\frac{d \log(e/\varepsilon) + \log(1/\delta)}{\varepsilon^2}\right).$$

The message contains just a proposed orthonormal k-frame. In particular the conclusion holds at every smaller Massart noise bound. The theorem includes parity, conjunction, and majority of any fixed number of orthogonal homogeneous halfspaces. Theorem C.7 makes the rank dependence explicit.

C.1 Geometry modulo the finite symmetry group

Let $\mathcal{S}_b$ be the signed permutation matrices Q for which $f(Q^\top u) = f(u)$ almost everywhere. For frames V, W define

$$r(V, W) = \min_{Q \in \mathcal{S}_b} \|W - VQ\|_F, \quad D_{V,W} = \{f(V^\top X) \ne f(W^\top X)\}, \quad B_s(V) = \{\min_i |v_i \cdot X| \le s\}.$$

Write $r_b(V, W) = r(V, W)$. Here v_i denotes the ith column of V. All joint probabilities depend on at most $2k$ Gaussian coordinates. For the geometry one may embed both frames in $\mathbb{R}^{2k}$ and fix $V = (I_k, 0)^\top$. When $d < 2k$ this is a containing, possibly larger geometry space; bounds uniform on it also hold for every realizable ambient frame pair.

Lemma C.2 (Quantitative exclusion from a pivotal facet)**.** *Let b be essential of rank $k \geq 1$, let c be any Boolean rule of rank $\ell \geq 1$, and let V, W be orthonormal frames of these ranks. Put $\Delta = \mathbb{P}(h_{b,V}(X) \neq h_{c,W}(X))$. For every $i \in [k]$,*

$$s_i := \min_{j \in [\ell]} \sqrt{1 - (v_i \cdot w_j)^2} \quad \textit{satisfies} \quad \Delta \geq \frac{4^{-k}}{8\ell} s_i.$$

The competing rule need not be essential.

Proof. If $s_i = 0$ the assertion is immediate. Otherwise decompose $X = Tv_i + Z$, where $T \sim N(0,1)$ is independent of the standard Gaussian Z on $v_i^{\perp}$. Choose a pivotal assignment of the other $k-1$ coordinates of b, and let O be its orthant event on Z. Orthogonality of V gives

$$q := \mathbb{P}(Z \in O) = 2^{-(k-1)}.$$

For each competing normal, $w_j \cdot Z$ is a centered Gaussian with standard deviation $\sigma_j = \sqrt{1 - (v_i \cdot w_j)^2} \geq s_i$. The Gaussian density bound gives, for $h > 0$,

$$\mathbb{P}(|w_j \cdot Z| \leq h) \leq \sqrt{\frac{2}{\pi}} \frac{h}{\sigma_j} \leq \frac{h}{s_i}.$$

Set $h = qs_i/(2\ell) \leq 1/2$ and remove these ℓ bands from O. The remaining event

$$F = O \cap \bigcap_{j=1}^{\ell} \{|w_j \cdot Z| > h\}$$

has probability at least $q/2$. No independence between the removed bands or their intersection with O is required.

For $z \in F$ and $0 < t < h$, the signs of every $w_j \cdot (z + tv_i)$ and $w_j \cdot (z - tv_i)$ coincide, because $t|w_j \cdot v_i| < h < |w_j \cdot z|$. Consequently the competing predictor has the same value at these two points. The anchor predictor has opposite values there: the other anchor signs realize its chosen pivotal assignment, and the ith sign changes. At least one of the two points therefore belongs to the disagreement set. Symmetry of the density of T implies

$$\begin{aligned} \Delta &\geq \int_0^h \varphi(t)\, \mathbb{P}(Z \in F)\, dt \\ &\geq \varphi(1) \frac{q}{2} h = \varphi(1) \frac{q^2 s_i}{4\ell} = \varphi(1) \frac{4^{-k} s_i}{\ell} \geq \frac{4^{-k} s_i}{8\ell}. \end{aligned}$$

Here $\varphi(1) > 1/8$ is a deliberately loose numerical bound. □

For $K \geq 1$ define

$$A_K = 8K4^K, \qquad \gamma_K = \frac{2^{-K}}{4KA_K} = \frac{2^{-3K}}{32K^2}, \qquad c_K = \frac{\gamma_K}{2\sqrt{K}} = \frac{2^{-3K}}{64K^{5/2}}. \tag{C.1}$$

Theorem C.3 (Small disagreement forces the same rule and a matched frame)**.** *Let b, c be essential rules with ranks $1 \leq k, \ell \leq K$, and let V, W be orthonormal frames. If*

$$\Delta := \mathbb{P}(h_{b,V}(X) \neq h_{c,W}(X)) < \gamma_K,$$

then $k = \ell$ and there is a signed relabeling of the competing columns such that its induced Boolean rule is exactly b. Denoting the relabeled competing frame by $\widetilde{W}$, one has

$$\|V - \widetilde{W}\|_F \leq \sqrt{2k}\, A_K \Delta.$$

In particular, if $b = c$ then $r_b(V, W) \leq \sqrt{2k}\, A_K \Delta$.

Proof. Let $\rho = A_K\Delta$. Lemma C.2, applied in both directions, says that every anchor normal has a competing normal with projective sine distance at most ρ, and conversely. Moreover,

$$\rho < A_K\gamma_K = \frac{2^{-K}}{4K} \le \frac{1}{8}.$$

Any such match has squared inner product at least $1-\rho^2 > 1/2$. Two distinct orthonormal anchor normals cannot match the same competing normal, by Bessel's inequality. Thus the matches give an injection from the k anchor normals to the ℓ competing normals, and the reverse argument gives the opposite injection. Therefore $k = \ell$ and the first matching is a bijection.

Relabel the competing columns according to that matching and reverse their signs so that $v_i \cdot \widetilde{w}_i > 0$. Their angles θ_i obey $\sin\theta_i \le \rho$ and $0 \le \theta_i < \pi/2$. Hence

$$\theta_i \le \arcsin\rho \le \frac{\pi}{2}\rho, \qquad \|v_i - \widetilde{w}_i\|^2 = 2(1-\cos\theta_i) \le 2\sin^2\theta_i \le 2\rho^2.$$

The Gaussian halfspace disagreement identity and the union bound give

$$\mathbb{P}\big(\operatorname{sign}(V^\top X) \ne \operatorname{sign}(\widetilde{W}^\top X)\big) \le \sum_i \frac{\theta_i}{\pi} \le \frac{k\rho}{2}.$$

Let $\widetilde{c}$ be the induced Boolean rule after relabeling, so that $h_{c,W} = h_{\widetilde{c},\widetilde{W}}$. By the triangle inequality,

$$\mathbb{P}\big(b(\operatorname{sign}(V^\top X)) \ne \widetilde{c}(\operatorname{sign}(V^\top X))\big) \le \Delta + \frac{kA_K\Delta}{2} < \gamma_K\left(1 + \frac{KA_K}{2}\right) < 2^{-K}.$$

The vector $\operatorname{sign}(V^\top X)$ is uniform on $\{-1,1\}^k$. Two distinct Boolean tables therefore disagree with probability at least $2^{-k} \ge 2^{-K}$. The displayed strict inequality forces $\widetilde{c} = b$. Summing the squared column-distance estimates proves the Frobenius bound. When the original rules coincide, a relabeling that preserves the induced table is a symmetry of b, proving the final claim. □

Corollary C.4 (Explicit gap between distinct rule orbits)**.** *Among all essential Boolean rules of rank at most K, including the two rank-zero constants, distinct signed-permutation equivalence classes have Gaussian disagreement at least γ_K, uniformly over their frames and ambient dimension.*

Proof. For two nonconstant rules this is the contrapositive of Theorem C.3. An essential nonconstant rank-k rule takes both values on at least one orthant, so its disagreement with either constant is at least $2^{-k} \ge 2^{-K} \ge \gamma_K$. The two constants disagree everywhere. □

Corollary C.5 (Explicit global disagreement metric)**.** *For every essential rank-k Boolean rule b, with $1 \le k \le K$, and every pair of orthonormal frames,*

$$c_K r_b(V,W) \le \mathbb{P}(h_{b,V}(X) \ne h_{b,W}(X)) \le \frac{\sqrt{k}}{2}\, r_b(V,W).$$

The lower bound needs neither a local facet expansion nor a compactness argument.

Proof. Write $r = r_b(V,W)$ and Δ for disagreement. If $\Delta \ge \gamma_K$, use $r \le 2\sqrt{k} \le 2\sqrt{K}$ to obtain $\Delta \ge \gamma_K r/(2\sqrt{K}) = c_K r$. If $\Delta < \gamma_K$, Theorem C.3 gives $r \le \sqrt{2k}A_K\Delta$. This implies $\Delta \ge c_K r$, since

$$c_K\sqrt{2k}A_K \le \frac{\gamma_K A_K}{\sqrt{2}} = \frac{2^{-K}}{4\sqrt{2}K} < 1.$$

For the upper bound, align a minimizing symmetry and let θ_i be the angles of corresponding columns. A change of the Boolean output requires at least one constituent sign change. Since $\theta_i/\pi \le \|v_i - w_i\|/2$ for every angle in $[0,\pi]$, the union bound and Cauchy–Schwarz give the claimed estimate. □

C.2 A uniform negative conditional allocation margin

Given $T = V^\top X$, put $q_W(T) = \mathbb{P}(D_{V,W} \mid T)$. The mass-only program in Section 3 has the explicit value

$$J_0(V, W; \eta) := J_0(V^\top W; \eta) = \mathbb{E}[2\min\{\eta, q_W(T)\} - q_W(T)].$$

Lemma C.6 (Uniform negative allocation margin)**.** *Put $\eta_K = 2^{-3K}/(2^{14}K^4)$. For every pair of frames for the same essential rule of rank at most K,*

$$J_0(V, W; \eta_K) \le -\kappa_K r, \qquad \kappa_K = c_K/2.$$

Proof. Set $L_K = \sqrt{8\log(32K/c_K)}$. For $r \le 1/4$, align a minimizing symmetry and write $w_i = \cos\theta_i v_i + \sin\theta_i n_i$. Then $\tan\theta_i \le 2r$. With $X_i = v_i \cdot X$ and $Z_i = n_i \cdot X$, a sign flip requires $|X_i| \le 2r|Z_i|$; X_i, Z_i are independent standard normals. The candidate-measurable band $B = \{\min_i |X_i| \le L_K r\}$ therefore satisfies

$$\mathbb{P}(B) \le 2K\varphi(0)L_K r, \qquad \mathbb{P}(D \setminus B) \le 2Kre^{-L_K^2/8} = c_K r/16.$$

For example the second inequality follows by integrating the one-sided flip interval conditional on Z_i and using $\mathbb{E}[|G|\mathbf{1}\{|G| > s\}] = 2\varphi(s)$, then summing over i. For integer $K \ge 1$, $\log K \le (K-1)\log 2$ gives $L_K^2 \le 112(\log 2)K < 81K$. Thus

$$\eta_K \le \frac{c_K}{16KL_K}, \quad \eta_K \le c_K/16, \quad \eta_K \le \gamma_K/8.$$

Write $q(T) = \mathbb{P}(D \mid V^\top X = T)$. Since B is measurable in T,

$$\begin{aligned} J_0 &= \mathbb{E}[2\min\{\eta_K, q(T)\} - q(T)] \\ &\le 2\eta_K \mathbb{P}(B) + 2\mathbb{P}(D \setminus B) - \mathbb{P}(D) \\ &\le (\varphi(0)/4 + 1/8 - 1)c_K r \le -c_K r/2. \end{aligned}$$

For $r \ge 1/4$, simply use $J_0 \le 2\eta_K - c_K r \le -c_K r/2$. □

For a single rule of rank k, take $K = k$, $\eta_b = \eta_K$, $c_b = c_K$, $C_b = \sqrt{K}$, and $\kappa = \kappa_K$. The bounds also hold at every smaller noise allowance.

An exact scalar certificate. Use the fixed allowance η_b in the protocol and let $\alpha_W(T) = \mathbf{1}\{q_W(T) > \eta_b\}$, $E_V = \mathbf{1}\{Y \ne f(V^\top X)\}$, and $A_W = \mathbb{E}[\alpha_W(T)(E_V - \eta_b)]$. For arbitrary stochastic labels, the pointwise dual inequality gives

$$R(h_V) - R(h_W) \le J_0(V, W; \eta_b) + 2A_W \le -\kappa r + 2A_W. \tag{C.2}$$

Indeed $eD \le \alpha e + (D - \alpha)_+$ for every $e \in [0, 1]$, and the chosen α attains the mass-only dual. In addition,

$$\mathbb{P}(\alpha_W = 1) \le \mathbb{P}(D_{V,W})/\eta_b \le C_b r/\eta_b. \tag{C.3}$$

Thus the bounded scalar observation in A_W has variance $O_b(r)$, under arbitrary labels as well as under honest ones.

C.3 Finite protocol and sample accounting

Fix the candidate frame V. Let $m = \min\{k, d-k\}$. Any competitor has the form $W = VA + NB$, where N is an orthonormal residual m-frame, $A \in \mathbb{R}^{k\times k}$, $B \in \mathbb{R}^{m\times k}$, and $A^\top A + B^\top B = I_k$; extend a smaller residual span as needed. For $m = 0$ omit N, B. Set $M = \begin{pmatrix} A \\ B \end{pmatrix}$ and

$$r(M) = \min_{Q\in\mathcal{S}_b} \left\| M - \begin{pmatrix} Q \\ 0 \end{pmatrix} \right\|_F .$$

This equals $r(V, W)$ and is independent of N. Conditional on T, the disagreement probability is the known fixed-dimensional integral

$$q_M(T) = \mathbb{P}_{G\sim N(0,I_m)}\{f(A^\top T + B^\top G) \neq f(T)\}.$$

Thus α_M and the scalar test do not require a residual orientation.

Enlarge C_b to at least one and choose $r_0 = \varepsilon/(4C_b)$. Cover $[r_0, 2\sqrt{k}]$ by dyadic annuli $[r_j, 2r_j]$. For each nonempty annulus choose once and for all a maximal νr_j-separated net of its feasible coefficient frames. It has size at most $|\mathcal{S}_b|(1 + 8/\nu)^{2k^2}$: every point lies in one of finitely many balls of radius $2r_j$ about the symmetry frames, so disjoint packing balls of radius $\nu r_j/2$ fit in slightly larger balls. The same fixed net is used by both parties. No continuity of α_M is assumed.

For each net point at scale r_j, estimate A_M to accuracy $\kappa r_j/8$ and require its estimate to be at most $\kappa r_j/4$. Equation (C.3) and Bernstein's inequality give this accuracy with $O_b(r_j^{-1}\log(JF_b/\delta))$ examples, where $J = O_b(\log(e/\varepsilon))$ and F_b is the per-annulus net bound. The observations can be reused over the finite net at a scale; a union bound needs no independence between those tests. Independent batches over scales have total size $O_b(\varepsilon^{-1}\log(e/(\varepsilon\delta)))$.

On their simultaneous accuracy event, acceptance implies $A_M \le 3\kappa r_j/8$, so (C.2) gives gain at most $-\kappa r_j/4$ for every tested competitor orientation. Rounding a competitor in that annulus keeps its residual orientation N and changes its frame by at most νr_j. The union bound on constituent halfspace disagreements changes risk by at most $C_b\nu r_j$. Choose $\nu \le \kappa/(8C_b)$. Every competitor with $r \ge r_0$ then has nonpositive gain. For $r < r_0$, Corollary C.5 bounds gain by $C_b r_0 \le \varepsilon/4$. This proves soundness. All verifier samples are drawn after the entire message, so the argument remains valid for every maliciously chosen candidate frame.

Put $q_* = \kappa r_0/16$. The honest prover uses a Frobenius net of the Stiefel manifold at mesh $q_*/(4C_b)$, where C_b is the constituent-halfspace disagreement constant. Its cardinality is at most $(C_b'/\varepsilon)^{dk}$ for a rule-dependent C_b', and some net point has risk at most $R(h_*) + q_*/4$. Hoeffding uniform convergence to tolerance $q_*/8$ uses $O_b((d\log(e/\varepsilon) + \log(1/\delta))/\varepsilon^2)$ samples. On this event an empirical risk minimizer has excess risk at most $q_*/2$. Since $R(h) - R(h_*) \ge (1 - 2\eta_b)\mathbb{P}(h \neq h_*)$ and $1 - 2\eta_b \ge 1/2$, its disagreement is at most q_*. The uniform-convergence event fails with probability at most $\delta/2$. For every feature-measurable $0 \le \alpha(T) \le 1$, the candidate's error rate obeys

$$\mathbb{E}[\alpha(T)(E_V - \eta_b)] \le \mathbb{P}(h_V \neq h_*) \le \kappa r_0/16.$$

Consequently every honest scalar estimate passes on its accuracy event, whose failure is at most $\delta/2$. Reject malformed frames. These are deterministic sample budgets, proving Theorem C.1.

C.4 The Boolean composition can also be unknown

Theorem C.7 (Quantitative verification of the unknown Boolean family)**.** *There is a universal $C > 0$ such that, for all integers $K, d \ge 1$, $0 < \varepsilon < 1$, and $0 < \delta \le 1/4$, the class $\mathcal{C}_{K,d}$ has a proper*

one-message verifier at noise allowance

$$\eta_K = \frac{2^{-3K}}{2^{14}K^4}$$

with deterministic budgets

$$m_V \le \frac{2^{CK}}{\varepsilon}\log\frac{e}{\varepsilon\delta}, \qquad m_P \le \frac{2^{CK}}{\varepsilon^2}\left(d\log\frac{e}{\varepsilon}+\log\frac{1}{\delta}\right).$$

Completeness permits full ambient Massart noise; soundness permits arbitrary labels and compares against the entire union. Both the Boolean table and the orthonormal frame are unknown and supplied in the message. There is no runtime claim.

Proof. Use $C_K = \sqrt{K}$ and the constants from (C.1). Delete inessential coordinates and fix one representative of each signed-permutation orbit of the Boolean rules of ranks at most K. There are at most $2+\sum_{k=1}^{K} 2^{2^k}$ representatives, so the logarithm of their number is $O(2^K)$. Candidates use this fixed list. For a nonconstant candidate of rank k, define $T = V^\top X$ and $E_V = \mathbf{1}\{Y \neq h_V(X)\}$. At every feasible coefficient frame $M = (A^\top, B^\top)^\top$ with $A^\top A + B^\top B = I_k$, let

$$q_M(T) = \mathbb{P}_G\{b(\operatorname{sign}(A^\top T + B^\top G)) \neq b(\operatorname{sign} T)\}, \quad \alpha_M(T) = \mathbf{1}\{q_M(T) > \eta_K\}.$$

Here G has dimension $m = \min(k, d-k)$; zero rows can be padded or omitted. For any residual orthonormal frame N, the competitor is $W = VA + NB$, and its quotient distance $r(M)$ is independent of N. The exact scalar dual gives, for all labels and all such N,

$$R(h_V) - R(h_W) \le -\kappa_K r(M) + 2A_M, \quad A_M = \mathbb{E}[\alpha_M(T)(E_V - \eta_K)].$$

Moreover $\operatorname{Var}(\alpha_M(T)(E_V - \eta_K)) \le \mathbb{P}(\alpha_M = 1) \le C_K r(M)/\eta_K$.

Set $r_0 = \varepsilon/(4C_K)$ and $\nu = \kappa_K/(8C_K)$. On each nonempty dyadic annulus $r_j \le r(M) \le 2r_j$, from r_0 through $2\sqrt{k}$, take a maximal νr_j-separated net inside the feasible set. Each annulus lies in $|\mathcal{S}_b|$ balls of radius $2r_j$ in Euclidean dimension at most $2K^2$. Consequently its net size F obeys

$$F \le 2^K K!\,(1+8/\nu)^{2K^2}, \qquad \log F = O(K^3).$$

The number of scales is $J = O(\log(eK/\varepsilon))$. At each scale use an independent batch, reused across its coefficient net, to estimate all A_M to tolerance $\kappa_K r_j/8$. Accept only if each estimate is at most $\kappa_K r_j/4$. Bernstein and a union bound permit a batch of size at most

$$C\left(\frac{C_K}{\eta_K\kappa_K^2}+\frac{1}{\kappa_K}\right)\frac{1}{r_j}\log\frac{CJF}{\delta},$$

with integer rounding absorbed. Since $\sum_j r_j^{-1} \le 2/r_0$, $\eta_K^{-1}, \kappa_K^{-1} \le 2^{O(K)}$, and $\log(JF/\delta) \le O(K^3\log(e/(\varepsilon\delta)))$, the total is at most $2^{CK}\varepsilon^{-1}\log(e/(\varepsilon\delta))$.

On the concentration event, accepted net competitors have gain at most $-\kappa_K r_j/4$. Rounding a competitor keeps its residual orientation and changes its predictions with probability at most $C_K\nu r_j \le \kappa_K r_j/8$. Thus all competitors at distance at least r_0 have nonpositive gain; closer ones have gain at most $C_K r_0 = \varepsilon/4$. A constant candidate needs no within-orbit test.

One further batch estimates the candidate risk within $\gamma_K/16$; accept only if its estimate is at most $\eta_K + \gamma_K/8$. This costs $O(\gamma_K^{-2}\log(1/\delta)) \le 2^{O(K)}\log(1/\delta)$. For every cross-orbit competitor, $R(h_V) - R(h_W) \le 2R(h_V) - \gamma_K$. On acceptance this is at most $2\eta_K - 5\gamma_K/8 < 0$. This proves soundness

against the entire union. Since all private samples are drawn after the message, conditioning on any malicious candidate justifies the concentration bounds without a union bound over candidate messages.

For completeness put $q_* = \min\{\kappa_K r_0/16, \gamma_K/16\}$. Use, for every representative rule of rank $k \le \min(K, d)$, a finite Stiefel net of Frobenius mesh $q_*/(4C_K)$. A volume packing argument in the radius-$\sqrt{k}$ ball of $\mathbb{R}^{dk}$ gives

$$\log N \le C\{2^K + dK \log(CK/q_*)\}.$$

The union contains a candidate with target disagreement at most $q_*/4$. Uniform empirical risk estimation within $q_*/8$ followed by empirical minimization gives excess risk at most $q_*/2$. Full ambient Massart noise implies excess risk at least $(1 - 2\eta_K)$ times target disagreement. Since $\eta_K \le 1/4$, the selected candidate has target disagreement at most q_*. The required prover budget is

$$Cq_*^{-2}\left(2^K + dK\log(CK/q_*) + \log(1/\delta)\right) \le \frac{2^{CK}}{\varepsilon^2}\left(d\log(e/\varepsilon) + \log(1/\delta)\right).$$

Constants can be enlarged universally to cover all $K \ge 1$. The orbit gap forces this candidate into the target's orbit. For every $0 \le \alpha \le 1$, $\mathbb{E}[\alpha(E_V - \eta_K)] \le q_* \le \kappa_K r_0/16$, so all within-orbit tests pass. Its risk is at most $\eta_K + q_* \le \eta_K + \gamma_K/16$, so the global test also passes. Allocate failure probability $\delta/3$ to learning, within-orbit concentration, and the risk test. All budgets are fixed in advance. □

C.5 Growing rank and necessary noise decay

At fixed accuracy and confidence, $K = o(\log d)$ gives $m_V = d^{o(1)}$, $m_P = d^{1+o(1)}$, and $\eta_K = d^{-o(1)}$. We next show that exponential noise decay in rank is necessary in order. At independent overlap $A = 0$, let $p = \mathbb{P}(f(G) = -1) \in (0, 1/2]$.

Proposition C.8 (Independent-overlap obstruction)**.** *For every such rule,*

$$J_0(0;\eta) = 2\big[(1-p)\min\{\eta, p\} + p\eta - p(1-p)\big].$$

It is positive if and only if $\eta > p(1-p)$. If the minority event $S = \{f = -1\}$ is centrally symmetric, then $J_1(0;\eta) = J_0(0;\eta)$.

Proof. On an anchor slice with $f(U) = +1$, disagreement is S and has mass p. On a slice with $f(U) = -1$, it is S^c and has mass $1 - p$. The scalar objective is $\min\{\eta, p\}$ in the first case and η in the second, because $\eta < 1/2 \le 1 - p$. Averaging these values and subtracting the unconditional disagreement $2p(1-p)$ gives the formula. For $\eta \le p$ it simplifies to $2[\eta - p(1-p)]$, and for $\eta \ge p$ it simplifies to $2p\eta$; the assertion about positivity follows.

If S is centrally symmetric and $\eta \le p$, the allocation $(\eta/p)\mathbf{1}_S$ attains the first slice's scalar optimum. If $\eta > p$, use $\mathbf{1}_S + ((\eta - p)/(1-p))\mathbf{1}_{S^c}$ instead. On the second slice use $(\eta/(1-p))\mathbf{1}_{S^c}$. Each allocation lies in $[0, 1]$, has mass η, is even, and hence has zero Gaussian first moments. Therefore the first-moment program attains the same optimum. □

Corollary C.9 (Noise ceilings for the whole orthogonal Boolean family)**.** *Let $\mathcal{C}_{K,d}$ consist of all Boolean combinations of at most K orthogonal homogeneous halfspaces. Any positive noise allowance supporting dimension-independent verification of this entire family at every accuracy must obey*

$$\eta_K \le 2^{-K}(1 - 2^{-K}).$$

For $K \geq 2$, if

$$\eta > 2^{1-K}(1 - 2^{1-K}),$$

then this entire family has $\Theta_{K,\eta,\varepsilon}(d)$ private verifier sample complexity at constant confidence and sufficiently small fixed accuracy.

Proof. For the first claim take a rule with minority event a single orthant, of mass $p = 2^{-K}$. Proposition C.8 gives a positive J_0 above the claimed ceiling, so Theorem B.1 rules out dimension-independent verification at sufficiently small fixed accuracy. For the second take the essential rule whose minority event consists of the all-positive and all-negative orthants. This centrally symmetric event has mass $p = 2^{1-K}$, so J_1 is positive at the stated noise level and the linear lower bound applies. In both cases the completeness null and the improving comparators belong to $\mathcal{C}_{K,d}$, so the lower bound remains valid when soundness compares against the whole union. Ordinary finite-class/net or VC learning gives the matching linear upper bound when all other parameters are fixed. □

The upper and lower noise exponents are not matched. These lower bounds fix K before choosing accuracy; their positive witness can vanish as K grows. They do not assert a linear lower bound at fixed accuracy for growing rank.

C.6 Optimal inverse-accuracy dependence

The following elementary two-target argument supplies the accuracy lower bound for Theorem 4.1.

Proposition C.10 (Accuracy lower bound from noiseless halfspaces)**.** *Let $d \geq 2$ and $0 < \varepsilon, \delta < 1/2$. For any comparator class containing all homogeneous halfspaces, every (ε, δ) verifier that is complete and sound on noiseless Gaussian halfspace laws has worst-case sample count*

$$m_V \geq \frac{\log(1/(2\delta))}{-\log(1 - 2\varepsilon)}.$$

The bound allows arbitrary interaction and improper outputs. For $0 < \varepsilon, \delta \leq 1/4$, it implies $m_V \geq (8\varepsilon)^{-1} \log(1/\delta)$. In particular it applies to $\mathcal{C}_{K,d}$ for every $K \geq 1$, even when completeness is required only at one noiseless halfspace law.

Proof. Fix $h_0(x) = \operatorname{sign}(x_1)$ and any $p \in (2\varepsilon, 1)$. Let h_1 have normal $\cos(\pi p)e_1 + \sin(\pi p)e_2$, so $\mathbb{P}(h_0(X) \neq h_1(X)) = p$. Under P_i, features are $N(0, I_d)$ and $Y = h_i(X)$. Both laws have comparator optimum zero. For every measurable Boolean predictor h,

$$R_{P_0}(h) + R_{P_1}(h) \geq p > 2\varepsilon.$$

Consequently no output, proper or improper, is accurate for both laws.

Fix an honest prover strategy for P_0. Run it in both worlds by simulating its independent samples from the known P_0; in the P_1 world this is an admissible malicious prover. Couple the two verifier executions using identical feature streams, verifier coins, and simulated-prover samples and coins. Pad an adaptive sample schedule to its worst-case budget $m = m_V$. With probability $(1-p)^m$, none of these m features falls in the disagreement region. On this event all observed labels, all interactive messages, and the final output coincide.

Completeness under P_0 fails with probability at most δ. Thus with probability at least $(1-p)^m - \delta$, this common output is nonrejecting and has P_0-risk at most ε. It then has P_1-risk greater than ε. Soundness under P_1 bounds this probability by δ, so $(1-p)^m \leq 2\delta$. Taking logarithms and letting $p \downarrow 2\varepsilon$ proves the exact bound. If $\varepsilon, \delta \leq 1/4$, then $-\log(1 - 2\varepsilon) \leq 4\varepsilon$ and $\log(1/(2\delta)) \geq \frac{1}{2}\log(1/\delta)$, giving the stated simpler bound. □

C.7 An exact balanced quadratic obstruction

For $d \geq 2$ and $t \geq 0$, define

$$f_t(s,u) = \operatorname{sign}(s)\operatorname{sign}(u-t), \qquad \mathcal{Q}_{t,d} = \{x \mapsto f_t(v\cdot x, w\cdot x) : \|v\| = \|w\| = 1,\ v \perp w\}.$$

Proposition C.11 (Exact conditional envelopes)**.** *For every $0 < \eta < 1/2$, set $p = \eta/2$ and $t = \Phi^{-1}(1-p) > 0$. Then*

$$J_0^{f_t}(\operatorname{diag}(1,0);\eta) = J_1^{f_t}(\operatorname{diag}(1,0);\eta) = \eta^2.$$

Consequently, at constant confidence and every fixed $0 < \varepsilon \leq \eta^2/8$, verification of $\mathcal{Q}_{t,d}$ requires $\Omega_{\eta,\varepsilon}(d)$ private observations, even with interaction and improper outputs. Ordinary proper learning gives a matching $O_{\eta,\varepsilon}(d)$ upper bound.

Proof. At overlap $A = \operatorname{diag}(1,0)$, the sign gate $\operatorname{sign}(s)$ is shared, and the second competing coordinate is an independent standard normal G. The first innovation coordinate is unused. Conditional on an anchor coordinate $u \leq t$, disagreement is $D = \mathbf{1}\{G > t\}$, of probability p. Choose

$$e_-(G) = \mathbf{1}\{|G| > t\}.$$

This allocation has mass $2p = \eta$, has zero first moment by symmetry, and attains objective $\mathbb{E}[e_- D] = p = \min\{\eta, p\}$. For $u > t$, disagreement is $D = \mathbf{1}\{G \leq t\}$, of probability $1-p$. Choose

$$e_+(G) = \frac{\eta}{1-2p}\mathbf{1}\{|G| \leq t\}.$$

It has mass η, is even, and takes values in $[0,1]$ because $\eta/(1-2p) = \eta/(1-\eta) < 1$. Its support is contained in the disagreement event, so its objective is $\eta = \min\{\eta, 1-p\}$. Both allocations are independent of the unused innovation coordinate, so the entire first-moment vector vanishes. They attain the mass-only optima in both anchor cases. Averaging over u therefore gives

$$\begin{aligned} J_1^{f_t}(A;\eta) = J_0^{f_t}(A;\eta) &= 2\{(1-p)p + p\eta\} - 2p(1-p) \\ &= 2p\eta = \eta^2. \end{aligned}$$

Theorem B.1 applies with this positive witness and $\varepsilon \leq \eta^2/8$, giving the unrestricted lower bound. For the upper bound one may use the finite-net learner from Theorem A.1; alternatively, the VC bound below gives ordinary proper agnostic learning with $O(d)$ observations at fixed accuracy and confidence. □

Lemma C.12 (Matched structural parameters)**.** *For every fixed $t \geq 0$, every member of $\mathcal{Q}_{t,d}$ is balanced under the standard Gaussian law, has minimum projection rank two, and has minimum polynomial threshold degree two. Moreover,* $\operatorname{VCdim}(\mathcal{Q}_{t,d}) = \Theta(d)$ *as $d \to \infty$. The known noisy anchor and every alternative in the preceding lower-bound construction also have balanced labels.*

Proof. Independence of $v\cdot X$ and $w\cdot X$ gives $\mathbb{E} f_t(v\cdot X, w\cdot X) = 0$ for every t. The two affine boundary hyperplanes

$$H_1 = \{v\cdot x = 0\}, \qquad H_2 = \{w\cdot x = t\}$$

each contain a relatively open facet across which the prediction changes. If the predictor were a function of a projection onto a subspace L, translation by any direction in $L^\perp$ would preserve it almost everywhere. Such a direction must be tangent to each active hyperplane: otherwise

a sufficiently small translation moves one of its facets and changes the prediction on a positive-volume set away from the other hyperplane. Hence $L^\perp \subseteq v^\perp \cap w^\perp$, so $\dim L \geq 2$. The displayed representation attains rank two.

The predictor is the sign of the degree-two polynomial $(v \cdot x)(w \cdot x - t)$. Any nonzero polynomial whose sign agrees almost everywhere with the predictor must vanish on both active facets, by continuity and the change of sign on their two sides. Its restriction to each hyperplane is therefore identically zero, since a polynomial vanishing on a relatively open set vanishes identically. Thus it is divisible by both distinct affine linear factors. The factors are relatively prime, so its degree is at least two. This proves the minimum degree assertion.

The number of labelings of n points by two affine halfspaces, and hence by their parity, is at most the square of the halfspace growth bound. For $n \geq d+1$ this is at most $(en/(d+1))^{2(d+1)}$, yielding VC dimension $O(d)$. For the reverse inequality fix $v = e_1$ and $x_i = e_1 + Re_i$, $2 \leq i \leq d-1$. The gate is positive on all these points. For $t > 0$, choose $w_i = 0$ or $2t/R$ independently, according to the desired negative or positive label, and take $R > 2t\sqrt{d}$. Set $w_1 = 0$ and use coordinate d to make w a unit vector. For $t = 0$, instead choose $w_i = \pm 1/R$ and $R > \sqrt{d}$, again normalizing in coordinate d. These choices shatter $d-2$ points and satisfy $w \perp v$.

For label balance, instantiate the random-frame construction in Appendix B with exactly the displayed allocations e_-, e_+. Write the known-null regression as $(1-2\eta)\operatorname{sign}(s)\operatorname{sign}(u-t)$. Its mean is zero by the independent sign gate. Every constructed alternative replaces η by $e_-(G)$ or $e_+(G)$ according to u, independently of s. The alternative regression is still $\operatorname{sign}(s)$ times a function of independent coordinates and therefore also has mean zero. □

Proof of Theorem 4.3. Apply Theorem C.1 to the essential two-bit parity rule and let $\eta_* > 0$ be its noise allowance, reduced to at most $1/4$. For every fixed $0 < \eta \leq \eta_*$, its protocol remains complete at noise bound η and retains the stated dimension-independent rate. Proposition C.11 supplies the shifted class at the same noise bound and its matching linear rate. Lemma C.12 proves all the matched structural assertions. The offset $t = \Phi^{-1}(1-\eta/2)$ is fixed before dimension and accuracy vary. □

Corollary C.13 (The class allowing one offset)**.** *Let $\mathcal{Q}_d = \bigcup_{t\in\mathbb{R}} \mathcal{Q}_{t,d}$, where the same definition of $\mathcal{Q}_{t,d}$ is used for negative t. For every fixed $0 < \eta < 1/2$, every fixed $0 < \varepsilon \leq \eta^2/8$, and constant confidence, the private verifier complexity of $\mathcal{Q}_d$ is $\Theta_{\eta,\varepsilon}(d)$, with unrestricted interaction and improper outputs allowed in the lower bound.*

Proof. For the lower bound take the fixed offset $t = \Phi^{-1}(1-\eta/2)$ from Proposition C.11. The known completeness null and each improving comparator in its lower-bound construction belong to $\mathcal{Q}_d$. Enlarging the comparator class can only strengthen the soundness requirement: $\mathrm{OPT}(\mathcal{Q}_d) \leq \mathrm{OPT}(\mathcal{Q}_{t,d})$. Thus the same null-simulation and validation proof gives the claimed lower bound. For the upper bound, the growth bound for parity of two affine halfspaces applies to the entire offset union and gives VC dimension $O(d)$. Ordinary proper agnostic learning on the verifier's sample therefore uses $O_\varepsilon(d + \log(1/\delta))$ observations. Every rule remains balanced by independence of its homogeneous gate. This proves Corollary 4.4 as well. □

D Halfspace geometry and a localized dual certificate

Let ϕ and Φ denote the standard Gaussian density and distribution function, and write $\overline{\Phi} = 1 - \Phi$. Throughout this section $G \sim N(0,1)$ and

$$A = \sqrt{2\log 2}, \qquad \phi_0 = \phi(0), \qquad a_c = 2\Phi(A) - 1.$$

We derive a conditional moment certificate for halfspaces. Its essential property is a strictly negative angular margin that persists after revealing any number of orthogonal Gaussian coordinates. The dual coefficients are bounded, and their nonconstant part is localized near the candidate's boundary.

D.1 The signed conditional moment problem

For $0 < \eta < 1/2$ and every real u, define

$$M_\eta(u) = \sup\left\{\mathbb{E}[e(G)\mathbf{1}\{G < -u\}] : 0 \le e \le 1,\ \mathbb{E}e(G) = \eta,\ \mathbb{E}[Ge(G)] = 0\right\}. \tag{D.1}$$

The supremum is over measurable functions on $\mathbb{R}$. Put

$$b_\eta = \Phi^{-1}\left(\frac{1+\eta}{2}\right), \qquad k_\eta = \Phi^{-1}(1-\eta/2).$$

Since $\eta < 1/2$, one has $0 < b_\eta < k_\eta$.

Lemma D.1 (Signed optimizer and bounded dual)**.** *If $u \le -b_\eta$, then $M_\eta(u) = \eta$; if $u \ge k_\eta$, then $M_\eta(u) = \overline{\Phi}(u)$. For $-b_\eta < u < k_\eta$, there are unique positive roots $\ell, z > |u|$ satisfying*

$$\phi(\ell) + \phi(z) = \phi(u), \qquad \overline{\Phi}(u) - \overline{\Phi}(\ell) + \overline{\Phi}(z) = \eta. \tag{D.2}$$

In this interval an optimizer and the optimal value are

$$e_u^*(g) = \mathbf{1}\{-\ell < g < -u\} + \mathbf{1}\{g > z\}, \qquad M_\eta(u) = \overline{\Phi}(u) - \overline{\Phi}(\ell).$$

Define dual coefficients by

$$(\alpha_\eta(u), \gamma_\eta(u)) = \begin{cases} (1,0), & u \le -b_\eta, \\ (z/(\ell+z), -1/(\ell+z)), & -b_\eta < u < k_\eta, \\ (0,0), & u \ge k_\eta. \end{cases}$$

The optimizers and coefficients may be chosen jointly measurable in (η, u, g). For every real x, u and $e \in [0,1]$,

$$e\mathbf{1}\{x < -u\} \le \left(\mathbf{1}\{x < -u\} - \alpha_\eta(u) - \gamma_\eta(u)x\right)_+ + \alpha_\eta(u)e + \gamma_\eta(u)xe. \tag{D.3}$$

Equality holds after expectation at the optimizer, so

$$M_\eta(u) = \mathbb{E}\left(\mathbf{1}\{G < -u\} - \alpha_\eta(u) - \gamma_\eta(u)G\right)_+ + \eta\alpha_\eta(u). \tag{D.4}$$

Moreover

$$0 \le \alpha_\eta \le 1, \quad |\gamma_\eta| \le A^{-1}, \quad \{\alpha_\eta \ne 0\} \subset \{u < k_\eta\}, \quad \{\gamma_\eta \ne 0\} \subset \{|u| < k_\eta\}. \tag{D.5}$$

Proof. For $u \le -b_\eta$, choose $e_u^*(g) = \mathbf{1}\{|g| \le b_\eta\}$. Its mass is η, its first moment is zero, and it is supported inside $\{g < -u\}$ up to null sets. It attains the objective upper bound η. For $u \ge k_\eta$, choose $e_u^*(g) = \mathbf{1}\{|g| \ge k_\eta\}$. This has the required moments and includes all of $\{g < -u\}$, attaining the upper bound $\overline{\Phi}(u)$.

In the intervening interval, let ℓ increase from $|u|$ to infinity. The first equation in (D.2) determines $z > |u|$ uniquely and continuously; z decreases from infinity to $|u|$. Hence $\overline{\Phi}(u) - \overline{\Phi}(\ell) + \overline{\Phi}(z)$ strictly increases from $\overline{\Phi}(u) - \overline{\Phi}(|u|)$ to $\overline{\Phi}(u) + \overline{\Phi}(|u|)$. For $u < 0$ these endpoints are $2\Phi(|u|) - 1 < \eta$ and

one. For $u \geq 0$ they are zero and $2\overline{\Phi}(u) > \eta$. There is a unique solution. The zeroth moment of the displayed indicator is the second root equation, and its first moment is $\phi(\ell) - \phi(u) + \phi(z) = 0$.

For these roots the function $c(g) = \mathbf{1}\{g < -u\} - \alpha_\eta(u) - \gamma_\eta(u)g$ is strictly positive precisely on $(-\ell, -u) \cup (z, \infty)$, apart from its endpoints. Indeed, on $g < -u$ it equals $(\ell + g)/(\ell + z)$, and on $g \geq -u$ it equals $(g - z)/(\ell + z)$. The inequality $ec \leq c_+$ proves (D.3) and, for every feasible function in (D.1), gives

$$\mathbb{E}[e(G)\mathbf{1}\{G < -u\}] = \mathbb{E}[e(G)c(G)] + \eta\alpha_\eta(u) \leq \mathbb{E}c(G)_+ + \eta\alpha_\eta(u).$$

The proposed indicator attains equality. The same statement holds for the two saturated cases with their specified coefficients.

The strictly monotone root equation and continuity of its data imply continuous dependence of the finite roots on (η, u) inside the open intermediate region. Defining the optimizers separately on the two closed saturated regions gives joint measurability everywhere. Finally $\phi(\ell) + \phi(z) = \phi(u) \leq \phi(0)$ implies $\max\{\ell, z\} \geq A$, since $\phi(A) = \phi(0)/2$. Thus $|\gamma| \leq A^{-1}$. The remaining coefficient bounds and support statements follow from their definitions and $b_\eta < k_\eta$. □

D.2 An exact threshold and a single crossing

For $u \geq 0$ define

$$F_\eta(u) = 2M_\eta(u) - \overline{\Phi}(u), \qquad J_\eta = \int_0^\infty F_\eta(u)\, du. \tag{D.6}$$

These integrals are finite because $|F_\eta(u)| \leq \overline{\Phi}(u)$ and $\int_0^\infty \overline{\Phi}(u)\, du = \phi_0$. The root representation also gives the explicit formula

$$J_\eta = \phi_0 - 2\int_0^{k_\eta} \overline{\Phi}(\ell_\eta(u))\, du. \tag{D.7}$$

Theorem D.2 (The conditional moment threshold)**.** *There is a unique $\eta_\dagger \in (0, 1/2)$ with $J_{\eta_\dagger} = 0$. Set $a_\dagger = 1 - 2\eta_\dagger$. Then $J_\eta < 0$ for $\eta < \eta_\dagger$ and $J_\eta > 0$ for $\eta > \eta_\dagger$. The constants satisfy*

$$\frac{1}{2} < 4\Phi(A) - 3 < a_\dagger < a_c = 2\Phi(A) - 1 < 1, \qquad A = \sqrt{2\log 2}. \tag{D.8}$$

In particular $\eta_\dagger < 1/4$.

Proof. For fixed $u \geq 0$ in the unsaturated region, increasing η strictly increases $\ell_\eta(u)$, by the monotonic root construction in Lemma D.1; thus it strictly increases $M_\eta(u)$. At saturation the value is $\overline{\Phi}(u)$. For any two values of η in $(0, 1/2)$, there is a positive interval of u on which both programs are unsaturated. Hence J_η is strictly increasing. The root formula gives pointwise continuity, including at saturation, and dominated convergence gives continuity of J. It also gives $J_\eta \longrightarrow -\phi_0$ as $\eta \downarrow 0$, since $M_\eta(u) \leq \eta$.

We locate a zero strictly between $\eta_c = \overline{\Phi}(A)$ and $\eta_s = 2\overline{\Phi}(A)$. Dropping the first-moment constraint yields $M_\eta(u) \leq \min\{\eta, \overline{\Phi}(u)\}$. If $v = \Phi^{-1}(1 - \eta)$, integration gives

$$\int_0^\infty [2\min\{\eta, \overline{\Phi}(u)\} - \overline{\Phi}(u)]\, du = 2\phi(v) - \phi_0.$$

For example, $\int_0^\infty \min\{\eta, \overline{\Phi}(u)\}\, du = \eta v + \int_v^\infty \overline{\Phi}(u)\, du = \phi(v)$. At $\eta = \eta_c$ the displayed upper bound is zero. For every $0 < u < A$, it is strictly greater than $F_{\eta_c}(u)$: attaining objective η_c would require all positive error mass to lie in $G < -u$, which gives a strictly negative first moment. The optimum is attained by Lemma D.1. Thus $J_{\eta_c} < 0$.

At $\eta = \eta_s$, the feasible symmetric indicator $e(g) = \mathbf{1}\{|g| \geq A\}$ has conditional objective $\overline{\Phi}(\max\{u, A\})$. Its integrated gain is zero, because

$$\int_0^\infty \overline{\Phi}(\max\{u, A\})\, du = A\overline{\Phi}(A) + \int_A^\infty \overline{\Phi}(u)\, du = \phi(A) = \phi_0/2.$$

For every $0 < u < A$, this symmetric indicator differs on a set of positive Gaussian measure from the unique maximizing indicator in Lemma D.1. More explicitly, it assigns value one sufficiently far in the negative tail, where the dual function c in that proof is strictly negative and the optimizer is zero. The dual inequality is consequently strict. Therefore $J_{\eta_s} > 0$.

For completeness we verify analytically that $\eta_s < 1/4$. The first two terms of $\log 2 = 2\sum_{j\geq 0}(1/3)^{2j+1}/(2j+1)$ give $\log 2 > 56/81$, hence $A > 7/6$. The third-order Taylor lower bound for e^{-x} gives

$$\int_0^{7/6} e^{-t^2/2}\, dt \geq \frac{7}{6} - \frac{(7/6)^3}{6} + \frac{(7/6)^5}{40} - \frac{(7/6)^7}{336} = \frac{63643027}{67184640} > \frac{947}{1000}.$$

Since $\pi < 22/7$, one has $\phi_0 > 199/500$. It follows that

$$\Phi(A) - \frac{1}{2} > \frac{947}{1000}\frac{199}{500} > \frac{3}{8}.$$

Thus $2\overline{\Phi}(A) < 1/4$. Continuity and strict increase now prove existence and uniqueness of a zero in (η_c, η_s), and conversion to signal parameters gives (D.8). □

Lemma D.3 (Single crossing)**.** *For $0 < \eta < 1/4$, the continuous function F_η on $[0, \infty)$ has a unique zero $u_\eta \in (0, k_\eta)$. It is negative before that zero and positive after it.*

Proof. At zero, $F_\eta(0) \leq 2\eta - 1/2 < 0$. For $u \geq k_\eta$, it equals $\overline{\Phi}(u) > 0$. As $u \uparrow k_\eta$, the roots satisfy $\ell \to \infty$ and $z \to k_\eta$, so the formulas join continuously. On $0 < u < k_\eta$, differentiating (D.2) gives

$$\phi(\ell)\ell' = \frac{u+z}{\ell+z}\phi(u), \qquad F_\eta'(u) = \phi(u)\frac{2u+z-\ell}{\ell+z}.$$

At a zero of F_η, one has $\overline{\Phi}(\ell) = \overline{\Phi}(u)/2$ and $\ell > u$. The Gaussian Mills ratio is strictly decreasing on $[0, \infty)$, as follows directly from

$$\frac{\overline{\Phi}(x)}{\phi(x)} = \int_0^\infty e^{-xt - t^2/2}\, dt.$$

Consequently $\phi(\ell) > \phi(u)/2$. The first root equation then implies $\phi(z) < \phi(\ell)$, so $z > \ell$ and $F_\eta'(u) > 0$. Every zero is therefore a strict crossing from negative to positive. Since the function starts negative and is positive at k_η, there is exactly one such zero. □

D.3 A uniform angular margin

Let T be standard Gaussian and let Z be an independent standard Gaussian vector. For $0 < \theta < \pi/2$, set

$$I_\eta(\theta) = \mathbb{E}F_\eta(|T|\cot\theta) = 2\tan\theta \int_0^\infty \phi(u\tan\theta)F_\eta(u)\, du. \tag{D.9}$$

If an error rule $e(T, Z) \in [0, 1]$ satisfies

$$\mathbb{E}[e \mid T] = \eta, \qquad \mathbb{E}[Ze \mid T] = 0, \tag{D.10}$$

interpret e as the conditional probability that the label differs from $\operatorname{sign}(T)$. Its gain against a competitor $\operatorname{sign}(T\cos\theta + n \cdot Z \sin\theta)$ is at most $I_\eta(\theta)$. Indeed, condition on $T = t \neq 0$ and $G = \operatorname{sign}(t) n \cdot Z$. The conditional expectation of e given this G is feasible for (D.1), while disagreement is $\{G < -|t| \cot\theta\}$. The paired-loss identity makes conditional gain at most $F_\eta(|t| \cot\theta)$, proving the claim after integration.

Lemma D.4 (Strict global angular margin)**.** *Fix $a \in (a_\dagger, 1)$ and $\eta = (1-a)/2$. There is $g_a > 0$ such that every error rule satisfying (D.10) has improvement at most $-g_a\theta$ against every halfspace making angle $\theta \in [0,\pi]$ with $\operatorname{sign}(T)$.*

Proof. Theorem D.2 gives $\eta < 1/4$ and $J_\eta < 0$. Let u_η be the crossing from Lemma D.3, and put

$$\bar\theta = \pi(3/4 - a/2) \in (0, \pi/2), \qquad g_a = \min\left\{\frac{a}{\pi}, 2\phi(u_\eta \tan\bar\theta)(-J_\eta), \frac{a-1/2}{2\pi}\right\} > 0.$$

For $0 < \theta \le \bar\theta$, the density $u \mapsto \phi(u\tan\theta)$ is decreasing. The single crossing implies

$$\int_0^\infty \phi(u\tan\theta) F_\eta(u)\, du \le \phi(u_\eta \tan\theta) J_\eta \le \phi(u_\eta \tan\bar\theta) J_\eta.$$

Use (D.9) and $\tan\theta \ge \theta$ to obtain $I_\eta(\theta) \le -g_a\theta$. For $\theta \ge \bar\theta$, Gaussian rotational invariance gives disagreement probability θ/π: in the two-dimensional span of the normals the direction is uniform, and the disagreement wedges have total angle 2θ. Since $\mathbb{E}e = \eta$, the gain is at most

$$\mathbb{E}[(2e-1)\mathbf{1}\{\text{disagreement}\}] \le 2\eta - \theta/\pi \le -(a-1/2)/2 \le -g_a\theta.$$

Zero angle gives zero gain, completing the proof. □

D.4 Revealed coordinates and the certificate

Let $X \sim N(0, I_d)$, choose a unit candidate normal v, and let S be a matrix with orthonormal columns perpendicular to v. Write

$$T = v \cdot X, \qquad U = S^\top X, \qquad Z = (I - vv^\top - SS^\top)X.$$

These Gaussian components are independent. For a unit coefficient tuple $\lambda = (c, \beta, \sigma)$ with $\sigma \ge 0$, and a residual unit n when $\sigma > 0$, put

$$h_0 = \operatorname{sign}(T), \qquad h_w = \operatorname{sign}(cT + \beta \cdot U + \sigma n \cdot X), \qquad \theta = \arccos c.$$

Let $D_\lambda = \mathbf{1}\{h_0 \neq h_w\}$. For $\sigma > 0$ define

$$u_\lambda = \frac{\operatorname{sign}(T)(cT + \beta \cdot U)}{\sigma}, \quad \alpha_\lambda = \alpha_\eta(u_\lambda), \quad \gamma_\lambda = \gamma_\eta(u_\lambda), \quad N = n \cdot X.$$

If $\sigma = 0$, instead set $\alpha_\lambda = D_\lambda$, $\gamma_\lambda = 0$; the residual direction is unnecessary.

Theorem D.5 (Gaussian baseline and risk certificate)**.** *Fix $a \in (a_\dagger, 1)$ and $\eta = (1-a)/2$. For every dimension, candidate, revealed frame, coefficient tuple, and residual direction as above, define for $\sigma > 0$*

$$H_\lambda = 2(D_\lambda - \alpha_\lambda - \gamma_\lambda \operatorname{sign}(T) N)_+ + 2\eta\alpha_\lambda - D_\lambda.$$

For $\sigma = 0$, define $H_\lambda = -aD_\lambda$. Then, with g_a from Lemma D.4,

$$\mathbb{E}H_\lambda \le -g_a\theta \qquad \text{for every } \theta \in [0,\pi]. \tag{D.11}$$

For any stochastic labeling let $E_0 = \mathbf{1}\{Y \neq h_0(X)\}$ *and*

$$A_\lambda = \mathbb{E}[\alpha_\lambda(E_0 - \eta)], \qquad V_\lambda = \mathbb{E}[\gamma_\lambda \operatorname{sign}(T) Z E_0].$$

Then

$$R(h_0) - R(h_w) \leq -g_a\theta + 2A_\lambda + 2\|V_\lambda\|. \tag{D.12}$$

The signed coefficients obey the pointwise dual inequality (D.3)*, the value identity* (D.4)*, and the localization bounds* (D.5)*. All constants are independent of the revealed frame dimension.*

Proof. Suppose $\sigma > 0$. Conditional on (T, U), $\operatorname{sign}(T)N$ is standard Gaussian and $D_\lambda = \mathbf{1}\{\operatorname{sign}(T)N < -u_\lambda\}$ up to null sets. By (D.4),

$$\mathbb{E}[H_\lambda \mid T, U] = 2M_\eta(u_\lambda) - \overline{\Phi}(u_\lambda).$$

We bound its expectation by lifting an optimizer. Let e_u^* be the jointly measurable optimizer of Lemma D.1 and define

$$e(T, U, Z) = e^*_{u_\lambda(T,U)}(\operatorname{sign}(T) n \cdot Z).$$

Conditional on (T, U) this error rule has mean η and first moment zero in every residual coordinate. The latter follows in direction n from the optimizer's first-moment constraint and in all other residual directions from Gaussian independence. Integrating over U gives

$$\mathbb{E}[e \mid T] = \eta, \qquad \mathbb{E}[Ue \mid T] = \eta\mathbb{E}[U \mid T] = 0, \qquad \mathbb{E}[Ze \mid T] = 0.$$

Thus it obeys (D.10) in the full space transverse to v. Its improvement against h_w equals $\mathbb{E}H_\lambda$. Lemma D.4 proves (D.11). If $\sigma = 0$, the baseline is exactly $-a\mathbb{P}(h_0 \neq h_w) = -a\theta/\pi \leq -g_a\theta$.

For arbitrary labels, the paired-loss identity is $R(h_0) - R(h_w) = \mathbb{E}[(2E_0 - 1)D_\lambda]$. Applying the pointwise dual (D.3) with $e = E_0$ and $x = \operatorname{sign}(T)N$ gives

$$R(h_0) - R(h_w) \leq \mathbb{E}H_\lambda + 2A_\lambda + 2n \cdot V_\lambda \leq -g_a\theta + 2A_\lambda + 2\|V_\lambda\|.$$

For $\sigma = 0$, $V_\lambda = 0$ and the same identity becomes $R(h_0) - R(h_w) = -a\theta/\pi + 2A_\lambda$. The remaining assertions were proved in Lemma D.1. □

D.5 The complete halfspace dimension classification

Proof of Corollary 3.2. Every rank-one overlap is $A = \cos\theta$ for $\theta \in [0, \pi]$. For $0 < \eta < \eta_c$, Lemma E.3 gives $J_0(\cos\theta; \eta) < 0$ when $0 < \theta \leq \pi/4$. At larger angles, the conditional mass constraint and Gaussian halfspace disagreement formula give

$$J_0(\cos\theta; \eta) \leq 2\eta - \theta/\pi \leq 2\eta - 1/4 < 0,$$

since $\eta_c < 1/8$, as proved in Appendix E. At $\theta = 0$ the envelope is zero. Thus every scalar envelope is nonpositive below η_c. Likewise, Lemma D.4 gives nonpositive first-moment envelopes at every angle whenever $0 < \eta < \eta_\dagger$. The 2-Lipschitz continuity in the noise parameter from Section A.8 extends these respective inequalities to η_c and $\eta_\dagger$, for every overlap. Theorem A.1 gives the dimension-independent and square-root upper bounds, including those endpoints, and the linear upper bound for all remaining noise levels.

For the converse, (H.2) gives a strictly positive small-angle scalar slope for every $\eta > \eta_c$. Hence some $J_0(\cos\theta; \eta) > 0$. For the first-moment envelope, (D.9) and dominated convergence give

$$\lim_{\theta \downarrow 0} \frac{J_1(\cos\theta; \eta)}{\theta} = 2\phi(0)J_\eta.$$

Here $|F_\eta(u)| \le \overline{\Phi}(u)$ supplies an integrable majorant. Theorem D.2 makes this slope positive exactly when $\eta > \eta_\dagger$ and proves $\eta_c < \eta_\dagger < 2\eta_c$. Apply Theorem B.1 to each positive witness: for sufficiently small fixed accuracy depending on η, it gives respectively $\Omega(\sqrt{d})$ and $\Omega(d)$, allowing interaction and improper outputs. This proves all three regimes at constant confidence.

The endpoint step uses continuity of envelope values, not a uniform strict angular margin. It therefore gives no extension of the sharper inverse-accuracy upper bounds in Theorems E.1 and G.1 to their respective critical noise levels. □

E Dimension-independent halfspace verification at low noise

E.1 Model and constants

Let $X \sim N(0, I_d)$, $Y \in \{-1, 1\}$, and

$$h_v(x) = \operatorname{sign}\langle v, x\rangle, \qquad \mathcal{H}_d = \{h_v : v \in \mathbb{S}^{d-1}\}.$$

Write $R(h) = \mathbb{P}(h(X) \ne Y)$ and $\mathrm{OPT} = \inf_{h\in\mathcal{H}_d} R(h)$. For soundness, $r(x) = \mathbb{E}[Y \mid X = x]$ can be any measurable function in $[-1, 1]$.

Let Φ, φ denote the standard Gaussian distribution function and density. Define

$$A = \sqrt{2\log 2}, \qquad a_c = 2\Phi(A) - 1.$$

Fix a known constant $a_0 \in (a_c, 1)$, and set

$$\eta_0 = (1 - a_0)/2, \qquad k = \Phi^{-1}(1 - \eta_0), \tag{E.1}$$

$$J = \varphi(0) - 2\varphi(k) > 0, \qquad \kappa = 2\varphi(k)J > 0, \qquad g = a_0 - 3/4 > 0. \tag{E.2}$$

Here $k > A$ ensures $J > 0$, and $a_c > 3/4$ ensures $g > 0$. The last inequality has an elementary analytic certificate: $\log 2 > 2(1/3 + 1/81) = 56/81 > 49/72$ gives $A > 7/6$. Using $e^{-u} \ge 1 - u + u^2/2 - u^3/6$ and $\pi < 22/7$, with $b = 7/6$ and $P = b - b^3/6 + b^5/40 - b^7/336 = 63643027/67184640$,

$$a_c = \sqrt{2/\pi}\int_0^A e^{-x^2/2}\,dx > \sqrt{7/11}\,P > 3/4.$$

The final strict inequality is the rational comparison $(7/11)P^2 > 9/16$; the bound on $\log 2$ follows from its positive series $\log 2 = 2\sum_{j\ge 0} 3^{-(2j+1)}/(2j+1)$.

The completeness model consists of all distributions satisfying

$$r(x) = b(x)h_w(x), \qquad w \in \mathbb{S}^{d-1}, \qquad b(x) \in [a_0, 1] \text{ almost surely.} \tag{E.3}$$

The target direction and noise function are unknown. In particular, $b(x) \equiv a \ge a_0$ gives exact random classification noise. If a user supplies the stronger bound $a_0 = 1$, any fixed design parameter in $(a_c, 1)$ can instead be used, so excluding this endpoint from the formulas loses no realizable-case applicability.

Theorem E.1 (Globally sound high-signal verification)**.** *Fix $0 < \varepsilon, \delta < 1$ and $a_0 \in (a_c, 1)$. There is a one-message proper verifier using*

$$m_V = O_{a_0}\left(\frac{1}{\varepsilon}\log\frac{e}{\varepsilon\delta}\right)$$

private labeled samples such that:

1. *For every labeling distribution with marginal $N(0, I_d)$ and every prover, its probability of outputting a classifier with risk exceeding* $\mathrm{OPT} + \varepsilon$ *is at most* $\delta/2$.

2. *On every distribution satisfying* (E.3)*, a sample-only honest prover makes it accept with probability at least* $1 - \delta$*. The honest prover needs* $O_{a_0}((d + \log(1/\delta))/\varepsilon^2)$ *samples.*

The verifier evaluates one projection and a finite list of scalar band tests. No efficient honest learning procedure is asserted. The hidden constants need not remain bounded as $a_0 \downarrow a_c$.

E.2 An orientation-free lower bound from a band error count

Fix any proposed unit direction v, and let $T = \langle v, X \rangle$. For a competing unit direction u making angle $\theta \in (0, \pi/2)$ with v, define

$$q_\theta(t) = \Phi(-|t| \cot \theta), \qquad B_\theta = \{|T| \le k \tan \theta\}, \qquad p_\theta = \mathbb{P}(B_\theta).$$

The conditional disagreement probability of h_u, h_v given $T = t$ is $q_\theta(t)$, independently of the competitor's orientation within the subspace perpendicular to v. Unconditional disagreement is θ/π.

Lemma E.2 (A valid bound for arbitrary labels)**.** *For every stochastic labeling distribution with the known Gaussian marginal,*

$$R(h_u) - R(h_v) \ge \frac{\theta}{\pi} - 2\mathbb{E}[q_\theta(T)\mathbf{1}\{T \notin B_\theta\}] - 2\mathbb{P}(Y \ne h_v(X), B_\theta). \tag{E.4}$$

The right-hand side depends on the competitor only through θ.

Proof. Let $D = \{h_u(X) \ne h_v(X)\}$ and $E_v = \{Y \ne h_v(X)\}$. The exact paired-loss identity is

$$R(h_u) - R(h_v) = \mathbb{P}(D) - 2\mathbb{P}(D \cap E_v).$$

Inside B_θ, bound the intersection by $\mathbb{P}(E_v \cap B_\theta)$. Outside the band, bound it by $\mathbb{P}(D \setminus B_\theta)$, and condition on T. □

Lemma E.3 (Positive geometric margin)**.** *For* $0 < \theta \le \pi/4$*, define*

$$\begin{aligned} F_\theta &= \frac{\theta}{\pi} - 2\mathbb{E}[q_\theta(T)\mathbf{1}\{T \notin B_\theta\}] - 2\eta_0 p_\theta \\ &= \mathbb{E}[q_\theta(T) - 2\min\{q_\theta(T), \eta_0\}]. \end{aligned}$$

Then $F_\theta \ge \kappa\theta$.

Proof. The two formulas agree since $q_\theta(t) \ge \eta_0$ exactly when $|t| \le k \tan \theta$. Put

$$f(z) = \Phi(-z) - 2\min\{\Phi(-z), \eta_0\}, \qquad z \ge 0.$$

Its sign changes once, from nonnegative to nonpositive, at $s = \Phi^{-1}(1 - 2\eta_0) \in (0, k)$. This sign-change point is distinct from the split point k in the minimum. Direct integration gives

$$\int_0^\infty \Phi(-z)\, dz = \varphi(0), \qquad \int_0^\infty \min\{\Phi(-z), \eta_0\}\, dz = \varphi(k),$$

where the second equality uses $\int_k^\infty \Phi(-z)\, dz = \varphi(k) - k\eta_0$. Consequently $\int_0^\infty f(z)\, dz = J > 0$.

Changing variables $|T| = z\tan\theta$ yields

$$F_\theta = 2\tan\theta \int_0^\infty \varphi(z\tan\theta) f(z)\, dz.$$

The weight is decreasing in z. Before s it is at least $\varphi(s\tan\theta)$ and multiplies a nonnegative quantity; after s it is at most that value and multiplies a nonpositive quantity. Thus

$$F_\theta \geq 2\tan\theta\, \varphi(s\tan\theta) J \geq 2\theta\, \varphi(k) J = \kappa\theta.$$

The last step uses $\theta \leq \pi/4$, $\tan\theta \geq \theta$, and $s\tan\theta \leq k$. □

E.3 The finite-angle protocol

Let

$$M = \lceil 2/\varepsilon \rceil, \qquad h = \frac{\pi}{4M}, \qquad \theta_j = jh \quad (1 \leq j \leq M), \qquad L = \log\frac{4(M+1)}{\delta}.$$

Then

$$\pi\varepsilon/12 \leq h \leq \pi\varepsilon/8. \tag{E.5}$$

For the proposed direction v, write

$$E_j = \mathbb{P}(Y \neq h_v(X),\ |T| \leq k\tan\theta_j), \qquad p_j = \mathbb{P}(|G| \leq k\tan\theta_j).$$

The prover sends only v, before the verifier draws any private samples. The verifier rejects malformed directions. For a valid direction it forms empirical estimates $\widehat{R}_v, \widehat{E}_j$ on

$$n = \left\lceil L \max\left\{ \frac{256k/\kappa^2 + 16/(3\kappa)}{h}, \frac{128}{g^2} \right\} \right\rceil \tag{E.6}$$

fresh independent labeled observations. It accepts and outputs h_v precisely when all of

$$\widehat{R}_v \leq \eta_0 + g/8, \tag{E.7}$$

$$\widehat{E}_j \leq \eta_0 p_j + \kappa\theta_j/4 \quad (1 \leq j \leq M) \tag{E.8}$$

hold. The sample count in (E.6) has the order claimed in Theorem E.1 by (E.5).

Lemma E.4 (Simultaneous confidence event)**.** *Conditional on any prover message, with probability at least* $1 - \delta/2$*,*

$$|\widehat{R}_v - R(h_v)| \leq g/16, \qquad |\widehat{E}_j - E_j| \leq \kappa\theta_j/8 \quad \textit{for all } j. \tag{E.9}$$

Proof. Hoeffding bounds the global-test failure probability by $2\exp(-ng^2/128) \leq 2e^{-L}$. A band indicator has variance at most p_j. On $0 \leq \theta \leq \pi/4$, the inequalities $\tan\theta \leq 2\theta$ and $2\varphi(0) < 1$ give $p_j \leq 2k\theta_j$. Apply Bernstein with $t_j = \kappa\theta_j/8$:

$$\mathbb{P}(|\widehat{E}_j - E_j| > t_j) \leq 2\exp\left(-\frac{nt_j^2}{2(p_j + t_j/3)}\right).$$

The sufficient sample count for the exponent to be at least L is bounded by

$$\frac{2(p_j + t_j/3)L}{t_j^2} \leq \left(\frac{256k}{\kappa^2} + \frac{16}{3\kappa}\right)\frac{L}{\theta_j},$$

which is at most (E.6). A union bound over all $M+1$ estimates gives failure probability at most $\delta/2$. □

E.4 Global soundness

Soundness in Theorem E.1. Assume the verifier accepts and (E.9) holds. No assumption on the label regression is imposed. The band tests give $E_j \le \eta_0 p_j + 3\kappa\theta_j/8$. Combining Lemmas E.2 and E.3, for every halfspace h_u at angle exactly θ_j from v,

$$R(h_u) - R(h_v) \ge F_{\theta_j} - 3\kappa\theta_j/4 \ge \kappa\theta_j/4 \ge 0.$$

This conclusion holds simultaneously for every orientation at that angle; no net of perpendicular directions is required.

For a competing halfspace at angle $0 < \theta < \pi/4$, take the next grid angle $\theta_j \ge \theta$ and the corresponding normal in the same two-dimensional plane as u, v, on the same side of v. Its angle from u is at most h. Gaussian disagreement and the elementary risk Lipschitz bound therefore give

$$R(h_u) \ge R(h_{u_j}) - h/\pi \ge R(h_v) - \varepsilon/8.$$

The comparator with $\theta = 0$ is h_v itself. In dimension one, the only angles are zero and π, so the intermediate-angle step is unnecessary.

For every angle $\theta \ge \pi/4$, use the global test instead. It implies $R(h_v) \le \eta_0 + 3g/16 \le \eta_0 + g/4$, and the paired-risk identity gives

$$R(h_u) - R(h_v) \ge \theta/\pi - 2R(h_v) \ge 1/4 - 2\eta_0 - g/2 = g/2 > 0.$$

Taking an infimum over all competitors proves $R(h_v) \le \mathrm{OPT} + \varepsilon/8$. The probability of an unsound output is therefore at most the confidence-event failure probability $\delta/2$. □

E.5 Sample-only Massart completeness

Completeness in Theorem E.1. Assume (E.3). Put $\eta(x) = (1 - b(x))/2 \le \eta_0$. The Bayes halfspace is h_w, and for every proposed h_v,

$$R(h_v) - R(h_w) = \mathbb{E}[b(X)\mathbf{1}\{h_v(X) \ne h_w(X)\}]. \tag{E.10}$$

Let

$$\zeta = \frac{1}{16}\min\{g, \kappa h\}.$$

An honest sample-only prover uses agnostic VC learning to obtain a proper h_v with excess risk at most ζ with probability at least $1 - \delta/2$. This uses $O((d + \log(1/\delta))/\zeta^2) = O_{a_0}((d + \log(1/\delta))/\varepsilon^2)$ samples.

Condition on that success event. Globally, $R(h_v) \le R(h_w) + \zeta \le \eta_0 + \zeta$. For each candidate-dependent band, the conditional error identity gives

$$\begin{aligned} E_j &= \mathbb{E}\big[(\eta(X) + b(X)\mathbf{1}\{h_v(X) \ne h_w(X)\})\mathbf{1}\{|T| \le k\tan\theta_j\}\big] \\ &\le \eta_0 p_j + R(h_v) - R(h_w) \le \eta_0 p_j + \zeta. \end{aligned}$$

On the verifier confidence event, the global estimate is at most $\eta_0 + \zeta + g/16 \le \eta_0 + g/8$. Every band estimate is at most

$$\eta_0 p_j + \zeta + \kappa\theta_j/8 \le \eta_0 p_j + 3\kappa\theta_j/16 < \eta_0 p_j + \kappa\theta_j/4.$$

Thus all acceptance conditions pass. Combining the honest-prover and verifier failure probabilities gives completeness error at most δ. □

F A nonlinear geometric separation: centered slabs

Fix $t > 0$, put $p = 2\Phi(t) - 1$, and define

$$h_w(x) = 2\mathbf{1}\{|w \cdot x| \le t\} - 1, \qquad \mathcal{S}_{t,d} = \{h_w : w \in \mathbb{S}^{d-1}\}.$$

The feature law is $N(0, I_d)$. Completeness has Massart error at most $\eta < 1/2$ around an unknown member of this class; soundness has arbitrary labels, with excess risk measured against $\mathcal{S}_{t,d}$. The honest prover has independent samples. Lower bounds allow arbitrary interaction and improper outputs.

F.1 An explicit linear-dimensional obstruction

Consider anchor h_{e_1} and an orthogonal competitor h_v. Conditional on the anchor projection, competitor disagreement is even in the hidden Gaussian coordinate. Maximizing error mass on that event therefore automatically permits zero first moment. The zeroth- and first-moment envelopes agree at orthogonal overlap:

$$J_0(0) = J_1(0) = 2\{p\min(\eta, 1-p) + (1-p)\min(\eta, p)\} - 2p(1-p). \tag{F.1}$$

Indeed, when the anchor is positive the disagreement mass is $1 - p$; when negative it is p. In either case allocate the conditional error budget η to disagreement first, filling its complement if needed. The allocation is even, has mean η, and has zero Gaussian first moment. This proves equality, not merely a lower bound.

For $\eta \le \min(p, 1-p)$ the right side is $2\eta - 2p(1-p)$. For $p \le \eta \le 1-p$ it is $2\eta p$; the other case follows by exchanging the two labels. It follows that

$$J_1(0) > 0 \quad \Longleftrightarrow \quad \eta > p(1-p). \tag{F.2}$$

Theorem F.1. *If $\eta > p(1-p)$, centered-slab verification requires $\Omega_{t,\eta,\varepsilon}(d)$ private observations at every fixed sufficiently small positive accuracy and confidence at least $2/3$. Completeness is needed only at the single known random-classification-noise anchor $\mathbb{E}[Y \mid X] = (1-2\eta)h_{e_1}(X)$. More explicitly, writing Γ for the positive expression in (F.1), for $d \ge 2$ and $0 < \varepsilon \le \Gamma/8$,*

$$m_V + \left\lceil \frac{512 \log 24}{\Gamma^2} \right\rceil > \frac{d-1}{100}.$$

An ordinary proper learner gives an upper bound linear in d at fixed accuracy and confidence, so the dimension exponent is sharp.

Proof. Apply Theorem B.1 with $k = 1$, $j = 1$, using the explicit even allocation above. Its labeled Hermite-mixture proof and fresh relative-risk validation cover unrestricted interaction and improper outputs. Substituting $g = \Gamma/2$ gives the displayed constants. For the upper bound, Gaussian continuity of rotations supplies a finite normal net of size $(O_t(1/\varepsilon))^d$, as also proved quantitatively below. Uniform convergence and empirical risk minimization on that net use $O_t((d\log(e/\varepsilon) + \log(1/\delta))/\varepsilon^2)$ observations and arbitrary-label soundness, without needing a prover. □

Corollary F.2 (Identical noise and dimension, different verification cost)**.** *Fix any $0 < \eta < \eta_c = (1 - a_c)/2$, and choose the fixed slab width so that $p = \eta/2$. For every fixed $0 < \varepsilon \le \eta^2/8$ and constant confidence:*

comparator class under $N(0, I_d)$	*private verifier samples*
homogeneous halfspaces	$\widetilde{O}_\eta(1/\varepsilon)$
$\mathcal{S}_{t,d}$	$\Theta_{\eta,\varepsilon}(d)$.

Both guarantees have full Massart completeness at the same error bound η and arbitrary-label Gaussian soundness. Both classes have VC dimension $\Theta(d)$ and depend on a single projection.

Proof. The halfspace upper bound is Theorem E.1. For slabs, $p < \eta < 1 - p$ gives $\Gamma = 2\eta p = \eta^2$, so Theorem F.1 applies. For the VC assertion, n points induce an arrangement of $2n$ affine hyperplanes $w \cdot x_i = \pm t$ in parameter space. The resulting number of sign patterns is $(O(n/d))^d$ when $n \geq d$, giving VC dimension $O(d)$. To shatter $d-1$ points, take $x_i = Re_i$ with $R > 2t\sqrt{d}$. Set $w_i = 0$ for points to be inside the slab and $w_i = 2t/R$ for those outside, and use the last coordinate to make w a unit vector. The standard halfspace VC bound is d. □

The slab width in this corollary is fixed before dimension grows. The accuracy restriction scales with its positive gain. This is not a claim that a vanishingly rare slab forces linear complexity at an unchanged accuracy.

F.2 Every fixed slab also has a positive low-noise regime

For two normals identify w and $-w$, and write their angle as $\theta \in [0, \pi/2]$. With independent $T, G \sim N(0,1)$, the competing projection is $T' = T\cos\theta + G\sin\theta$. Let D_θ be slab disagreement and $B_s = \{\, ||T| - t| \leq s \,\}$.

Lemma F.3. *There are $c_t, C_t, b_t > 0$ such that, uniformly for $0 < \theta \leq \pi/2$,*

$$c_t\theta \leq \mathbb{P}(D_\theta) \leq C_t\theta, \qquad \mathbb{P}(B_{b_t\theta}) \leq 4b_t\varphi(0)\theta, \qquad \mathbb{P}(D_\theta \setminus B_{b_t\theta}) \leq c_t\theta/16.$$

Proof. For $0 < \theta \leq 1/2$, disagreement implies, with $s = ||T| - t|$,

$$s \leq |T' - T| \leq (1 - \cos\theta)(t + s) + \sin\theta|G|, \qquad s \leq C\theta(t + |G|).$$

Conditioning on G and using the bounded density of T gives the upper bound $C_t\theta$. For a sufficiently small θ, take $T \in [t - \theta, t]$ and $G \geq 3 + t$. Then the anchor is inside and $T' > t$, giving $\mathbb{P}(D_\theta) \geq c_t'\theta$. For angles bounded away from zero, disagreement is positive and continuous: the two nonidentical slabs differ on an open set, and their boundaries have Gaussian measure zero. Compactness completes both angle bounds.

The band bound is immediate from its total interval length at most $4b_t\theta$. For $\theta \leq 1/2$, the first display gives

$$\mathbb{P}(D_\theta \setminus B_{b\theta}) \leq C\theta\mathbb{E}[(t + |G|)\mathbf{1}\{t + |G| > b/C\}],$$

whose coefficient tends to zero as $b \to \infty$. For $\theta \geq 1/2$ choose $b \geq 2t$, so outside the band $|T| > t + b\theta \geq t + b/2$. Its probability tends uniformly to zero and can be made at most $c_t/32 \leq c_t\theta/16$. Choose b_t large enough for both cases. □

Theorem F.4. *For each fixed $t > 0$ there is $\eta_t > 0$ such that every $0 \leq \eta \leq \eta_t$ admits a one-message proper slab verifier with full Massart completeness and arbitrary-label soundness using*

$$m_V = O_t\left(\varepsilon^{-1}\log\frac{e}{\varepsilon\delta}\right), \qquad m_P = O_t\left(\frac{d\log(e/\varepsilon) + \log(e/(\varepsilon\delta))}{\varepsilon^2}\right).$$

Proof. Put $C_B = 8b_t\varphi(0)$ and choose $\eta_t \leq \min\{1/4, c_t/(16C_B)\}$. For a candidate v, put $E_v = \mathbf{1}\{Y \neq h_v(X)\}$. On an annulus $r \leq \theta \leq 2r$, use $B = B_{2b_tr}$ and $A_r = \mathbb{E}[(E_v - \eta)\mathbf{1}_B]$. The paired-loss identity and Lemma F.3 give, for arbitrary labels,

$$\begin{aligned} R(h_v) - R(h_w) &\leq 2A_r + 2\eta\mathbb{P}(B) + 2\mathbb{P}(D_\theta \setminus B) - \mathbb{P}(D_\theta) \\ &\leq 2A_r + 2\eta C_B r - 7c_t\theta/8 \leq 2A_r - 3c_t r/4. \end{aligned}$$

Take dyadic annuli from $r_0 = \varepsilon/(4C_t)$ to $\pi/2$, adjusting constants for large accuracy. Estimate each A_r within $c_t r/64$ and require its estimate to be at most $c_t r/32$. On simultaneous accuracy, every annular competitor has negative gain. Smaller angles change predictions on mass at most $C_t r_0 \le \varepsilon/4$.

The scalar observation is bounded and has variance at most $\mathbb{P}(B) \le C_B r$. Bernstein and a union bound therefore use $O_t(r_0^{-1} \log(e/(\varepsilon\delta)))$ observations; the same private batch can serve all annuli. An honest independent proper ERM on a fine normal net gives disagreement $q \le c_t r_0/64$ from the Massart target with the displayed sample cost. Outside this disagreement its conditional error is at most η, hence $A_r \le q$ for every annulus, and all tests pass. Divide the failure budget between the candidate and private tests. This proves both guarantees. □

The orthogonal calculation is an explicit obstruction, not a claim that orthogonal overlap maximizes either envelope. The full critical curves for centered slabs are not determined here.

F.3 Additive dimension and accuracy costs

We now allow d and accuracy to vary jointly, keeping the slab width $t > 0$ and noise allowance fixed. Write $q(h, g) = \mathbb{P}_X(h(X) \ne g(X))$ for prediction disagreement. As above, $p = 2\Phi(t) - 1$ and $\eta_c = \overline{\Phi}(\sqrt{2\log 2})$.

Theorem F.5 (Additive joint complexity)**.** *Fix $t > 0$ and $0 < \eta < \eta_c$. There are positive constants $C_{t,\eta}$ and $\varepsilon_{t,\eta}$ such that for every $d \ge 2$, $0 < \varepsilon \le \varepsilon_{t,\eta}$, and $0 < \delta \le 1/4$, a proper one-message verifier has full Massart completeness and arbitrary-label soundness with*

$$m_V \le C_{t,\eta}\left[d + \frac{1}{\varepsilon}\log\frac{e}{\varepsilon\delta}\right], \qquad m_P \le C_{t,\eta}\,\frac{d\log(e/\varepsilon) + \log(1/\delta)}{\varepsilon^2}.$$

The message is just a unit normal. If additionally

$$p(1-p) < \eta, \tag{F.3}$$

then, after decreasing $\varepsilon_{t,\eta}$ if necessary, every verifier in this model satisfies

$$m_V \ge \frac{1}{400}\left[d + \frac{1}{\varepsilon}\log\frac{1}{\delta}\right].$$

Under (F.3), at constant confidence the optimal rate is $\widetilde{\Theta}_{t,\eta}(d+\varepsilon^{-1})$, with only an accuracy logarithm separating the displayed bounds. At every fixed $0 < \eta < \eta_c$, choosing $p = \eta/2$ satisfies (F.3).

The upper bound does not claim that its d term is necessary for all widths or noise allowances. The lower bound uses the strict inequality (F.3). The width and noise allowance are fixed before dimension and accuracy vary. No endpoint claim at η_c is made.

F.4 Geometry and an exact local scalar margin

Normals w and $-w$ represent the same slab. The angle between two slabs is therefore $\theta = \arccos|v \cdot w| \in [0, \pi/2]$. Let T, G be independent standard normals and put $T_\theta = T\cos\theta + G\sin\theta$. Write

$$D_\theta = \mathbf{1}\{\mathbf{1}_{\{|T|\le t\}} \ne \mathbf{1}_{\{|T_\theta|\le t\}}\}, \quad q_\theta(T) = \mathbb{E}[D_\theta \mid T], \quad Q_t(\theta) = \mathbb{E}q_\theta(T).$$

For $0 < \theta \le \pi/2$, all these quantities are explicitly Gaussian-integrable. In particular, if

$$b_\theta(x) = \Phi\left(\frac{t - x\cos\theta}{\sin\theta}\right) - \Phi\left(\frac{-t - x\cos\theta}{\sin\theta}\right),$$

then $q_\theta(x) = 1 - b_\theta(x)$ on $|x| \le t$ and $q_\theta(x) = b_\theta(x)$ otherwise.

Lemma F.6 (Prediction metric)**.** *For $0 \le \theta \le \pi/2$, $Q_t(\theta) \le 2\theta/\pi \le \theta$. For unit normals v, w,*

$$q(h_v, h_w) \le \|v - w\|_2.$$

The function Q_t is continuous, is positive for $\theta > 0$, and satisfies $Q_t(\pi/2) = 2p(1-p)$.

Proof. In the plane spanned by two normals, condition on the Gaussian radius. Its angle is uniform. For each fixed radius, membership in one slab is the union of at most two circular arcs. Rotating these arcs through θ changes their indicator on angular measure at most 4θ: each of at most four endpoints sweeps an arc of length θ. Averaging proves $Q_t(\theta) \le 2\theta/\pi$. The same bound applies for an oriented angle $\alpha \in [0, \pi]$; since $2\alpha/\pi \le 2\sin(\alpha/2) = \|v - w\|_2$, it gives the second claim. Degenerate two-dimensional representations follow by a limit, or directly for equal and opposite normals.

Continuity follows from dominated convergence, since each boundary $|T_\theta| = t$ has probability zero. For $\theta > 0$, the event $|T| < t/2$ and sufficiently large positive G has positive probability and contains slab disagreement. At $\theta = \pi/2$ the two slab membership indicators are independent Bernoulli(p), giving the last identity. □

Define

$$H_\eta(q) = 2\min\{\eta, q\} - q, \qquad J_0(\theta; \eta) = \mathbb{E}H_\eta(q_\theta(T)).$$

The function H_η is 1-Lipschitz on $[0, 1]$, vanishes at zero, and satisfies $|H_\eta(q)| \le q$.

Lemma F.7 (Local slab envelope)**.** *For every fixed $t > 0$ and $0 < \eta < 1/2$, with $z_\eta = \Phi^{-1}(1 - \eta)$,*

$$\lim_{\theta \downarrow 0} \frac{J_0(\theta; \eta)}{\theta} = 4\varphi(t)\,[2\varphi(z_\eta) - \varphi(0)]. \tag{F.4}$$

Consequently, if $\eta < \eta_c$, there exist $\theta_0 \in (0, \pi/4]$ and $\kappa \in (0, 1]$, depending only on t, η, such that

$$J_0(\theta; \eta) \le -\kappa\theta \qquad (0 < \theta \le \theta_0). \tag{F.5}$$

Proof. Let $q_+(T)$ and $q_-(T)$ be the conditional probabilities of crossing the thresholds t and $-t$, respectively, between T and T_θ. Slab disagreement is the exclusive-or of those two crossing events. If their conditional intersection probability is $r(T)$, then $q_\theta = q_+ + q_- - 2r$, with $0 \le r \le \min(q_+, q_-)$. The Lipschitz and absolute bounds for H_η give

$$|H_\eta(q_\theta) - H_\eta(q_+) - H_\eta(q_-)| \le 2\min(q_+, q_-).$$

When $T \ge 0$ and $0 < \theta < \pi/2$, crossing the lower threshold has probability $\overline{\Phi}((t + T\cos\theta)/\sin\theta) \le \overline{\Phi}(t/\sin\theta)$. When $T < 0$, the analogous bound holds for crossing the upper threshold. By symmetry the two threshold envelopes agree. Writing $J_+(\theta) = \mathbb{E}H_\eta(q_+(T))$, we obtain

$$|J_0(\theta; \eta) - 2J_+(\theta)| \le 2\overline{\Phi}(t/\sin\theta) = o(\theta). \tag{F.6}$$

The last assertion follows, for example, from the Gaussian bound $\overline{\Phi}(x) \le \varphi(x)/x$ for $x > 0$.

Put $s = \tan\theta$ and $b = t\tan(\theta/2)$. On substituting $T = t + su$ on the upper side of the threshold and $T = t - su$ on the lower side, the conditional crossing probabilities are $\overline{\Phi}(u - b)$ and $\overline{\Phi}(u + b)$, respectively. Thus

$$\frac{J_+(\theta)}{s} = \int_0^\infty \big[\varphi(t + su) H_\eta(\overline{\Phi}(u - b)) + \varphi(t - su) H_\eta(\overline{\Phi}(u + b))\big]\, du.$$

For sufficiently small θ, $0 \le b \le 1$. The absolute value of the integrand is then at most $2\varphi(0)\overline{\Phi}(u-1)$, an integrable function on $[0,\infty)$. Dominated convergence, $s/\theta \to 1$, and (F.6) therefore show that the limit in (F.4) equals

$$4\varphi(t)\int_0^\infty H_\eta(\overline{\Phi}(u))\,du.$$

Finally,

$$\int_0^\infty \min\{\eta,\overline{\Phi}(u)\}\,du = \eta z_\eta + \int_{z_\eta}^\infty \overline{\Phi}(u)\,du = \varphi(z_\eta), \qquad \int_0^\infty \overline{\Phi}(u)\,du = \varphi(0).$$

These identities follow by integrating the normal density in the opposite order. They prove (F.4). Its right side is strictly negative precisely when $z_\eta > \sqrt{2\log 2}$, equivalently $\eta < \eta_c$. Taking, for example, any $0 < \kappa \le \min\{1, -2\varphi(t)[2\varphi(z_\eta) - \varphi(0)]\}$ and then a sufficiently small $\theta_0 \le \pi/4$ proves the margin. □

Lemma F.8 (Localized scalar certificate)**.** *Fix a candidate h_v, and let $T = v \cdot X$ and $E_v = \mathbf{1}\{Y \ne h_v(X)\}$. For $0 < \theta \le \theta_0$, set*

$$\alpha_\theta(T) = \mathbf{1}\{q_\theta(T) > \eta\}, \qquad A_\theta = \mathbb{E}[\alpha_\theta(T)(E_v - \eta)].$$

For every labeling and every slab h_w at angle θ from h_v,

$$R(h_v) - R(h_w) \le -\kappa\theta + 2A_\theta. \tag{F.7}$$

The summand defining A_θ is bounded in absolute value by one and has variance at most θ/η. If the labels satisfy Massart noise around h_*, then

$$A_\theta \le q(h_v, h_*) \qquad \text{for every } \theta. \tag{F.8}$$

Proof. For the disagreement indicator D between h_v, h_w, the paired risk identity is $R(h_v) - R(h_w) = \mathbb{E}[(2E_v - 1)D]$. For $e \in [0,1]$ and $\alpha \in \{0,1\}$, $eD \le (D-\alpha)_+ + \alpha e$. Conditional on T, Gaussian rotational invariance gives $\mathbb{E}[D \mid T] = q_\theta(T)$ independently of the competitor's residual direction. The chosen α_θ gives exactly

$$2\eta\alpha_\theta + 2\mathbb{E}[(D-\alpha_\theta)_+ \mid T] - q_\theta(T) = H_\eta(q_\theta(T)).$$

After integration this proves the certificate using Lemma F.7. Also $\mathbb{P}(\alpha_\theta = 1) \le \mathbb{E}q_\theta/\eta \le \theta/\eta$, which bounds the second moment and hence the variance of the summand. Under the honest promise, its conditional candidate error probability is at most $\eta + \mathbf{1}\{h_v(X) \ne h_*(X)\}$. Multiplication by the nonnegative α_θ proves (F.8). □

F.5 A finite-resolution check for distant competitors

The following elementary lemma isolates the part of the protocol that pays for ambient dimension. It applies to any known feature marginal and any Boolean hypothesis class, using disagreement q under that marginal. Its negative acceptance threshold is essential when the desired final accuracy tends to zero.

Lemma F.9 (Distant-competitor check)**.** *Fix $0 < a \le 1$, a candidate h, and $r \in (0,1]$. Suppose a finite set $\mathcal{N}$ is a disagreement ρ-net for all competitors, where $\rho \le ar/32$. Filter it to*

$$\mathcal{N}_h = \{g \in \mathcal{N} : q(h,g) \ge r/2\}.$$

Using

$$n_F = \left\lceil \frac{512}{a^2 r^2} \log \frac{2 \max\{1, |\mathcal{N}|\}}{\zeta} \right\rceil \tag{F.9}$$

independent labeled observations, estimate each paired gain $R(h) - R(g)$ and accept if all estimates are at most $-ar/4$. With probability at least $1 - \zeta$, acceptance implies $R(h) \le R(g)$ for every competitor with $q(h, g) \ge r$, for arbitrary labels.

For completeness, suppose there is a target h_* with Massart signal floor a, and $q(h, h_*) \le ar/[16(1 + a)]$. On the same simultaneous estimation event, every filtered test passes.

Proof. Each paired loss lies in $[-1, 1]$. Hoeffding and a union bound show that (F.9) estimates all filtered gains to error at most $ar/16$ with failure probability at most ζ. On this event an accepted filtered net point has true gain at most $-3ar/16$. For a competitor g with $q(h, g) \ge r$, choose $g' \in \mathcal{N}$ with $q(g, g') \le \rho$. Then $q(h, g') \ge r - \rho \ge r/2$, so it is filtered in, and

$$R(h) - R(g) \le R(h) - R(g') + q(g, g') \le -3ar/16 + ar/32 < 0.$$

For the honest claim, the Massart condition gives $R(g) - R(h_*) \ge a q(g, h_*)$ and $R(h) - R(h_*) \le q(h, h_*)$. By the disagreement triangle inequality,

$$R(h) - R(g) \le -a q(h, g) + (1 + a) q(h, h_*). \tag{F.10}$$

For every filtered g this is at most $-7ar/16$ under the assumed candidate accuracy. Adding estimation error $ar/16$ leaves an estimate at most $-3ar/8 < -ar/4$. □

The lemma is also a local/global gluing principle: a separate certificate may handle a neighborhood of the candidate, while this check handles its complement whenever every excluded competitor has disagreement at least a fixed $r > 0$. A fixed resolution net of size $\exp(O(d))$ then costs $O(d + \log(1/\zeta))$ observations. No ε-resolution ambient net is necessary.

F.6 The complete upper protocol

Fix θ_0, κ from Lemma F.7, and put

$$a = 1 - 2\eta, \qquad r_F = \min_{\theta \in [\theta_0, \pi/2]} Q_t(\theta) > 0, \qquad \rho_F = a r_F / 32.$$

These are fixed known constants depending on t, η. Take a Euclidean ρ_F-net of S^{d-1}, of cardinality at most $(1 + 2/\rho_F)^d$; the usual disjoint-ball volume argument proves this bound. Lemma F.6 turns it into a slab disagreement ρ_F-net. It will be used in Lemma F.9.

Choose

$$\beta = \min\left\{ \frac{1}{16}, \frac{\kappa}{256}, \frac{a r_F}{16(1 + a)} \right\} > 0, \qquad q_* = \beta\varepsilon. \tag{F.11}$$

For now require $0 < \varepsilon \le \min\{1/6, 8\theta_0\}$. The prover sends a unit normal v; invalid messages are rejected. All private data are drawn after this message.

Local tests. Put $r_0 = \varepsilon/8$ and $\nu = \min\{1/4, \kappa/8\}$. Cover $[r_0, \theta_0]$ by successive intervals

$$I_j = [r_j, \min\{2r_j, \theta_0\}], \qquad r_j = 2^j r_0,$$

stopping after the interval whose right endpoint is θ_0. If $r_0 = \theta_0$, use its singleton interval. In each interval, choose a grid including its endpoints with maximum gap at most νr_j. Let $\mathcal{G}$ be this finite

indexed collection; if the same angle occurs in two intervals its tests can be retained separately. Its size N satisfies

$$N \le (\lceil 1/\nu \rceil + 2) \max\{1, \lceil \log_2(\theta_0/r_0) \rceil\} = O_{t,\eta}(\log(e/\varepsilon)).$$

For every grid angle $\theta \in I_j$, estimate A_θ from one common private batch and accept its test only if

$$\widehat{A}_\theta \le \kappa r_j/8. \tag{F.12}$$

Use a batch of size

$$n_L = \left\lceil \frac{256(4/\eta + \kappa/12)}{\kappa^2 r_0} \log \frac{6N}{\delta} \right\rceil. \tag{F.13}$$

Distant tests. Independently use Lemma F.9 with $h = h_v$, $r = r_F$, the fixed net above, and $\zeta = \delta/3$. Its filtered set is determined after receiving v, using exact known-marginal disagreement. The verifier outputs h_v if and only if every local and distant test passes.

Lemma F.10 (Private guarantees)**.** *For every fixed candidate v, with probability at least $1 - 2\delta/3$ both of the following hold simultaneously:*

1. *If the verifier accepts, then $R(h_v) \le \mathrm{OPT} + \varepsilon/8$ under arbitrary labels.*
2. *If the labels satisfy Massart noise around h_* and $q(h_v, h_*) \le q_*$, all the tests pass.*

Proof. For $\theta \in I_j$, Lemma F.8 bounds the variance of the local summand by $2r_j/\eta$ and its centered absolute value by two. Bernstein's inequality at error tolerance $u_j = \kappa r_j/16$ gives failure probability at most

$$2\exp\left\{-\frac{n_L u_j^2}{4r_j/\eta + 4u_j/3}\right\} \le 2\exp\left\{-\frac{n_L \kappa^2 r_j}{256(4/\eta + \kappa/12)}\right\} \le \frac{\delta}{3N}.$$

Thus, except with probability $\delta/3$, every local estimate is accurate to u_j. The distant estimates have simultaneous failure probability at most $\delta/3$ by Lemma F.9. Work on the intersection of these events.

An accepted local test has $A_\theta \le 3\kappa r_j/16$. Its certificate therefore bounds the gain of every competitor at that grid angle by $-\kappa r_j + 3\kappa r_j/8 = -5\kappa r_j/8$. For any $\theta' \in I_j$, choose a grid angle within νr_j and keep the competitor's residual direction fixed while changing its angle. The resulting normals differ by that angle; their prediction disagreement is at most νr_j by Lemma F.6. Consequently the original competitor has gain at most $-5\kappa r_j/8 + \nu r_j \le -\kappa r_j/2 < 0$. Angles below r_0 have gain at most disagreement, hence at most $r_0 = \varepsilon/8$. Every angle at least θ_0 has disagreement at least r_F, and is controlled by the distant check. This proves soundness against all competing slabs, including cases where the infimum defining OPT is not attained.

For completeness, (F.8) and (F.11) give

$$A_\theta \le q_* \le \kappa\varepsilon/256 = \kappa r_0/32 \le \kappa r_j/32.$$

Adding estimation error $\kappa r_j/16$ leaves the estimate at most $3\kappa r_j/32 < \kappa r_j/8$, so every local test passes. Also $q_* \le a r_F/[16(1+a)]$ because $\varepsilon \le 1$. The distant tests therefore pass by Lemma F.9. □

Independent sample-only honest prover. Take a Euclidean net of S^{d-1} at radius $aq_*/4$ and choose an empirical risk minimizer on it using independent prover data. The net has at most $(1 + 8/(aq_*))^d$ elements. Hoeffding and a union bound ensure that every net risk is estimated to error $aq_*/8$ with probability at least $1 - \delta/3$ using

$$m_P = O_{t,\eta}\left(\frac{d\log(e/\varepsilon) + \log(1/\delta)}{\varepsilon^2}\right).$$

A net point within $aq_*/4$ of the target normal has risk at most $R(h_*) + aq_*/4$ by Lemma F.6. The empirical minimizer therefore has excess risk at most $aq_*/2$. Under Massart noise, excess risk is at least a times target disagreement, so its disagreement is at most $q_*/2$, in particular at most q_*. The target is Bayes, hence $\mathrm{OPT} = R(h_*)$.

Conditional on any sent normal, the verifier's independent private batches have the guarantees of Lemma F.10. For an arbitrary malicious normal, this proves soundness failure at most $2\delta/3 \le \delta$. For the honest prover, combining its failure budget $\delta/3$ with the private budget $2\delta/3$ gives completeness probability at least $1 - \delta$ and a correct nonreject output. The verifier uses one message and is proper. Finally (F.13) is $O_{t,\eta}(\varepsilon^{-1}\log(e/(\varepsilon\delta)))$, while (F.9) is $O_{t,\eta}(d + \log(1/\delta))$. Their sum proves the upper bound in Theorem F.5.

F.7 Unrestricted matching lower bounds

We give the two lower bounds separately. Their joint use is legitimate: a verifier for the stated completeness and soundness families must satisfy both, even though their honest null laws are different.

F.7.1 A linear dimension obstruction without validation cost

Let $a = 1 - 2\eta$ and define the positive gain

$$\Gamma = 2\{p\min(\eta, 1-p) + (1-p)\min(\eta, p) - p(1-p)\}. \tag{F.14}$$

It is positive exactly when $\eta > p(1-p)$. Indeed, when $\eta \le \min(p, 1-p)$ it is $2\eta - 2p(1-p)$; in the other two possible ranges with $\eta < 1/2$ it is respectively $2\eta p$ or $2\eta(1-p)$.

Fix $X = (T, Z) \sim N(0, I_d)$, $Z \in \mathbb{R}^D$, $D = d - 1$, and let $h_0(X) = 2\mathbf{1}\{|T| \le t\} - 1$. The known null Q has regression ah_0 and Bayes risk η. For $v \in S^{D-1}$, the improving comparator is $h_v(X) = 2\mathbf{1}\{|v \cdot Z| \le t\} - 1$. Given T, let $D_T(g)$ indicate disagreement between $h_0(T)$ and $2\mathbf{1}\{|g| \le t\} - 1$, and put $q_T = \mathbb{E}_G D_T(G)$, which equals $1 - p$ or p. Define

$$e_T(g) = \begin{cases} (\eta/q_T)D_T(g), & q_T \ge \eta, \\ D_T(g) + \dfrac{\eta - q_T}{1 - q_T}(1 - D_T(g)), & q_T < \eta. \end{cases}$$

These functions are even, lie in $[0, 1]$, have mean η and zero first Gaussian moment, and satisfy $\mathbb{E}[e_T(G)D_T(G)] = \min(\eta, q_T)$. Let P_v have regression

$$r_v(T, Z) = h_0(T)(1 - 2e_T(v \cdot Z)).$$

It has the promised feature marginal, while its labels need not satisfy Massart noise. Directly,

$$R_{P_v}(h_0) = \eta, \qquad R_{P_v}(h_0) - R_{P_v}(h_v) = \Gamma, \qquad \mathrm{OPT}_{P_v} \le \eta - \Gamma. \tag{F.15}$$

Lemma F.11 (Mixture bound)**.** *Let $\overline{P}_m = \mathbb{E}_v P_v^{\otimes m}$ for a uniform hidden direction. If $2m < D$, then*

$$1 + \chi^2(\overline{P}_m \| Q^{\otimes m}) \le (1 - 2m/D)^{-1/2}.$$

In particular $m \le D/100$ implies $\mathrm{TV}(\overline{P}_m, Q^{\otimes m}) < 1/6$.

Proof. The null gives both labels positive conditional probability. Writing $r_0 = ah_0$ and $\Delta_v = r_v - r_0$, the likelihood ratio is

$$\frac{dP_v}{dQ}(X, Y) = 1 + \frac{Y - r_0(X)}{1 - a^2}\Delta_v(X).$$

For two directions with inner product ρ, its cross moment is

$$K(v, w) = 1 + \mathbb{E}_T \mathbb{E}_Z[b_T(v \cdot Z) b_T(w \cdot Z)], \qquad b_T(g) = \frac{e_T(g) - \eta}{\sqrt{\eta(1 - \eta)}}.$$

The function b_T is even and centered, with squared Gaussian L^2 norm at most one, since $0 \le e_T \le 1$ and $\mathbb{E} e_T = \eta$. Expand it in normalized Gaussian Hermite polynomials. Only even degrees at least two occur. The Gaussian covariance identity $\mathbb{E}[H_j(G) H_\ell(H)] = \mathbf{1}_{j=\ell} \rho^j$ implies

$$0 \le \mathbb{E}[b_T(G) b_T(H)] \le \rho^2 \|b_T\|_2^2 \le \rho^2.$$

This identity follows by comparing coefficients in the joint Gaussian exponential generating function; approximation by finite Hermite sums gives it for L^2 functions. Consequently $1 \le K(v, w) \le 1 + \rho^2$. Independence conditional on the hidden direction gives

$$1 + \chi^2(\overline{P}_m \| Q^{\otimes m}) = \mathbb{E}_{v,w} K(v, w)^m \le \mathbb{E} e^{m\rho^2}.$$

For independent uniform sphere directions,

$$\mathbb{E}\rho^{2j} = \frac{(2j - 1)!!}{D(D + 2) \cdots (D + 2j - 2)} \le \frac{(2j - 1)!!}{D^j}.$$

The formula follows by integrating a spherical coordinate, and is also valid for $D = 1$. Expanding the exponential proves $\mathbb{E} e^{m\rho^2} \le (1 - 2m/D)^{-1/2}$ for $2m < D$. If $m \le D/100$, the resulting chi-square divergence is at most $\sqrt{50/49} - 1 < 1/9$. The inequality $\mathrm{TV} \le \frac{1}{2}\sqrt{\chi^2}$ proves the stated conclusion. □

Lemma F.12 (Interactive dimension lower bound)**.** *Suppose $\Gamma > 0$, $0 < \delta \le 1/4$, and*

$$0 < \varepsilon \le \frac{a\Gamma}{2(1 + a)}.$$

Any verifier complete at Q and sound for every P_v, even with interaction and improper outputs, has $m_V > (d - 1)/100$.

Proof. Freeze an honest prover strategy for the known null Q and simulate its samples independently from Q in every execution. At Q this is honest; at P_v it is an allowed malicious strategy, which may react to verifier messages. With a worst-case sample bound m, the entire transcript and output are one common randomized function of m unknown-law observations and independent auxiliary randomness. Adaptive sample requests can be padded to m.

Consider the event that the verifier outputs h and $q(h, h_0) \le \varepsilon/a$. Under Q,

$$R_Q(h) - R_Q(h_0) = a q(h, h_0)$$

for every Boolean h, so completeness gives this event probability at least $1-\delta$. Under every P_v, such an output satisfies

$$R_{P_v}(h) \geq R_{P_v}(h_0) - q(h, h_0) \geq \eta - \varepsilon/a > \eta - \Gamma + \varepsilon \geq \mathrm{OPT}_{P_v} + \varepsilon.$$

The strict inequality follows from the displayed accuracy restriction, which leaves at least $\Gamma/2$ of slack. Soundness therefore bounds the event probability by δ, also under the mixture. The probability gap is at least $1-2\delta \geq 1/2$. Data processing contradicts Lemma F.11 if $m \leq (d-1)/100$. The event is measurable under the known feature law; it requires no additional observations from the unknown labeled distribution. Randomized predictors, if allowed, obey the same identities with risks and disagreement averaged over prediction randomness. □

F.7.2 The accuracy and confidence obstruction

Lemma F.13 (Noiseless two-target bound)**.** *If $d \geq 2$, $0 < \delta \leq 1/4$, and*

$$0 < \varepsilon \leq \min\{1/6, p(1-p)/4\},$$

then every verifier complete and sound on noiseless Gaussian slab laws satisfies

$$m_V \geq \frac{1}{12\varepsilon} \log \frac{1}{\delta}.$$

The statement permits arbitrary interaction and improper outputs.

Proof. By Lemma F.6, the disagreement along a planar rotation is continuous from zero to $2p(1-p)$. Thus there are two slab predictors h_0, h_1 with disagreement exactly 3ε. Let P_i be their noiseless labeled laws. Their comparator optima are zero, and every Boolean predictor h obeys

$$R_{P_0}(h) + R_{P_1}(h) \geq 3\varepsilon > 2\varepsilon.$$

Fix an honest strategy for P_0 and run its independently simulated P_0 data in both worlds. Couple the verifier's $m = m_V$ padded iid feature observations and all protocol coins. With probability $(1-3\varepsilon)^m$, no feature falls in the disagreement set; then every observation, interactive message, and output agrees. Intersecting this event with correct nonreject output under P_0 loses at most δ probability. Every such common output has P_1 risk greater than ε. Soundness under P_1 gives

$$(1-3\varepsilon)^m - \delta \leq \delta, \qquad m \geq \frac{\log(1/(2\delta))}{-\log(1-3\varepsilon)}.$$

No independence between correct output and the no-disagreement event was used. Since $\varepsilon \leq 1/6$ and $\delta \leq 1/4$, $-\log(1-3\varepsilon) \leq 6\varepsilon$ and $\log(1/(2\delta)) \geq \frac{1}{2}\log(1/\delta)$, proving the claim. □

Completion of Theorem F.5. The upper proof used only $\eta < \eta_c$ and $\varepsilon \leq \min\{1/6, 8\theta_0\}$. When (F.3) holds, take

$$\varepsilon_{t,\eta} = \min\left\{\frac{1}{6}, 8\theta_0, \frac{p(1-p)}{4}, \frac{a\Gamma}{2(1+a)}\right\} > 0.$$

The full completeness family contains both the constant-noise null of Lemma F.12 and the noiseless laws of Lemma F.13; soundness contains all their alternatives. Because $d-1 \geq d/2$, the two lemmas give

$$m_V > \frac{d}{200}, \qquad m_V \geq \frac{1}{12\varepsilon}\log\frac{1}{\delta}.$$

The maximum of these two bounds is at least half their sum, hence at least $(d + \varepsilon^{-1}\log(1/\delta))/400$. Finally, if $p = \eta/2$, then $p(1-p) = \eta(1-\eta/2)/2 < \eta$, so the asserted example satisfies the hard-regime condition. □

G Verification under an unknown smooth feature law

The intermediate verification saving survives a fixed, nonparametric neighborhood of the Gaussian defined by curvature bounds. The density is unknown and no density, score, or conditional-mean oracle is used. The observable certificate is the same raw-coordinate certificate used for standard Gaussians. Its soundness is stable because each baseline depends on at most three projections and is boundary localized. Its honest compression is stable because raw witnesses differ only slightly from exact conditional-mean innovations. Those innovations are used solely in the proof.

G.1 Distribution family and guarantee

For $0 \le \tau < 1$, let $\mathcal{D}_\tau$ be the centered isotropic laws on $\mathbb{R}^d$ with a positive density $Z^{-1}e^{-V(x)}$, $V \in C^2(\mathbb{R}^d)$, satisfying

$$(1-\tau)I_d \preceq \nabla^2 V(x) \preceq (1+\tau)I_d \quad \text{for every } x. \tag{G.1}$$

Both completeness and soundness promise this marginal family. Soundness permits every measurable labeling. Completeness permits every Massart labeling with signal floor a, namely $\mathbb{E}[Y \mid X] = b(X)\,\mathrm{sign}(w \cdot X)$ for some unit w and $a \le b(X) \le 1$. All observations are counted, including those whose labels are ignored. The prover has only independent samples.

Theorem G.1. *Let $a_\dagger$ be the threshold in Theorem D.2. For every fixed $a \in (a_\dagger, 1)$ there is $\tau_a > 0$, independent of d, ε, δ, with the following property. Uniformly over unknown marginals in $\mathcal{D}_{\tau_a}$, there is a one-message proper verifier with full Massart completeness, arbitrary-label soundness, and error at most δ, using*

$$m_V = O_a\!\left(\frac{\sqrt{d}}{\varepsilon}\log\frac{e}{\varepsilon\delta}\right), \qquad m_P = O_a\!\left(\frac{d\log(e/\varepsilon)+\log(e/(\varepsilon\delta))}{\varepsilon^2}\right).$$

It outputs a candidate whose risk is at most the best homogeneous halfspace risk plus ε, or rejects. The message consists of the candidate and $O(\log(e/\varepsilon))$ orthonormal frames, each of dimension at most a constant K_a. It contains no description of the unknown density or its conditional means. No computational or finite-bit efficiency claim is made.

For $a_\dagger < a < a_c$, Theorem H.4 gives a matching $\Omega_a(\sqrt{d}/\varepsilon)$ lower bound, up to logarithms, in its stated dimension–accuracy range. This follows because $N(0, I_d) \in \mathcal{D}_\tau$ for every $\tau \ge 0$. The curvature and isotropy conditions are promises of both the completeness and soundness classes. The verifier does not test these conditions. The theorem gives a sample bound; computational and finite-precision complexity are outside its statement.

We use the signed Gaussian dual and angular margin of Theorem D.5, together with the classical Brascamp–Lieb variance inequality $\mathrm{Var}_\mu f \le \mathbb{E}_\mu[\nabla f^\top (\nabla^2 V)^{-1} \nabla f]$; see [CCEL13, Equation (1.3)]. The remaining perturbation and compression arguments are proved below.

G.2 Low-dimensional densities and conditional drift

Fix throughout a sufficiently small universal $\tau_0 > 0$, for example $\tau_0 = 1/16$, and assume $\tau \le \tau_0$.

Lemma G.2 (Projected density comparison)**.** *For every orthonormal $d \times q$ matrix P, let p_P denote the density of $P^\top X$. For each fixed q, there are constants $C_q, c_q > 0$ independent of d, P, τ such that*

$$p_P(x) \le C_q e^{-c_q\|x\|^2}, \tag{G.2}$$

$$|p_P(x) - \phi_q(x)| \le C_q\tau(1+\|x\|^2)e^{-c_q\|x\|^2}, \tag{G.3}$$

where ϕ_q is the standard Gaussian density. In particular all constants may be made universal for $q \le 3$.

Proof. Write $X = (U, Z)$ in orthogonal coordinates, and write $p_P(u) = e^{-W(u)}$ with the normalizing constant included in W. Differentiating under the integral gives

$$\nabla^2 W(u) = \mathbb{E}[\nabla^2_{UU} V \mid U = u] - \mathrm{Cov}(\nabla_U V \mid U = u).$$

For a unit vector $b \in \mathbb{R}^q$, conditional Poincare and the cross-Hessian bound $\|\nabla^2_{ZU} V\|_{\mathrm{op}} \le \tau$ give

$$\mathrm{Var}(b \cdot \nabla_U V \mid U = u) \le \frac{\tau^2}{1-\tau}.$$

Consequently

$$\left(1 - \frac{\tau}{1-\tau}\right) I_q \preceq \nabla^2 W(u) \preceq (1+\tau) I_q, \qquad \|\nabla^2 W - I_q\|_{\mathrm{op}} \le 2\tau.$$

Gaussian tails justify the differentiations and integration by parts; the bounded Hessian gives at most linear score growth.

Since U is centered and isotropic, $\mathbb{E}\nabla W(U) = 0$ and Taylor's formula for the gradient imply

$$\|b_0\| := \|\nabla W(0)\| \le 2\tau \mathbb{E}\|U\| \le 2\tau\sqrt{q}.$$

Thus $W(x) - W(0) = \|x\|^2/2 + s(x)$, with

$$|s(x)| \le 2\tau\sqrt{q}\,\|x\| + \tau\|x\|^2.$$

Put $Z_q = \mathbb{E}_{G\sim N(0,I_q)} e^{-s(G)}$. The elementary inequality $|e^{-z} - 1| \le |z| e^{|z|}$ gives $|Z_q - 1| \le C_q \tau$, and Z_q is bounded above and below by positive q-dependent constants uniformly for $\tau \le \tau_0$. Since $p_P(x) = \phi_q(x) e^{-s(x)}/Z_q$, absorbing the linear term in a small multiple of $\|x\|^2$ proves both claims. All integrals involved are finite uniformly for the stated τ_0. □

Lemma G.3 (Conditional-mean drift)**.** *For $U = P^\top X \in \mathbb{R}^k$ and $R = (I - PP^\top)X$, let $m(u) = \mathbb{E}[R \mid U = u]$. Then*

$$\|Dm(u)\|_{\mathrm{op}} \le \frac{\tau}{1-\tau}, \qquad \|m(u)\| \le \frac{\tau}{1-\tau}(\|u\| + \sqrt{k}).$$

Proof. For residual unit n and $b \in \mathbb{R}^k$, differentiating the conditional density gives

$$n^\top Dm(u) b = -\,\mathrm{Cov}(n \cdot R, b \cdot \nabla_U V \mid U = u).$$

Conditional Poincare bounds the two variances by $(1-\tau)^{-1}$ and $\tau^2\|b\|^2/(1-\tau)$, respectively. Cauchy–Schwarz proves the derivative bound. Since $\mathbb{E}m(U) = \mathbb{E}R = 0$,

$$\|m(0)\| = \|\mathbb{E}[m(0) - m(U)]\| \le \frac{\tau}{1-\tau}\mathbb{E}\|U\| \le \frac{\tau\sqrt{k}}{1-\tau}.$$

The Lipschitz estimate proves the pointwise conclusion. □

We will also use a direct consequence of the $q = 2$ case of Lemma G.2: for unit v, w of angle θ,

$$\mathbb{P}(\mathrm{sign}(v \cdot X) \ne \mathrm{sign}(w \cdot X)) \le C_{\mathrm{ang}}\theta. \tag{G.4}$$

Indeed the disagreement set in their two-dimensional span is a pair of wedges of total angular measure 2θ; integrate $Ce^{-c\rho^2}\rho\,d\rho$ in polar coordinates. Degenerate collinear cases follow directly.

G.3 The observable certificate remains valid at every scale

Fix $a > a_\dagger$, put $\eta = (1-a)/2$, and let $\alpha(u), \gamma(u)$ be the signed Gaussian dual weights. For precision, they obey the pointwise inequality, for every $e \in [0,1]$ and $x, u \in \mathbb{R}$,

$$e\mathbf{1}\{x < -u\} \le (\mathbf{1}\{x < -u\} - \alpha(u) - \gamma(u)x)_+ + \alpha(u)e + \gamma(u)xe, \tag{G.5}$$

and

$$0 \le \alpha \le 1, \quad |\gamma| \le B = (2\log 2)^{-1/2}, \quad k_\eta = \Phi^{-1}(1-\eta/2),$$
$$\{\alpha \ne 0\} \subset \{u < k_\eta\}, \quad \{\gamma \ne 0\} \subset \{|u| < k_\eta\},$$

They are dual optimizers for maximizing $\mathbb{E}[e(G)\mathbf{1}\{G < -u\}]$ subject to $0 \le e \le 1$, $\mathbb{E}e = \eta$, $\mathbb{E}[Ge] = 0$.

Let $P = (v, S)$ be a proposed orthonormal frame, with $U = (T, S^\top X) = P^\top X$, $T = v \cdot X$, and $R = (I - PP^\top)X$. A coefficient tuple is $\lambda = (c, \beta, \sigma)$ with $c^2 + \|\beta\|^2 + \sigma^2 = 1$, $\sigma \ge 0$. For a residual unit n, the associated competitor is $h_w = \operatorname{sign}(cT + \beta \cdot S^\top X + \sigma n \cdot X)$; its Euclidean angle from v is $\theta = \arccos c$. For $\sigma > 0$ set

$$u_\lambda = \frac{\operatorname{sign}(T)(cT + \beta \cdot S^\top X)}{\sigma}, \quad E_v = \mathbf{1}\{Y \ne \operatorname{sign}(T)\},$$

$$A_\lambda = \mathbb{E}[\alpha(u_\lambda)(E_v - \eta)], \qquad V_\lambda = \mathbb{E}[\gamma(u_\lambda)\operatorname{sign}(T)RE_v].$$

When $\sigma = 0$, use $\alpha_\lambda = \mathbf{1}\{\operatorname{sign}(T)(cT + \beta \cdot S^\top X) < 0\}$ and $\gamma_\lambda = 0$. These conventions define A, V as well.

Lemma G.4 (Localized baseline perturbation)**.** *There are $g_a, C_a > 0$, independent of the frame dimension, such that every labeling and every tuple satisfy*

$$R(h_v) - R(h_w) \le -(g_a - C_a\tau)\theta + 2A_\lambda + 2\|V_\lambda\|. \tag{G.6}$$

Proof. Let $N = n \cdot X$, $D = \mathbf{1}\{h_v \ne h_w\}$, and set

$$H_\lambda(T, S^\top X, N) = 2(D - \alpha - \gamma\operatorname{sign}(T)N)_+ + 2\eta\alpha - D.$$

The paired-loss identity and (G.5) imply

$$R(h_v) - R(h_w) \le \mathbb{E}H_\lambda + 2A_\lambda + 2n \cdot V_\lambda.$$

For $\sigma = 0$, define $H_\lambda = -aD$; the same inequality then holds with equality before replacing $n \cdot V$ by its norm.

Under a standard Gaussian marginal Theorem D.5 gives $\mathbb{E}_G H_\lambda \le -g_a\theta$ for every angle in $[0, \pi]$. One way to see the uniformity in the frame is to lift the conditional Gaussian optimizers at each $(T, S^\top X)$: the lifted error rule has conditional error η given T and zero first moment in every transverse coordinate. The original one-dimensional envelope thus applies. Its strictly negative linear margin holds for acute angles up to $\bar\theta = \pi(3/4 - a/2)$; larger angles use $2\eta - \theta/\pi \le -(a - 1/2)/2$ and a smaller g_a.

Crucially, H_λ depends only on at most three orthogonal projections: T, $Q = (S\beta/\|\beta\|) \cdot X$ when $\beta \ne 0$, and N. For $0 < \theta \le 1/4$, $c \ge 1/2$, $\|\beta\|, \sigma \le \theta$, and

$$|H_\lambda| \le C_a(1 + |N|)\mathbf{1}\{|T| \le C_a\theta(1 + |Q| + |N|)\}.$$

To check the support, $D = 1$ forces $c|T| \le \|\beta\||Q| + \sigma|N|$, while $\alpha \neq 0$ or $\gamma \neq 0$ forces $c|T| \le \|\beta\||Q| + k_\eta\sigma$. Off their union the baseline is zero. The stated magnitude bound follows from bounded dual weights. The same support bound holds for $\sigma = 0$.

Multiply this display by the $q \le 3$ density difference in Lemma G.2. Integrating first in T over an interval of length $O_a(\theta(1 + |Q| + |N|))$ leaves a finite Gaussian polynomial moment. Hence

$$|\mathbb{E}H_\lambda - \mathbb{E}_G H_\lambda| \le C_a\tau\theta.$$

For $\theta > 1/4$, use $|H_\lambda| \le C_a(1 + |N|)$ and the global density comparison, giving $C_a\tau \le 4C_a\tau\theta$. Enlarge C_a and combine the bounds. Zero angle is immediate. □

G.4 Localization and compression without Gaussian independence

For an angular annulus $r \le \theta \le \min\{2r, \pi\}$, write $\alpha_\lambda, \gamma_\lambda$ for the weights above.

Lemma G.5 (Uniform local moments)**.** *There are constants $L_a, C'_a > 0$ and a function $\omega_a(M) \downarrow 0$, all independent of $k, d, \tau \le \tau_0$, such that for every frame, annular tuple, labeling, and residual unit n,*

$$\mathbb{P}(\alpha_\lambda \neq 0) + \mathbb{P}(\gamma_\lambda \neq 0) \le L_a r, \tag{G.7}$$

$$\mathrm{Var}(\alpha_\lambda(E_v - \eta)) \le L_a r, \quad \mathrm{Cov}(\gamma_\lambda \,\mathrm{sign}(T) R E_v) \preceq L_a r I_d, \tag{G.8}$$

$$\mathbb{E}[(\gamma_\lambda n \cdot X)^2] \le L_a r, \tag{G.9}$$

$$\mathbb{E}[|\gamma_\lambda n \cdot X| \mathbf{1}\{|T| > Mr\}] \le r\omega_a(M), \tag{G.10}$$

$$\mathbb{E}[|\gamma_\lambda| \, \|m(U)\|] \le C'_a\tau\sqrt{k}\, r. \tag{G.11}$$

Proof. For $r \le 1/8$, the support of either weight lies in $|T| \le C_a r(1 + |Q|)$, as in the previous proof. The $q \le 3$ Gaussian upper envelopes in Lemma G.2 imply

$$\mathbb{P}(\alpha \neq 0) \le C_a r, \quad \mathbb{P}(\gamma \neq 0) \le C_a r, \quad \mathbb{E}[N^2 \mathbf{1}\{\gamma \neq 0\}] \le C_a r.$$

For example integrate the density in T on an interval of length $C_a r(1 + |Q|)$ and then integrate its polynomially weighted Gaussian envelope in (Q, N). The scalar and vector covariance bounds follow by discarding $E_v^2 \le 1$ and taking every residual direction. If $|T| > Mr$ on the support, then $1 + |Q| > M/C_a$. The same integration shows the tail in (G.10) is bounded by r times a Gaussian polynomial tail tending to zero with M. No label independence is used.

For (G.11), let $E = \{\gamma \neq 0\}$. Every coordinate U_j lies in a space of dimension at most three together with T, Q. In an orthonormal basis of that space, $U_j^2 \le T^2 + Q^2 + Z_j^2$ (with unused coordinates omitted). The same upper-envelope integration gives

$$\mathbb{E}[U_j^2 \mathbf{1}_E] \le C_a r, \qquad \mathbb{E}[\|U\|^2 \mathbf{1}_E] \le C_a k r.$$

Thus Cauchy–Schwarz, (G.7), and Lemma G.3 give

$$\mathbb{E}[|\gamma| \|m(U)\|] \le \frac{B\tau}{1 - \tau} \mathbb{E}[(\|U\| + \sqrt{k})\mathbf{1}_E] \le C'_a\tau\sqrt{k}\, r.$$

For $r > 1/8$ the moment bounds follow from global second moments and $r \ge 1/8$; the tail follows from the two-dimensional Gaussian upper envelope for (T, N) and $Mr \ge M/8$. This proves a common function $\omega_a(M) \downarrow 0$ over all annuli. □

Lemma G.6 (Stable constant rank)**.** *Fix $\lambda > 0$ and one measurable function $e : \mathbb{R}^d \to [0,1]$ used at every stage. There are an integer $K = K_{a,\lambda}$ and $\tau_{a,\lambda} > 0$, independent of r, d, such that the following cannot run for K steps when $\tau \le \tau_{a,\lambda}$. Starting with the candidate frame, at stage j choose an annular tuple relative to the current frame P_j and a residual unit n_j, append n_j, and require*

$$\mathbb{E}[e(X)F_j(X)] > \lambda r, \qquad F_j = \gamma_{\lambda_j} \operatorname{sign}(T) n_j \cdot X, \qquad 0 \le e \le 1.$$

Proof. Write $U_j = P_j^\top X$ and define the auxiliary witness

$$W_j = F_j - \mathbb{E}[F_j \mid U_j] = \gamma_{\lambda_j} \operatorname{sign}(T)(n_j \cdot X - \mathbb{E}[n_j \cdot X \mid U_j]).$$

Every previous W_i is measurable with respect to U_j, because its new direction was appended at step i. Hence $\mathbb{E}[W_j \mid U_j] = 0$ makes the W_j pairwise orthogonal. Conditional centering gives $\mathbb{E}W_j^2 \le \mathbb{E}F_j^2 \le L_a r$. Since $\{|T| > Mr\}$ is U_j-measurable, conditional Jensen gives

$$\mathbb{E}[|W_j|\mathbf{1}\{|T| > Mr\}] \le 2\mathbb{E}[|F_j|\mathbf{1}\{|T| > Mr\}] \le 2r\omega_a(M).$$

Choose $M \ge 1$ so that $2\omega_a(M) \le \lambda/4$. The one-dimensional density bound yields $\mathbb{P}(|T| \le Mr) \le C_T M r$, for a universal C_T. Choose an integer

$$K > 4L_a C_T M/\lambda^2.$$

All these choices precede the choice of τ and are independent of K through the constants L_a, ω_a. Choose $\tau_{a,\lambda} \le \tau_0$ so small that $C_a' \tau_{a,\lambda}\sqrt{K} \le \lambda/4$. For the first K attempted additions, $\dim U_j \le K$. By (G.11), $|\mathbb{E}[e(F_j - W_j)]| \le \lambda r/4$. Therefore

$$\mathbb{E}[e\mathbf{1}\{|T| \le Mr\}W_j] > \lambda r/2.$$

Bessel applied to these orthogonal W_j gives

$$K\frac{\lambda^2 r}{4L_a} < \sum_{j=1}^{K} \frac{|\mathbb{E}[e\mathbf{1}\{|T| \le Mr\}W_j]|^2}{\mathbb{E}W_j^2} \le \mathbb{E}[e^2\mathbf{1}\{|T| \le Mr\}] \le C_T M r,$$

contradicting the choice of K. This proof also covers adaptive realized sequences whenever their population correlations satisfy the displayed condition. It uses neither observed nor estimated conditional means. □

G.5 Finite-sample verifier and honest construction

Shrink τ_a first so that $C_a\tau_a \le g_a/2$ in Lemma G.4, and put $g = g_a/2$. Set $t_r = gr/32$ and $\lambda = g/512 = t_r/(16r)$. Obtain K from Lemma G.6, and shrink τ_a again to at most its $\tau_{a,\lambda}$. This is a noncircular order of constants: the baseline, local variance, and local tail constants use at most three-dimensional projections and are independent of the eventual rank.

For $0 < \varepsilon \le 1/64$, let $r_0 = \varepsilon/(8C_{\mathrm{ang}})$, decreased further by a universal constant if necessary, and use dyadic annuli $[r_0, 2r_0], [2r_0, 4r_0], \dots$ ending at π. There are $J = O(\log(e/\varepsilon))$ annuli. For each annulus the message supplies an orthonormal frame containing the candidate and at most $K - 1$ additional directions. Reject malformed messages. For each frame dimension and annulus, fix in advance one maximal νr-separated set within that annulus on the feasible coefficient sphere, shared by the honest prover and verifier and expressed in the ordered frame's coordinates, where $\nu > 0$ is a sufficiently small constant depending only on a. The annulus lies within Euclidean distance $2r$

of the candidate coefficient vector, so the net has at most $F_a = (1 + 4/\nu)^{K+1}$ points. If the frame exhausts the ambient space, use only $\sigma = 0$.

By (G.4) and the elementary bound $\angle(x, y) \le (\pi/2)\|x - y\|$ for unit vectors, choose ν such that rounding coefficients changes risk by at most $gr/16$. Keep the same residual direction while rounding. Thus no net over ambient residual directions is required and no continuity claim for the dual weights is required.

Private tests. All verifier observations are drawn after the entire message. For every net tuple estimate A_λ to additive accuracy t_r, and require $\widehat{A}_\lambda \le 2t_r$. Test $\|V_\lambda\|$ with completeness radius $t_r/4$ and soundness radius t_r. Lemma G.5 and Bernstein give scalar cost $O_a(r^{-1}\log(JF_a/\delta))$. For the vector observation $Z = \gamma_\lambda \operatorname{sign}(T) R E_v$, the off-diagonal mean-norm statistic on N observations has

$$\mathbb{E}U_N = \|\mu\|^2, \qquad \operatorname{Var}(U_N) = \frac{4\mu^\top \Sigma_Z \mu}{N} + \frac{2\operatorname{tr}(\Sigma_Z^2)}{N(N-1)}.$$

Since $\Sigma_Z \preceq L_a r I_d$, a block of

$$N_0(r) = \left\lceil \max\left\{2, \frac{512 L_a r\sqrt{d}}{t_r^2}\right\}\right\rceil = O_a(\sqrt{d}/r)$$

observations distinguishes these two radii with error at most $1/16$ by thresholding U_N at $t_r^2/2$. Indeed, when $\|\mu\| \le t_r/4$ the expectation lies at least $7t_r^2/16$ below the threshold, and when $\|\mu\| > t_r$ it lies at least $\|\mu\|^2/2$ above it; the displayed variance formula and Chebyshev's inequality give the claim uniformly over both ranges. Taking the median of $O(\log(JF_a/\delta))$ independent blocks and thresholding at $t_r^2/2$ gives either error at most $\delta/(4JF_a)$. Observations can be reused across the finite tuple set in a block; union bounds require no independence among those tuple tests. Summing the geometric costs over annuli gives the stated m_V.

With failure at most $\delta/2$, every accepted tuple has $A_\lambda \le 3t_r$, $\|V_\lambda\| \le t_r$. Its gain is at most $-gr + 8t_r = -3gr/4$. Coefficient rounding preserves strictly negative gain. Competitors below the smallest angle differ from the candidate on mass at most $C_{\mathrm{ang}} r_0 \le \varepsilon/8$. Hence acceptance implies ε-optimality for arbitrary labels.

Sample-only honest prover. First find a proper candidate with disagreement $q \le c_a\varepsilon$ from the Massart target, with probability at least $1 - \delta/4$. A Euclidean normal net, the disagreement bound (G.4), uniform convergence, and empirical risk minimization suffice with $O_a((d\log(e/\varepsilon) + \log(1/\delta))/\varepsilon^2)$ samples. Explicitly, a net approximation to the target has excess risk $O_a(\varepsilon)$, and under Massart noise every candidate's excess risk is at least a times its disagreement with the target. Choose constants so that $q \le t_{r_0}$. For every feature-measurable weight $0 \le \alpha(X) \le 1$,

$$\mathbb{E}[\alpha(E_v - \eta)] \le q,$$

because outside the candidate–target disagreement the conditional candidate error is at most η. This ensures scalar completeness.

For each annulus separately start with the candidate frame and allow at most K estimation stages and $K - 1$ additions. If a Kth addition is requested, abort and send a rejecting message. At each stage estimate all net vector means to Euclidean accuracy $t_r/16$ using fresh observations. Stop if every estimated norm is at most $t_r/8$; otherwise append the unit direction of a violating estimated mean. These means and estimates are in the current residual space. On the accuracy event every addition has population correlation greater than $t_r/16 = \lambda r$ with $e(X) = \mathbb{P}(Y \ne h_v(X) \mid X)$.

Lemma G.6 forces termination before K additions on the accuracy event, so the deterministic cap does not affect completeness. It bounds the sample count on every execution. At termination every true vector norm is at most $3t_r/16 < t_r/4$, ensuring vector-test completeness. Exhausting the ambient space forces all remaining vectors to zero.

For completeness, a finite-sample estimate needs no model-fitting oracle. For any unit b the scalar projection of a vector observation is bounded in magnitude by $B|b \cdot R|$. The one-dimensional upper envelope in Lemma G.2 therefore bounds its sub-Gaussian norm by a universal constant. Centering, independent averaging, and a 1/2-net of the ambient sphere give

$$\mathbb{P}(\|\widehat{\mu} - \mu\| > x) \le 2 \cdot 5^d \exp(-c_a N x^2).$$

Thus $O_a((d + \log(JF_aK/\delta))/r^2)$ fresh observations per stage suffice. A union bound over the KJ possible stages and their nets makes all estimates accurate with failure at most $\delta/4$. Their geometrically summed cost is $O_a((d+\log(e/(\varepsilon\delta)))/\varepsilon^2)$. Combine with the candidate cost. The total prover failure is at most $\delta/2$, and the verifier's honest-test failure is at most $\delta/2$. Larger ε can use the protocol at the fixed accuracy 1/64, with a change in constants. This proves Theorem G.1.

G.6 Nonparametric scope

The permissible curvature radius is a positive constant determined after the noise-dependent angular margin and compression rank. It does not shrink with accuracy or dimension. In particular the argument introduces no $O(\tau)$ accuracy floor: both the baseline perturbation and the conditional-mean drift are proportional to the current boundary scale. The distribution family has an unknown function V, not merely an unknown covariance matrix.

Lemma G.7 (The family is not globally close in total variation)**.** *For every $0 < \tau \le 1/16$, there are laws $\mu_d \in \mathcal{D}_\tau$ with $\mathrm{TV}(\mu_d, N(0, I_d)) \longrightarrow 1$ as $d \longrightarrow \infty$.*

Proof. Fix $\rho = \tau/8$ and let Z have density proportional to $\exp(-z^2/2 - \rho\cos z)$. It is symmetric and non-Gaussian. Write $s^2 = \mathbb{E}Z^2$ and let $X = Z/s$, so X is centered and has variance one. The density ratio of Z to a standard Gaussian lies between $e^{-2\rho}$ and $e^{2\rho}$, giving $e^{-2\rho} \le s^2 \le e^{2\rho}$. The potential of X has second derivative

$$s^2(1 - \rho\cos(sx)) \in [e^{-2\rho}(1-\rho), e^{2\rho}(1+\rho)] \subset [1-4\rho, 1+4\rho] \subset [1-\tau, 1+\tau].$$

The elementary middle inclusion holds for $0 \le \rho \le 1/128$. Let p be the density of X and take $\mu_d = p^{\otimes d}$. Its potential has diagonal Hessian with the required bounds. Since $p \ne \phi$, its Hellinger affinity $A = \int \sqrt{p\phi}$ is strictly smaller than one. The inequality $\min(f, g) \le \sqrt{fg}$ gives

$$\mathrm{TV}(p^{\otimes d}, \phi^{\otimes d}) = 1 - \int \min(p^{\otimes d}, \phi^{\otimes d}) \ge 1 - A^d \longrightarrow 1.$$

□

Universality over strongly log-concave marginals itself is already present in [GKSV23]: its Theorems 1.2–1.3 give polynomial sample tester-learners with $O(\mathrm{opt}) + \varepsilon$ error for arbitrary labels and $\mathrm{opt} + \varepsilon$ when Massart noise is assumed. Noise-testable Gaussian learning in [GKSV26, Theorem 10] certifies near-optimal empirical risk for every accepted dataset, with polynomial sample complexity under Gaussian Massart completeness. The guarantee here combines the sharp sublinear verifier sample count, full Massart completeness, and additive arbitrary-label risk guarantee under the stated unknown feature-family promise. Lemma G.7 shows that the curvature condition does not imply a dimension-uniform total-variation neighborhood of the standard Gaussian. The localized proof uses low-dimensional projected densities and conditional drift, rather than a bound on full-dimensional total variation.

H A localized square-root lower bound for halfspaces

This section strengthens the mass-only conditional obstruction for halfspaces by tracking accuracy. For every fixed $0 < a < a_c$, we prove an $\Omega_a(\sqrt{d}/\varepsilon)$ verifier sample bound for sufficiently small accuracy and sufficiently large dimension, with thresholds depending only on a. Completeness is required only at one random-classification-noise null, while the lower bound allows improper outputs and arbitrary interaction. The alternatives modify labels only in a boundary strip of mass $O_a(\varepsilon)$; this localizes both the likelihood second moment and the fresh validation cost. Throughout, $d \geq 2$, $X = (T, Z) \sim N(0, I_d)$, and $h_0(X) = \operatorname{sign}(T)$. Here OPT_P denotes the infimum of risk under P over all homogeneous halfspaces.

H.1 Conditional allocation and its small-angle gain

Write ϕ, Φ for the standard normal density and cdf and $\phi_0 = \phi(0)$. Fix $0 \leq a \leq 1$ and let $\mathcal{R}_a$ be the set of measurable regressions $r : \mathbb{R}^d \to [-1, 1]$ such that

$$\mathbb{E}_Z[r(t, Z)] = a \operatorname{sign}(t) \quad \text{for almost every } t.$$

For a regression r, labels are generated by $\mathbb{P}(Y = y \mid X = x) = (1 + yr(x))/2$. Every $r \in \mathcal{R}_a$ has $R_r(h_0) = (1 - a)/2$. Let h_θ be any homogeneous halfspace whose normal makes angle $\theta \in [0, \pi]$ with the first coordinate vector. Conditional on $|T| = t$, its disagreement probability with h_0 is

$$q_\theta(t) = \Phi(-t \cot \theta), \qquad 0 < \theta < \pi,$$

with the natural values zero and one at $\theta = 0, \pi$.

Lemma H.1 (Exact projected-marginal extremum)**.** *For $0 < \theta < \pi$,*

$$\sup_{r \in \mathcal{R}_a} \{R_r(h_0) - R_r(h_\theta)\} = I_a(\theta) := 2 \int_0^\infty \phi(t) \min\{q_\theta(t), 1 - a - q_\theta(t)\}\, dt.$$

The supremum is attained by a measurable stochastic-label regression.

Proof. Put $\eta = (1 - a)/2$ and $S = Y h_0(X)$. The projected-marginal constraint states that $\mathbb{P}(S = -1 \mid T) = \eta$. Write $D = \mathbf{1}_{\{h_\theta(X) \neq h_0(X)\}}$. The risk improvement equals $-\mathbb{E}[SD]$. At a fixed $T = t$, let $q = \mathbb{P}(D = 1 \mid T = t)$. The largest possible joint error mass $\mathbb{P}(S = -1, D = 1 \mid T = t)$ is $\min\{\eta, q\}$. Hence the conditional improvement is at most $2 \min\{\eta, q\} - q = \min\{q, 2\eta - q\}$.

This bound is attainable. For $0 < \eta < 1$ let the conditional error probability $z = \mathbb{P}(S = -1 \mid X)$ be

$$z = \begin{cases} (\eta/q) D, & q \geq \eta, \\ D + \dfrac{\eta - q}{1 - q}(1 - D), & q < \eta. \end{cases}$$

Both cases have $0 \leq z \leq 1$ and $\mathbb{E}[z \mid T] = \eta$ and attain the bound. Take $r(X) = h_0(X)(1 - 2z(X))$. The endpoints in a follow directly or by limits. Integrating over the half-normal distribution of $|T|$ proves the result. □

For $0 \leq a < 1$, define

$$\eta_a = \frac{1 - a}{2}, \qquad k_a = \Phi^{-1}(1 - \eta_a) = \Phi^{-1}\left(\frac{1 + a}{2}\right), \qquad q(u) = \Phi(-u),$$

and

$$g_a(u) = \min\{q(u), 1 - a - q(u)\}, \qquad J_a = \int_0^\infty g_a(u)\,du.$$

The two expressions in the minimum coincide at $u = k_a$. Using $\int_k^\infty q(u)\,du = \phi(k) - kq(k)$ and $\int_0^\infty q(u)\,du = \phi_0$ gives

$$J_a = (1-a)k_a - \phi_0 + 2\{\phi(k_a) - k_a\eta_a\} = 2\phi(k_a) - \phi_0. \tag{H.1}$$

For acute angles, the substitution $t = u\tan\theta$ gives

$$I_a(\theta) = 2\tan\theta \int_0^\infty \phi(u\tan\theta) g_a(u)\,du.$$

The function g_a is integrable. Dominated convergence and (H.1) yield

$$\lim_{\theta\downarrow 0} \frac{I_a(\theta)}{\theta} = 2\phi_0 J_a = \frac{2e^{-k_a^2/2} - 1}{\pi}. \tag{H.2}$$

Thus the small-angle slope is positive exactly when $a < a_c = 2\Phi(\sqrt{2\log 2}) - 1$. The Gaussian constant $\sqrt{2\log 2}$ also appears in the linear-programming decoding analysis of Dwork–McSherry–Talwar [DMT07]. Their corruption threshold is $1 - a_c$, twice our noise rate $\eta_c = (1 - a_c)/2$; the Gaussian cutoff itself is an established quantity.

H.2 Local alternatives for every amplitude below the threshold

Fix for the rest of the section $0 < a < a_c$, so $J_a > 0$. Choose a finite $M = M_a \geq \max\{1, k_a\}$ so large that

$$(1+a)\int_M^\infty q(u)\,du \leq J_a/4.$$

Put

$$H_a = \int_0^M u^2 |g_a(u)|\,du > 0, \qquad \theta_a = \min\left\{\frac{\pi}{4}, \arctan\sqrt{\frac{J_a}{2H_a}}\right\}, \qquad \varepsilon_a = \frac{\phi_0 J_a \theta_a}{4}.$$

Take $0 < \varepsilon \leq \varepsilon_a$ and define

$$\theta = \frac{4\varepsilon}{\phi_0 J_a}, \qquad b = M\tan\theta, \qquad p = \mathbb{P}(|T| \leq b) = 2\Phi(b) - 1.$$

The null Q has regression $r_0 = ah_0$ and Bayes risk $B = (1-a)/2$. For each $v \in S^{d-2}$, use the tilted halfspace

$$h_v(T, Z) = \operatorname{sign}(T + \tan\theta\, v \cdot Z).$$

Inside the strip $|T| \leq b$, use the conditional error allocation from Lemma H.1 for this halfspace, with $q = \Phi(-|T|\cot\theta)$, and let its regression be $r_v = h_0(1 - 2z_v)$. Outside the strip set $r_v = r_0$. Write P_v for the resulting joint labeled distribution.

Lemma H.2. *These alternatives are measurable, bounded by one, and satisfy*

$$\mathbb{E}_Z[r_v(t, Z)] = a\operatorname{sign}(t), \qquad R_{P_v}(h_0) = B, \qquad \mathrm{OPT}_{P_v} \leq R_{P_v}(h_v) \leq B - 4\varepsilon.$$

Moreover, with $C_a = 16M/J_a$,

$$p \leq C_a\varepsilon, \qquad \|r_v - r_0\|_2^2 \leq (1-a^2)p \leq (1-a^2)C_a\varepsilon.$$

Proof. The conditional error construction proves the first claims and preserves the anchor's risk. Inside the strip the improvement is the extremal value; outside it the tilted halfspace loses a times its disagreement probability. Therefore the total improvement is

$$2\tan\theta\left\{\int_0^M \phi(u\tan\theta)g_a(u)\,du - a\int_M^\infty \phi(u\tan\theta)q(u)\,du\right\}.$$

Since $M \geq k_a$, $g_a = q$ beyond M. Replacing all Gaussian weights by ϕ_0 gives bracket at least $3\phi_0 J_a/4$ by the choice of M. For the first integral, the replacement error is at most $\phi_0 \tan^2\theta H_a/2 \leq \phi_0 J_a/4$. For the negative second integral, using $\phi(u\tan\theta) \leq \phi_0$ is already a valid lower bound. The bracket is thus at least $\phi_0 J_a/2$, proving improvement at least $\phi_0 J_a \tan\theta \geq 4\varepsilon$.

The inequalities $p \leq 2\phi_0 M\tan\theta \leq 4\phi_0 M\theta$ give the bound on p. At each $T = t$ in the strip, $r_v \in [-1,1]$ and its conditional mean is $r_0(t)$, so

$$\mathbb{E}_Z[(r_v - r_0)^2 \mid T = t] = \mathbb{E}_Z[r_v^2 \mid T = t] - a^2 \leq 1 - a^2.$$

Outside the strip the difference vanishes. Integrating proves the L^2 bound. □

H.3 A rank-one mixture bound, without an SQ restriction

Let $D = d - 1$ and let v, w be independent uniform directions on S^{D-1}. For t examples write $\overline{P}_t = \mathbb{E}_v P_v^{\otimes t}$.

Lemma H.3. *For $4tp < D$,*

$$1 + \chi^2(\overline{P}_t \| Q^{\otimes t}) \leq \exp(t^2p^2/D)(1 - 4tp/D)^{-1/4}.$$

In particular, if $tp \leq \sqrt{D}/20$, then $\mathrm{TV}(\overline{P}_t, Q^{\otimes t}) < 1/6$.

Proof. Write $\delta_v = r_v - r_0$. The likelihood ratio is

$$\frac{dP_v}{dQ}(X,Y) = 1 + \frac{Y - r_0(X)}{1 - a^2}\delta_v(X).$$

Its cross moment with the ratio for w is $1 + \mathbb{E}_X[\delta_v(X)\delta_w(X)]/(1-a^2)$. At a fixed $T = t$, expand the function $z \mapsto \delta_v(t,Z)/\sqrt{1-a^2}$, which depends on Z only through $z = v \cdot Z$, in orthonormal Hermites. Its constant coefficient vanishes because the conditional mean $\mathbb{E}_Z\delta_v(t,Z)$ is zero. Hermite orthogonality then shows that the cross moment equals $1 + K(v \cdot w)$, where

$$K(\rho) = \sum_{k\geq 1} b_k \rho^k, \qquad b_k \geq 0, \qquad \sum_{k\geq 1} b_k = \frac{\mathbb{E}_X \delta_v(X)^2}{1 - a^2} \leq p.$$

The nonnegative coefficients include the integral over T of each squared Hermite coefficient. In particular $K(\rho) \leq b_1\rho + p\rho^2$ for $|\rho| \leq 1$.

For the spherical inner product $\rho = v \cdot w$, the standard moment formula gives

$$\mathbb{E}\rho^{2j} \leq (2j-1)!!/D^j, \qquad \mathbb{E}e^{s\rho} \leq e^{s^2/(2D)}, \qquad \mathbb{E}e^{s\rho^2} \leq (1 - 2s/D)^{-1/2} \quad (2s < D).$$

These follow by expanding the respective moment-generating functions; odd moments vanish. Cauchy–Schwarz therefore gives

$$\begin{aligned} 1 + \chi^2 &= \mathbb{E}(1 + K(\rho))^t \\ &\leq \mathbb{E}e^{tb_1\rho + tp\rho^2} \\ &\leq (\mathbb{E}e^{2tb_1\rho})^{1/2}(\mathbb{E}e^{2tp\rho^2})^{1/2} \\ &\leq e^{t^2b_1^2/D}(1 - 4tp/D)^{-1/4}, \end{aligned}$$

which proves the first statement since $b_1 \le p$.

If $tp \le \sqrt{D}/20$, then $t^2p^2/D \le 1/400$ and $4tp/D \le 1/5$. Using $e^{1/400} \le 400/399$ and $(5/4)^{1/4} \le 17/16$ gives $1 + \chi^2 \le 425/399$, hence $\chi^2 \le 26/399 < 1/9$ and TV $\le \frac{1}{2}\sqrt{\chi^2} < 1/6$. □

H.4 Implication for interactive verification with local soundness

The sample-based verifier below has at most m private iid labeled examples. Its private randomness and any honest-prover samples are independent initially. The prover may receive information through arbitrary interaction, know the distribution exactly, and have unbounded computation. Output is any measurable Boolean predictor or rejection. Unlimited fresh samples from the known Gaussian marginal do not affect the argument.

Theorem H.4. *Fix $0 < a < a_c$ and $0 < \varepsilon \le \varepsilon_a$ as above. Suppose a protocol is complete with probability at least $2/3$ at Q, and for every displayed P_v and every prover strategy is sound with probability at least $2/3$ at excess risk ε. Then, with*

$$L_a = 1 + \log(24)(8C_a + 8/3),$$

its verifier sample complexity satisfies

$$m > \frac{1}{\varepsilon}\left(\frac{\sqrt{d-1}}{20C_a} - L_a\right).$$

In particular, for $d - 1 \ge (40C_aL_a)^2$,

$$m \ge \frac{\sqrt{d-1}}{40C_a\varepsilon} = \Omega_a(\sqrt{d}/\varepsilon).$$

For any prescribed local radius $\tau > 0$, these soundness alternatives all satisfy $\|r_v - r_0\|_2 \le \tau$ whenever additionally

$$\varepsilon \le \frac{\tau^2}{(1-a^2)C_a}.$$

Proof. Use one fixed honest prover strategy for the known Q regardless of the unknown input distribution. Its samples can be simulated from Q independently of the verifier's data. Under every P_v, this is an allowed malicious strategy. If the verifier outputs h, clamp it to h_0 outside the strip, obtaining $\widetilde{h}$. This never increases risk under Q or P_v, because all their regressions equal ah_0 outside the strip. On fresh validation data use

$$W = \mathbf{1}_{\{\widetilde{h}(X) \ne Y\}} - \mathbf{1}_{\{h_0(X) \ne Y\}}.$$

Conditional on the transcript, W is bounded in $[-1, 1]$ and has variance at most p, since it is zero outside the strip. With

$$\ell = \left\lceil \log(24)\left(\frac{8p}{\varepsilon^2} + \frac{8}{3\varepsilon}\right)\right\rceil \le L_a/\varepsilon,$$

Bernstein's inequality estimates $\mathbb{E}W$ to within $\varepsilon/2$ with failure probability at most $1/12$. The final inequality uses $p \le C_a\varepsilon$ and $\varepsilon \le \varepsilon_a < 1$.

Declare null if and only if the verifier did not reject and the estimated relative loss is at least -2ε. Under Q, the true relative loss is nonnegative because h_0 is Bayes, so the null declaration has probability at least $7/12$. Under P_v, on soundness the relative loss of a nonreject output after

clamping is at most -3ε by Lemma H.2. Null declaration consequently has probability at most $5/12$.

The gap is at least $1/6$, also against the uniform mixture of alternatives. All interaction, known-null simulation, and validation are a randomized function of $t = m+\ell$ unknown-distribution examples and independent auxiliary randomness. Data processing and Lemma H.3 force $tp > \sqrt{d-1}/20$. Using $p \le C_a\varepsilon$ and the bound on ℓ proves the sample lower bound. The local-radius assertion follows from Lemma H.2. □

I Further comparison with prior work

Our results separate three aspects of statistical certification: the feature distribution, the label law, and the information available to each party. The Gaussian classification holds for a fixed feature marginal and arbitrary stochastic labels in soundness; completeness requires a positive Massart signal. Both parties obtain independent ordinary examples. These choices make it possible to classify the verifier's statistical complexity while retaining unrestricted computation and unrestricted dishonest provers. The following comparisons explain how this question relates to verification, property testing, and testable learning.

I.1 PAC verification and statistical certificates

Goldwasser, Rothblum, Shafer, and Yehudayoff introduced PAC verification and established both separations between learning and verification and examples where verification requires essentially the learning sample complexity [GRSY21]. Their work includes a one-round sample separation and an interactive protocol for Fourier-sparse Boolean functions whose honest prover uses membership queries. Thus neither the possibility of statistical savings through an untrusted prover nor one-message verification is new. The question addressed here is which geometry determines those savings for Gaussian projection classes under a fixed positive noise allowance.

Mutreja and Shafer extended verification to statistical algorithms and proved a distribution-free $\Omega(\sqrt{v}/\epsilon^2)$ lower bound for classes of VC dimension v [MS26]. Their lower bound simulates the honest prover at a fixed null distribution and validates the returned hypothesis, allowing improper outputs. We use this reduction principle, but construct the null inside the promised Gaussian Massart model. The resulting alternatives have the same feature marginal, and the conditional allocation programs determine which residual moments can be hidden. A lower bound supported on a shattered set or based on fair labels does not establish this promised-marginal claim.

Bun, Desai, and Ferreira Pinto connect verification with identity testing in fooling distance and give general upper and lower bounds in terms of bounded distinguishers [BDFP26]. Their Theorem 5.4 and Corollary 5.8 provide conversions from testing to verification; Algorithm 3 supplies the honest prover with a description of the joint distribution. This framework explains an important route to verification, but leaves the relevant Gaussian loss-testing problem to be solved. Our conditional programs solve the geometric part of that problem, and the upper bounds include construction of an honest message from finitely many examples. In the lower bounds, the null retains a positive classification signal rather than independent fair labels.

Ngo and Kim give publicly verifiable certificates for statistical-query algorithms [NK26]. Their Theorem 3.4 certifies a deterministic algorithm using B queries of tolerance τ with $O(\tau^{-2}\log(B/\delta))$ verifier samples. The certified algorithm must itself solve its stated learning goal for every admissible accurate SQ oracle. Consequently, certifying a learner that is correct only under Massart noise does not supply arbitrary-label excess-risk soundness. Applying the theorem to an agnostic or noise-testable learner also requires retaining its query count and tolerances. Our protocols instead

certify the candidate directly through conditional risk inequalities. Localization makes the variance of each scalar test proportional to the disagreement scale, yielding the ϵ^{-1} dependence for the orthant family.

The protocols of [GJK+24] give dimension-independent verifier sample bounds for coordinate juntas on the Boolean cube, as well as general conversions from suitable query algorithms to interactive learning proofs. Their query-hiding mechanisms use deterministic labels and require the relevant query distributions to satisfy the conversion hypotheses. Their Theorem 7.1 and Protocol 9 also give a generic $O(\epsilon^{-1}\log(1/\delta))$ labeled-sample verifier for finite circuit classes, with an unbounded query prover and additional unlabeled examples depending on VC dimension. Under a known Gaussian marginal, unlabeled features could be generated internally; the decisive obstruction to applying that protocol here is its deterministic labeling oracle and query-based honest prover. Its hidden-label consistency check does not supply completeness at a fixed positive Massart allowance. Thus inverse-linear labeled-sample dependence alone is not the novelty of our Boolean-family theorem. More recently, Amit, Goldwasser, and Rothblum extend interactive proofs to separate persistent noisy views of the data [AGR26]. Their Theorems 5.15 and 5.22 verify heavy Fourier coefficients on the Boolean cube with dimension-independent verifier samples, three messages, and a noisy membership-query honest prover; the permitted noise bound is $O(\epsilon^4)$. These results show that noisy query conversion is possible. Our family theorem has a different guarantee: a fixed positive Massart allowance independent of accuracy, ordinary samples for both parties, one message, and soundness against arbitrary Gaussian stochastic label laws.

Karchmer, Pawelczyk, and Neel study PAC verification of linear data-attribution models under squared loss, measuring verifier cost through model evaluations and retrainings [KPN25]. This provides another setting in which verification avoids recovering all coefficients of a high-dimensional model. Couteau, Melissaris, and Paris study interactive proofs for model-evaluation statistics, separating the cost of sampling inputs from evaluating model outputs [CMP26]. For an interior tolerance gap g, their black-box sampling regime requires $\Theta(g^{-2})$ samples while reducing evaluations to $\widetilde{O}(g^{-1})$; chosen-randomness access to a sampler permits further savings. Their objective is to certify an evaluation statistic. Our objective is excess risk relative to an entire hypothesis class, and our resource bounds charge every example available to the verifier.

I.2 Low-dimensional Gaussian functions and halfspace testing

De, Mossel, and Neeman prove dimension-independent query testing for induced classes of low-dimensional Gaussian functions [DMN19]. Their class framework explicitly includes arbitrary Boolean combinations of a fixed number of halfspaces, with the combining rule unknown. Their subsequent robust-testing theorem gives tolerant testing of smooth linear-junta classes and estimates maximum correlation with a class using point queries [DMN21]. In particular, the latter framework can query a bounded real-valued function. This is a strong form of low-dimensional testability, but ordinary stochastic examples do not provide evaluations of the regression function at chosen points. Our theorem supplies a dimension-independent ordinary-example verifier for the entire class of Boolean rules on a bounded number of orthogonal homogeneous halfspaces. Orthogonality and homogeneity are explicit restrictions of this evaluation of the general classification. Their induced-class framework already allows a collection of unknown intrinsic rules. Theorem A.5 supplies necessary and sufficient cross-rule conditional criteria for ordinary-example verification, with matching unrestricted lower bounds. Its total-boundedness assumption ensures finite intrinsic nets, and its constants can depend on their entropy and uniform rotation modulus.

The square-root-dimensional statistic used in Gaussian halfspace testing has an earlier history. Balcan, Blais, Blum, and Yang give a passive Gaussian halfspace tester using $O(\sqrt{d}\log d)$ labels at

fixed accuracy and a nearly matching lower bound [BBBY12]. Their statistic averages $Y_iY_j\langle X_i, X_j\rangle$ over distinct pairs. Harms extends passive halfspace testing to unknown rotation-invariant distributions, with square-root dependence on dimension [Har19]. These membership-testing results concern deterministic functions; their completeness requirements do not give additive excess-risk verification above a fixed positive Massart floor. Rotation invariance also differs from the Hessian perturbation promise used here: the latter allows nonradial laws. Our unknown-marginal result preserves a localized noisy verification certificate and its $\sqrt{d}/\epsilon$ sample rate under a perturbation whose size is independent of dimension and accuracy.

Chen, De, Huang, Nadimpalli, Servedio, and Yang study relative-error testing of halfspaces, including a square-root-dimensional algorithm using samples from the positive preimage of a deterministic function and knowledge of its Gaussian volume [CDH+26]. Their other algorithms also use point queries. Positive-preimage sampling and ordinary stochastic-label sampling provide different information. Kong and Valiant estimate optimal linear squared loss from sublinear data and obtain classification results under a structured logistic label model [KV18]. Their mean-square estimators are a related statistical ingredient; optimal squared loss and arbitrary-label halfspace classification risk are different objectives.

I.3 Testable learning, Massart noise, and polynomial thresholds

Rubinfeld and Vasilyan introduced testable learning to replace an assumed feature law by conditions that the learner can check [RV23]. Gollakota, Klivans, and Kothari develop a general moment-matching and sandwiching-polynomial approach and characterize its statistical sample complexity through Rademacher complexity [GKK23]. Their Theorem 5.2 testably learns arbitrary Boolean functions of a constant number of halfspaces, including nonorthogonal and affine halfspaces, with Gaussian sample and time complexity $d^{\widetilde{O}(1/\epsilon^2)}$. Slot, Tiegel, and Wiedmer extend Gaussian testable learning to every fixed degree of polynomial threshold functions, with sample and time complexity $d^{\mathrm{poly}(1/\epsilon)}$ [STW24]. Alman, Patel, and Servedio give a subexponential-time distribution-free realizable PAC learner for Boolean functions of a fixed number of arbitrary affine halfspaces [APS26]. Their Theorems 1 and 18 do not assert full Massart completeness or private-verifier sample guarantees. This is a computational advance for a broader class, whereas our quantitative theorem controls the statistical cost of an assisted verifier. The testable-learning results above cover broader function classes and test the feature assumption. Our family theorem assumes the Gaussian marginal and shifts the cost of recovering a hypothesis to the prover, reducing the verifier's ordinary-example cost to $2^{O(K)}\epsilon^{-1}\log(e/(\epsilon\delta))$ at the explicit allowance $2^{-3K}/(2^{14}K^4)$. The growing-rank consequence uses a decreasing allowance; it does not give fixed positive noise for unbounded K.

Efficient halfspace tester-learners also exploit localization and label-dependent tests. Diakonikolas, Kane, Kontonis, Liu, and Zarifis give a polynomial-time tester-learner with error $O(\mathrm{OPT})+\epsilon$ under Gaussian features [DKK+23]. Gollakota, Klivans, Stavropoulos, and Vasilyan give efficient tester-learners for homogeneous halfspaces, with an additive guarantee under the Massart promise and a multiplicative guarantee for arbitrary labels [GKSV24]; their universal algorithms handle unknown families including strongly log-concave Poincaré laws [GKSV23]. Diakonikolas, Kane, Liu, and Zarifis obtain polynomial-time testable learning of general halfspaces with error $\widetilde{O}(\sqrt{\mathrm{OPT}})+\epsilon$ [DKLZ24]. The fixed positive noise floor matters here: a multiplicative factor on OPT cannot be absorbed into an arbitrarily small additive tolerance.

Goel, Klivans, Stavropoulos, and Vasilyan test noise assumptions directly and obtain a polynomial-sample noise-testable learner for Gaussian homogeneous halfspaces [GKSV26]. It permits full feature-dependent Massart noise and certifies arbitrary-label population excess risk (Definition 1.3), as well as near-optimality on arbitrary accepted datasets (Theorem 2.7). Thus the asymmetry

between Massart completeness and arbitrary-label soundness already appears in this literature. Diakonikolas, Iakovidis, Kane, and Liu extend noise-testable learning to general Gaussian halfspaces [DIKL26]; their Theorem 1.4 gives a quasipolynomial algorithm whose complexity depends on the halfspace bias, and their lower bounds concern SQ computation. Our quadratic separation is an unrestricted statistical separation between balanced classes with the same intrinsic rank and minimum threshold degree. It changes one offset while retaining the Gaussian marginal and the noise allowance.

Lange and Qiao show that membership queries do not yield superpolynomial computational improvements over sample-based distribution-specific learning in their testable-learning model [LQ26]. Their reductions concern learning and refutation without an untrusted prover. They do not preclude the verification savings obtained here, but reinforce the need to keep query access, ordinary examples, and prover assistance distinct.

I.4 Geometric and lower-bound ingredients

Gaussian boundary estimates and noise sensitivity are established tools. Kane proves sharp Gaussian surface-area and noise-sensitivity bounds for polynomial threshold functions, including tight examples given by products of homogeneous linear forms [Kan11]. Our orthant argument requires additional information: a lower bound on disagreement after quotienting by all exact symmetries, concentration of disagreement near candidate facets, and control of the worst admissible conditional error allocation. Together these give a strictly negative allocation envelope at every nonzero frame distance and scalar certificates with localized variance. A global noise sensitivity bound alone supplies neither the allocation inequality nor its verification consequence. The quantitative facet-exclusion bound additionally controls matching of normals across unequal ranks and separates inequivalent truth tables. It yields explicit sample factors and a growing-rank range. The opposite-orthant obstructions show exponential noise decay is necessary in order; they do not match the exact noise exponent or the verifier's rank dependence.

For fixed thin slabs, Theorem 4.5 determines the additive rate $\widetilde{\Theta}(d + 1/\epsilon)$. Its upper bound combines a localized scalar margin with a fixed-resolution distant-comparator check. The latter uses negative acceptance thresholds justified by Massart completeness, so it costs $O(d)$ rather than an accuracy-resolved ambient-net cost. The two-target accuracy lower bound and known-null simulation are established techniques; the joint geometric upper bound is the additional content. Constants depend on the fixed width and noise, and no uniform statement for vanishing width is asserted.

Hidden-direction moment matching is a standard source of statistical and SQ lower bounds; Diakonikolas, Kane, and Stewart give both types of consequences in robust Gaussian estimation [DKS17]. Conditional Hermite information and sequential subspace discovery also arise in the generative-leap theory of Damian, Lee, and Bruna [DLB25]. That theory concerns efficient recovery when the conditional label law depends on the planted subspace, with low-degree lower bounds. Here the noise may depend on every ambient coordinate, and the lower bounds apply to unrestricted interactive verification. The allocation constraints identify exactly when the noise can remove the residual first-order information; this is the matching obstruction to the upper certificates.

Finally, the curvature argument uses classical Brascamp–Lieb covariance estimates [CCEL13], and the halfspace threshold contains the Gaussian trimming constant also encountered in LP decoding [DMT07]. The contribution is the resulting verification characterization and its localized stability, with these geometric and probabilistic ingredients credited separately.

References

[AGR26] Noga Amit, Shafi Goldwasser, and Guy N. Rothblum. Interactive proofs with noisy data. Technical Report TR26-209, Electronic Colloquium on Computational Complexity, 2026. URL: https://eccc.weizmann.ac.il/report/2026/209/.

[APS26] Josh Alman, Shyamal Patel, and Rocco A. Servedio. Learning functions of halfspaces. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing*, 2026. Full version: arXiv:2603.08700v2. URL: https://arxiv.org/abs/2603.08700.

[BBBY12] Maria-Florina Balcan, Eric Blais, Avrim Blum, and Liu Yang. Active property testing. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science*, pages 21–30, 2012. Full version: arXiv:1111.0897v2. URL: https://www.cs.cmu.edu/~ninamf/papers/active-testing.pdf.

[BDFP26] Mark Bun, Rathin Desai, and Renato Ferreira Pinto, Jr. Testing distributions against bounded distinguishers. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing*, 2026. Full version: arXiv:2607.15645. URL: https://arxiv.org/abs/2607.15645.

[CCEL13] Eric A. Carlen, Dario Cordero-Erausquin, and Elliott H. Lieb. Asymmetric covariance estimates of Brascamp–Lieb type and related inequalities for log-concave measures. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 49(1):1–12, 2013. URL: https://www.numdam.org/item/AIHPB_2013__49_1_1_0/, doi:10.1214/11-AIHP462.

[CDH+26] Xi Chen, Anindya De, Yizhi Huang, Shivam Nadimpalli, Rocco A. Servedio, and Tianqi Yang. Sublinear-query relative-error testing of halfspaces. In *53rd International Colloquium on Automata, Languages, and Programming*, volume 374 of *Leibniz International Proceedings in Informatics*, pages 61:1–61:19, 2026. URL: https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ICALP.2026.61, doi:10.4230/LIPIcs.ICALP.2026.61.

[CMP26] Geoffroy Couteau, Nikolas Melissaris, and Tamara Paris. Interactive proofs of proximity for model evaluation. Technical Report TR26-215, Electronic Colloquium on Computational Complexity, 2026. URL: https://eccc.weizmann.ac.il/report/2026/215/.

[DIKL26] Ilias Diakonikolas, Giannis Iakovidis, Daniel M. Kane, and Sihan Liu. Testable learning of general halfspaces under Massart noise. arXiv:2602.22300, 2026. URL: https://arxiv.org/abs/2602.22300.

[DKK+23] Ilias Diakonikolas, Daniel M. Kane, Vasilis Kontonis, Sihan Liu, and Nikos Zarifis. Efficient testable learning of halfspaces with adversarial label noise. In *Advances in Neural Information Processing Systems*, volume 36, pages 39470–39490, 2023. URL: https://papers.nips.cc/paper_files/paper/2023/hash/7c319b62e2257b34cb0e1040ced2e007-Abstract-Conference.html, doi:10.52202/075280-1714.

[DKLZ24] Ilias Diakonikolas, Daniel M. Kane, Sihan Liu, and Nikos Zarifis. Testable learning of general halfspaces with adversarial label noise. In *Proceedings of the Thirty-Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 1308–1335. PMLR, 2024. URL: https://proceedings.mlr.press/v247/diakonikolas24a.html.

[DKS17] Ilias Diakonikolas, Daniel M. Kane, and Alistair Stewart. Statistical query lower bounds for robust estimation of high-dimensional Gaussians and Gaussian mixtures. In *Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science*, 2017. Full version: arXiv:1611.03473. URL: https://arxiv.org/abs/1611.03473.

[DLB25] Alex Damian, Jason D. Lee, and Joan Bruna. The generative leap: Sharp sample complexity for efficiently learning Gaussian multi-index models. arXiv:2506.05500, 2025. URL: https://arxiv.org/abs/2506.05500.

[DMN19] Anindya De, Elchanan Mossel, and Joe Neeman. Is your function low dimensional? In *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 979–993. PMLR, 2019. URL: https://proceedings.mlr.press/v99/de19a.html.

[DMN21] Anindya De, Elchanan Mossel, and Joe Neeman. Robust testing of low-dimensional functions. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, 2021. URL: https://arxiv.org/abs/2004.11642, doi:10.1145/3406325.3451115.

[DMT07] Cynthia Dwork, Frank McSherry, and Kunal Talwar. The price of privacy and the limits of LP decoding. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, 2007. URL: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/lp-decoding.pdf.

[GJK+24] Tom Gur, Mohammad Mahdi Jahanara, Mohammad Mahdi Khodabandeh, Ninad Rajgopal, Bahar Salamatian, and Igor Shinkar. On the power of interactive proofs for learning. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, 2024. URL: https://arxiv.org/abs/2404.08158, doi:10.1145/3618260.3649784.

[GKK23] Aravind Gollakota, Adam R. Klivans, and Pravesh K. Kothari. A moment-matching approach to testable learning and a new characterization of Rademacher complexity. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 1657–1670, 2023. URL: https://arxiv.org/abs/2211.13312, doi:10.1145/3564246.3585206.

[GKSV23] Aravind Gollakota, Adam Klivans, Konstantinos Stavropoulos, and Arsen Vasilyan. Tester-learners for halfspaces: Universal algorithms. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL: https://proceedings.neurips.cc/paper_files/paper/2023/hash/204d9a9a4816a45909010587ffc3204b-Abstract-Conference.html, doi:10.52202/075280-0443.

[GKSV24] Aravind Gollakota, Adam R. Klivans, Konstantinos Stavropoulos, and Arsen Vasilyan. An efficient tester-learner for halfspaces. In *Proceedings of the Twelfth International Conference on Learning Representations*, 2024. URL: https://proceedings.iclr.cc/paper_files/paper/2024/file/a759e253661be0fa3ffe3e37959ecc5e-Paper-Conference.pdf.

[GKSV26] Surbhi Goel, Adam Klivans, Konstantinos Stavropoulos, and Arsen Vasilyan. Testing noise assumptions of learning algorithms. In *Proceedings of the Thirty-Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pages 2804–2853. PMLR, 2026. URL: https://proceedings.mlr.press/v336/goel26a.html.

[GRSY21] Shafi Goldwasser, Guy N. Rothblum, Jonathan Shafer, and Amir Yehudayoff. Interactive proofs for verifying machine learning. In *12th Innovations in Theoretical Computer Science Conference*, volume 185 of *Leibniz International Proceedings in Informatics*, pages 41:1–41:19, 2021. URL: https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ITCS.2021.41, doi:10.4230/LIPIcs.ITCS.2021.41.

[Har19] Nathaniel Harms. Testing halfspaces over rotation-invariant distributions. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 694–713, 2019. URL: https://arxiv.org/abs/1811.00139, doi:10.1137/1.9781611975482.44.

[Kan11] Daniel M. Kane. The Gaussian surface area and noise sensitivity of degree-d polynomial threshold functions. *Computational Complexity*, 20:389–412, 2011. URL: https://cseweb.ucsd.edu/~dakane/noisesensitvity.pdf, doi:10.1007/s00037-011-0012-6.

[KPN25] Ari Karchmer, Martin Pawelczyk, and Seth Neel. Efficiently verifiable proofs of data attribution. arXiv:2508.10866, 2025. URL: https://arxiv.org/abs/2508.10866.

[KV18] Weihao Kong and Gregory Valiant. Estimating learnability in the sublinear data regime. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL: https://papers.nips.cc/paper_files/paper/2018/hash/8bd39eae38511daad6152e84545e504d-Abstract.html.

[LQ26] Jane Lange and Mingda Qiao. Limitations of membership queries in testable learning. In *17th Innovations in Theoretical Computer Science Conference*, volume 362 of *Leibniz International Proceedings in Informatics*, pages 91:1–91:23, 2026. URL: https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ITCS.2026.91, doi:10.4230/LIPIcs.ITCS.2026.91.

[MS26] Saachi Mutreja and Jonathan Shafer. PAC verification of statistical algorithms. *Theory of Computing*, 22(5):1–40, 2026. Conference version in COLT 2023. URL: https://theoryofcomputing.org/articles/v022a005/, doi:10.4086/toc.2026.v022a005.

[NK26] Michael Ngo and Michael P. Kim. Publicly-verifiable certificates for statistical algorithms. arXiv:2607.15528v1, 2026. URL: https://arxiv.org/abs/2607.15528.

[RV23] Ronitt Rubinfeld and Arsen Vasilyan. Testing distributional assumptions of learning algorithms. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 1643–1656, 2023. URL: https://arxiv.org/abs/2204.07196, doi:10.1145/3564246.3585117.

[STW24] Lucas Slot, Stefan Tiegel, and Manuel Wiedmer. Testably learning polynomial threshold functions. In *Advances in Neural Information Processing Systems*, volume 37, pages 3781–3831, 2024. URL: https://papers.nips.cc/paper_files/paper/2024/hash/06e9029d3d4d6cee71c5d9b8502f891b-Abstract-Conference.html, doi:10.52202/079017-0124.