
Sharp Root Anti-Concentration for Online Algorithm Configuration

Pahan Dewasurendra
Johns Hopkins University

Abstract

Online algorithm configuration needs a local guarantee that a random transition boundary does not cross a short parameter interval. Existing polynomial bounds assume a bounded joint coefficient density and are quadratic in the degree. Recent work asks for natural necessary and sufficient coefficient conditions, polynomial guarantees under independent bounded marginals, and a normalization for Pfaffian boundaries. We address these questions through coefficient flux. For an affine boundary $a(t) + \langle X, u(t) \rangle = 0$ with $\|u(t)\|_2 = 1$, we identify two Radon quantities, a slice density and a tangential first moment. Their maximum is exactly the best root-hitting constant uniform over all normalized affine boundaries. If the coefficient law is dominated by an isotropic log-concave law, both quantities are universally bounded. This gives a dimension-free bound controlled only by the total variation of a and the spherical length of u . For monic degree- d polynomials, the resulting joint-density bound is linear rather than quadratic in d . For independent coefficients with marginal densities at most κ and support radius R , a Rogozin–Ball argument gives $O(\kappa R d^{3/2})$, and we construct product laws attaining this rate. Consequently, online losses with random monic transitions have sharp dispersion coefficient $\Theta(\kappa R d^{3/2})$ and $\tilde{O}(\sqrt{T})$ regret. Marginal density without independence can force a deterministic root. Finally, the normalized feature curve $F/\|F\|_2$ supplies a sufficient Pfaffian normalization and yields general dispersion and regret bounds. For exponential-kernel graph learning, the previously hidden conditioning is exactly controlled by the exponent range.

1 Introduction

Many data-driven algorithm families have performance functions that are smooth except across transition boundaries. Examples include changes in a clustering merge, an active set, or a graph prediction. In online algorithm configuration, concentration of these boundaries determines whether the best parameter in hindsight can be learned with sublinear regret (Balcan et al., 2018, 2020, 2025).

A recent invited open problem isolates the missing probabilistic step (Balcan et al., 2026). Let

$$P_\alpha(t) = t^d + \sum_{i=0}^{d-1} \alpha_i t^i, \quad \alpha \in [-R, R]^d.$$

If the joint density is at most κ , the known guarantee for every interval I is

$$\mathbb{P}\{P_\alpha \text{ has a root in } I\} \leq A \left\{ \frac{d}{2R} + \frac{d(d-1)}{4} \right\} |I|, \\ A = \kappa(2R)^d. \quad (1)$$

The open questions ask which coefficient laws admit a finite or polynomial constant, whether independent bounded marginal densities give a polynomial bound without passing through an exponential product density, and what normalization extends the result to Pfaffian feature vectors.

We give an exact answer for unrestricted normalized affine families, settle the requested independent-coordinate polynomial rate, and prove the proposed Pfaffian normalization sufficient. The correct geometric object is not the raw feature derivative. It is the motion of the affine hyperplane after whitening the coefficient law and normalizing the feature vector. Write

$$h_X(t) = a(t) + \langle X, u(t) \rangle, \quad u(t) \in \mathbb{S}^{N-1}. \quad (2)$$

The event that (2) has a root is the event that X belongs to a set swept out by affine hyperplanes. Its probability is controlled by two fluxes: mass crossing under normal translation and mass crossing under rotation. We prove that these two fluxes are also necessary for a theorem uniform over all normalized boundaries.

This characterization turns distributional assumptions into root bounds. Domination by an isotropic log-concave law controls both fluxes in every dimension. Domination is broad enough to include an arbitrary density bounded on a cube, after comparison with the uniform law. Independence permits a different route. The sharp density bound for a normalized sum of independent bounded-density variables (Rudelson and Vershynin, 2015) combines with bounded support and a midpoint argument.

Our main consequences are as follows.

- For arbitrary dependence and joint density at most κ on $[-R, R]^d$, we replace (1) by

$$CA \left(d - 1 + \frac{\sqrt{3d}}{R} \right) |I|.$$

The leading degree dependence improves from d^2 to d . The same linear rate is necessary in the log-concave model, already for iid Gaussian coefficients.

- For independent coordinates with marginal densities at most κ , we obtain

$$\sqrt{2d} \kappa \{1 + R(d - 1)\} |I|.$$

This is polynomial in d, R, κ . A matching construction proves that the leading $\kappa R d^{3/2}$ rate cannot be improved for this distribution class.

- For a nonvanishing feature vector F , bounded spherical speed of $F/\|F\|_2$ is sufficient for root anti-concentration. This verifies the normalization conjectured for Pfaffian boundaries. Controlling that speed from formal Pfaffian complexity remains application-specific.
- For one-dimensional online losses whose transitions are random monic polynomials, the dispersion coefficient is $\Theta(\kappa R d^{3/2})$. Continuous exponential weights therefore has $\tilde{O}(K \kappa R d^{3/2} \sqrt{T})$ regret for fixed Lipschitz and domain parameters. A matching construction makes the degree dependence sharp. For exponential-kernel graph learning, we explicitly bound the feature conditioning by half the exponent range. The general Pfaffian result retains explicit feature conditioning.

The flux identity itself has classical roots. Edelman and Kostlan (1995) derive the homogeneous arbitrary-density expected-zero intensity from spherical hyperplane sweeps. In that homogeneous setting, our upper integrand is the same classical first-moment object followed by hit probability at most expected

count. We do not claim that step as new. Coarea and Kac–Rice formulas give related identities for polynomial zero counts (Götze et al., 2018; Feliu and Sadeghimanesh, 2022). Our contributions are the affine translation–rotation decomposition, exact necessity for every normalized boundary, rough-density and distribution-class criteria, sharp finite-degree consequences, and the resulting online guarantees.

2 Coefficient Flux

Let X have density f on $\mathbb{R}^N$. Write $d\mathcal{H}_u(x)$ for $d\mathcal{H}^{N-1}(x)$ on the slice $\langle u, x \rangle = s$. For $u, v \in \mathbb{S}^{N-1}$, define

$$\mathcal{R}_0(f) = \sup_{u \in \mathbb{S}^{N-1}} \operatorname{ess\,sup}_{s \in \mathbb{R}} \int_{\langle u, x \rangle = s} f(x) d\mathcal{H}_u(x), \quad (3)$$

$$\mathcal{R}_1(f) = \sup_{\substack{u, v \in \mathbb{S}^{N-1} \\ u \perp v}} \operatorname{ess\,sup}_{s \in \mathbb{R}} \int_{\langle u, x \rangle = s} |\langle v, x \rangle| f(x) d\mathcal{H}_u(x). \quad (4)$$

For $N = 1$, set $\mathcal{R}_1(f) = 0$. Every absolutely continuous curve in $\mathbb{S}^0$ is constant, so there is no rotational term. The inner integrals are almost-everywhere Radon derivatives. Essential suprema therefore make the definitions invariant to the version of f . Equivalently, they are the supremal probability and weighted-moment ratios over slabs of positive width. This operational form also fixes exceptional Radon slices. The first quantity is the largest density of a one-dimensional projection. The second is its tangential first-moment analogue. More explicitly, if J ranges over bounded intervals of positive length, then

$$\mathcal{R}_0(f) = \sup_{u, J} \frac{\mathbb{P}\{\langle u, X \rangle \in J\}}{|J|}, \quad (5)$$

$$\mathcal{R}_1(f) = \sup_{u \perp v, J} \frac{\mathbb{E}[|\langle v, X \rangle| \mathbf{1}\{\langle u, X \rangle \in J\}]}{|J|}. \quad (6)$$

These formulas are often easier to verify than the Radon definitions.

For a compact interval I and absolutely continuous $a : I \rightarrow \mathbb{R}$ and $u : I \rightarrow \mathbb{S}^{N-1}$, let $\operatorname{TV}_I(a) = \int_I |a'(t)| dt$ and $\operatorname{Len}_I(u) = \int_I \|u'(t)\|_2 dt$. Write

$$\mathcal{A}_I(a, u) = \{x : \exists t \in I, a(t) + \langle x, u(t) \rangle = 0\}.$$

Theorem 1 (Affine flux bound). *For every coefficient density f and normalized absolutely continuous boundary (a, u) ,*

$$\mathbb{P}\{X \in \mathcal{A}_I(a, u)\} \leq \mathcal{R}_0(f) \operatorname{TV}_I(a) + \mathcal{R}_1(f) \operatorname{Len}_I(u). \quad (7)$$

Moreover, define the universal root constant

$$C_{\text{univ}}(f) = \sup_{I, a, u} \frac{\mathbb{P}\{X \in \mathcal{A}_I(a, u)\}}{\operatorname{TV}_I(a) + \operatorname{Len}_I(u)},$$

where zero denominators are omitted. Then

$$C_{\text{univ}}(f) = \max\{\mathcal{R}_0(f), \mathcal{R}_1(f)\}. \quad (8)$$

Thus a finite linear bound for every normalized affine boundary exists if and only if both fluxes are finite.

The proof uses the area formula rather than root counting. Let $H_t = \{x : \langle x, u(t) \rangle = -a(t)\}$. An absolutely continuous orthonormal frame $Q(t)$ for $u(t)^\perp$ parameterizes the sweep by

$$\Phi(t, z) = -a(t)u(t) + Q(t)z.$$

Its N -dimensional Jacobian is

$$J_\Phi(t, z) = |a'(t) + \langle u'(t), \Phi(t, z) \rangle|.$$

Discarding multiplicity in the area formula yields

$$\mathbb{P}\{X \in \mathcal{A}_I(a, u)\} \leq \int_I \int_{H_t} |a'(t) + \langle u'(t), x \rangle| f(x) d\mathcal{H}^{N-1}(x) dt. \quad (9)$$

Equation (7) follows by writing $u' = \|u'\|v$ with $v \perp u$. For necessity, translated parallel hyperplanes recover (29) exactly. For orthogonal u, v , take $u_\theta = u \cos \theta + v \sin \theta$ and $a(\theta) = -s$. If $Y = \langle u, X \rangle$ and $Z = \langle v, X \rangle$, averaging the root probability over s in an interval J , dividing by $h|J|$, and sending $h \downarrow 0$ gives

$$\frac{\mathbb{E}[|Z| \mathbf{1}\{Y \in J\}]}{|J|}.$$

Thus infinitesimal great-circle rotations recover the slab form of (4). Truncation and dominated-convergence details are in the supplement.

Theorem 1 avoids a bound on the number of roots and does not require simple roots. For a restricted feature family, (7) uses only the directions and tangents reached by that family. Necessity holds for every translated or rotational subfamily it contains.

The Gaussian case makes the two roles transparent.

Corollary 2 (Exact Gaussian constants). *If $X \sim \mathcal{N}(0, I_N)$, then every normalized affine boundary satisfies*

$$\mathbb{P}\{X \in \mathcal{A}_I(a, u)\} \leq \frac{\text{TV}_I(a)}{\sqrt{2\pi}} + \frac{\text{Len}_I(u)}{\pi}. \quad (10)$$

Both coefficients are sharp, and $C_{\text{univ}} = 1/\sqrt{2\pi}$. For homogeneous boundaries, the sharp constant is $1/\pi$.

Indeed, orthogonal Gaussian projections Y, Z are independent. Hence $p_u(s) = (2\pi)^{-1/2}e^{-s^2/2}$ and $r_{u,v}(s) = p_u(s)\mathbb{E}|Z| = \pi^{-1}e^{-s^2/2}$. The claim follows from Theorem 1 and its translation and rotation constructions.

The homogeneous statement recovers the local content of the spherical-length formula of Edelman and Kostlan (1995), now as a hit-probability bound.

The tangential flux is genuinely separate from ordinary projection anti-concentration. Let $X = (Y, Z)$ with Y uniform on $[-1, 1]$ and Z an independent standard Cauchy variable. Every unit projection has a uniformly bounded density. Indeed, convolution bounds its density by the smaller of the rescaled uniform and Cauchy density bounds. Hence $\mathcal{R}_0(f) < \infty$. In contrast, for $u = (1, 0)$ and $v = (0, 1)$,

$$r_{u,v}(s) = p_Y(s)\mathbb{E}|Z| = \infty \quad \text{for almost every } s \in [-1, 1].$$

Thus bounded small-ball probabilities for fixed projections do not control rotating boundaries. The quantity $\mathcal{R}_1$ is necessary even in dimension two.

3 Log-Concave Domination

The exact criterion becomes useful when its two Radon quantities can be controlled. A probability law μ is A -dominated by ν if $d\mu/d\nu \leq A$ almost everywhere.

Lemma 3 (Log-concave flux). *There is a universal $C < \infty$ such that every isotropic log-concave density g on $\mathbb{R}^N$ satisfies*

$$\mathcal{R}_0(g) \leq C, \quad \mathcal{R}_1(g) \leq C.$$

Only two-dimensional marginals are needed. For orthogonal u, v , the density of $(\langle X, u \rangle, \langle X, v \rangle)$ is isotropic and log-concave. A fixed-dimensional envelope gives $g_{u,v}(y, z) \leq C_1 e^{-c_1 \sqrt{y^2 + z^2}}$ by the two-dimensional case of Lemma 2 in Balcan and Long (2013), based on Lovász and Vempala (2007). Integrating in z controls both the slice mass and its first tangential moment. The constants do not depend on N .

Theorem 4 (Dominated reference law). *Let ν be log-concave with mean m and positive covariance Σ . Suppose the coefficient law μ is A -dominated by ν . For absolutely continuous $\psi : I \rightarrow \mathbb{R}$ and $F : I \rightarrow \mathbb{R}^N$ with $F(t) \neq 0$, let*

$$\phi_\alpha(t) = \psi(t) + \langle \alpha, F(t) \rangle,$$

define

$$\begin{aligned} q(t) &= \|\Sigma^{1/2}F(t)\|_2, & u(t) &= \frac{\Sigma^{1/2}F(t)}{q(t)}, \\ a(t) &= \frac{\psi(t) + \langle m, F(t) \rangle}{q(t)}. \end{aligned} \quad (11)$$

Then for every interval I ,

$$\mathbb{P}\{\exists t \in I : \phi_\alpha(t) = 0\} \leq CA\{\text{TV}_I(a) + \text{Len}_I(u)\}. \quad (12)$$

This is coordinate-free. It remains valid for nonsmooth coefficient densities and arbitrary dependence. The reference law supplies the geometry, while A measures the allowed departure from it.

The normalization has two useful invariances. An invertible affine change of coefficient coordinates is absorbed by m and Σ . Also, multiplying both F and ψ by a positive scalar function changes neither a nor u . A locally nonzero scalar changes both only by its locally constant sign. In the centered homogeneous case, the pointwise speed is

$$\|u'(t)\|_2 = \frac{\|(I - u(t)u(t)^\top)\Sigma^{1/2}F'(t)\|_2}{q(t)}. \quad (13)$$

This is the component of feature motion that rotates the decision hyperplane. Radial growth of F is irrelevant, which is why a raw bound on $\|F'\|$ is not the right condition.

3.1 Bounded joint density on a cube

Suppose μ is supported on $[-R, R]^N$ and has density at most κ . It is $A = \kappa(2R)^N$ dominated by the uniform law on the cube, whose covariance is $(R^2/3)I$. Put $\hat{\psi} = \psi/\|F\|_2$ and $\hat{F} = F/\|F\|_2$. Theorem 4 gives

$$\mathbb{P}\{\phi_\alpha \text{ has a zero in } I\} \leq CA \left\{ \frac{\sqrt{3}}{R} \text{TV}_I(\hat{\psi}) + \text{Len}_I(\hat{F}) \right\}. \quad (14)$$

There is no explicit dimension or support-radius penalty beyond the normalized density A and affine offset. In the homogeneous case $\psi = 0$, spherical feature length is the only conditioning quantity.

Corollary 5 (Normalized Pfaffian boundary). *Let $F = (F_1, \dots, F_N)$ be a nonvanishing Pfaffian feature vector on an interval J . Under the cube law above, if $F/\|F\|_2$ is L_F -Lipschitz, then*

$$\mathbb{P}\{\exists t \in I : \langle \alpha, F(t) \rangle = 0\} \leq CAL_F|I|$$

for every $I \subseteq J$.

This proves that the normalized Lipschitz condition proposed by Balcan et al. (2026) is sufficient. Pfaffian complexity is relevant for bounding L_F and the path-hitting VC dimension, but not for the anti-concentration step itself.

4 Independent Marginals

Domination by the uniform cube can be expensive for product laws because $A \leq (2R\kappa)^N$. Independence instead gives direct projection anti-concentration.

Theorem 6 (Independent coefficients). *Suppose $\alpha_1, \dots, \alpha_N$ are independent, supported on $[-R, R]$, and have densities at most κ . Let $a : I \rightarrow \mathbb{R}$ and $u : I \rightarrow \mathbb{S}^{N-1}$ be L_a - and L_u -Lipschitz. Then*

$$\begin{aligned} \mathbb{P}\{\exists t \in I : a(t) + \langle \alpha, u(t) \rangle = 0\} \\ \leq \sqrt{2}\kappa(L_a + R\sqrt{N}L_u)|I|. \end{aligned} \quad (15)$$

The density of every unit projection of α is at most $\sqrt{2}\kappa$ by the sharp Rogozin–Ball comparison (Rudelson and Vershynin, 2015). If a root occurs, evaluating the boundary at the midpoint of I places that fixed projection in an interval of width at most $(L_a + R\sqrt{N}L_u)|I|$. This proves (15) without Kac–Rice regularity.

Neither independence nor a projection condition can be omitted. Let $r = 1/4$, draw $Z \sim \text{Unif}[-1/4, 1/4]$, and set

$$\alpha_1 = Z, \quad \alpha_0 = -r^2 - rZ.$$

Both marginals have bounded densities and support, but $t^2 + \alpha_1 t + \alpha_0$ has the root r almost surely. Thus coordinatewise marginal bounds alone permit an atom in the root distribution.

5 Sharp Polynomial Bounds

For the monic polynomial, set

$$\begin{aligned} F(t) &= (1, t, \dots, t^{d-1}), \quad u(t) = \frac{F(t)}{\|F(t)\|_2}, \\ a_0(t) &= \frac{t^d}{\|F(t)\|_2}. \end{aligned}$$

Lemma 7 (Monomial conditioning). *Globally on $\mathbb{R}$,*

$$\text{Lip}(u) \leq d - 1, \quad \text{Lip}(a_0) \leq \sqrt{d}.$$

The first inequality follows by projecting F' orthogonally to F . For the second, on $[0, 1]$ monotonicity gives $a'_0 \leq dt^{d-1}/\|F\|_2 \leq \sqrt{d}$. For $t \geq 1$, let $y = t^{-2}$ and $Q(y) = 1 + \dots + y^{d-1}$. Then

$$a'_0(t) = \frac{\sum_{j=0}^{d-1} (j+1)y^j}{Q(y)^{3/2}} \leq \sqrt{Q(y)} \leq \sqrt{d},$$

because the first d coefficients of Q^2 are $j+1$. Symmetry handles negative t .

Corollary 8 (Monic polynomials). *For coefficients supported on $[-R, R]^d$ with joint density at most κ , write $A = \kappa(2R)^d$. Then*

$$\mathbb{P}\{P_\alpha \text{ has a root in } I\} \leq CA \left(d - 1 + \frac{\sqrt{3d}}{R} \right) |I|. \quad (16)$$

If the coordinates are independent and each marginal density is at most κ , then

$$\mathbb{P}\{P_\alpha \text{ has a root in } I\} \leq \sqrt{2d}\kappa\{1 + R(d-1)\}|I|. \quad (17)$$

Equation (16) improves the leading d^2 dependence in (1) to d . Equation (17) answers the polynomial-bound part of Open Question 1 without an exponentially large joint density. The next result shows that its leading rate is minimax sharp.

First, the linear degree in the log-concave result cannot be improved.

Proposition 9 (Gaussian monic lower bound). *Let $\alpha_0, \dots, \alpha_{d-1}$ be iid standard Gaussian and $d \geq 2$. Let H_ϵ denote the event that P_α has a root in $[1, 1 + \epsilon]$. Then*

$$\lim_{\epsilon \downarrow 0} \frac{\mathbb{P}(H_\epsilon)}{\epsilon} \geq \frac{e^{-1/(2d)}\sqrt{d^2 - 1}}{\pi\sqrt{12}}. \quad (18)$$

In particular, the local root constant is $\Omega(d)$.

To see the scale, $(P_\alpha(1), P'_\alpha(1))$ is jointly Gaussian. Given $P_\alpha(1) = 0$, the derivative has mean $(d+1)/2$ and variance $d(d^2 - 1)/12$. The Kac–Rice density and $\mathbb{E}|\mathcal{N}(m, \sigma^2)| \geq \sigma\sqrt{2/\pi}$ give (18). The supplement justifies the passage from root intensity to local hit probability.

Theorem 10 (Product-law lower bound). *There are universal $c, c' > 0$ such that for every even $d \geq 4$, $R \geq 1$, and $\kappa R \geq 4$, one can choose independent coefficient laws supported on $[-R, R]$, each with density κ , for which*

$$\mathbb{P}\{P_\alpha \text{ has a root in } [1, 1 + \epsilon]\} \geq c\kappa R d^{3/2} \epsilon \quad (19)$$

for all sufficiently small $\epsilon \leq c'/d$.

The construction shifts the lower half of the coefficients near $-R/2$ and the upper half near $R/2$. At $t = 1$, the polynomial value is a centered sum of d uniform noises, with density $\Omega(\kappa/\sqrt{d})$ at zero. Its derivative has deterministic size $\Omega(Rd^2)$. A one-sided value interval therefore forces a crossing with probability $\Omega(\kappa R d^{3/2} \epsilon)$. The proof is direct and does not infer a hit probability from an expected root count.

Here is the quantitative core. Let d be even, $M = R/2$, and $\delta = (2\kappa)^{-1}$. Center coefficients with indices below $d/2$ at $-M - 1/d$ and the rest at $M - 1/d$, then add independent $\text{Unif}[-\delta, \delta]$ noises ξ_i . The assumptions place every coefficient in $[-R, R]$. At one,

$$P_\alpha(1) = S := \sum_{i=0}^{d-1} \xi_i, \\ P'_\alpha(1) = \frac{Md^2}{4} + \frac{d+1}{2} + \sum_{i=0}^{d-1} i\xi_i \geq \frac{Rd^2}{16}.$$

The symmetric unimodal density q_d of S obeys $q_d(0) \geq 3\sqrt{3}/(16\delta\sqrt{d})$. For sufficiently small ϵ , a negative interval of width $\Theta(Rd^2\epsilon)$ for S forces a crossing before $1 + \epsilon$. Multiplying width by density gives (19). The supplement controls the derivative throughout the interval and verifies the support constants.

Linear spherical-speed dependence is also unavoidable. If $Q(t) = \alpha_0 + \alpha_1 t^d$ with iid $\text{Unif}[-1, 1]$ coefficients, direct integration gives, for $q = (1 + \epsilon)^d$,

$$\mathbb{P}\{Q \text{ has a root in } [1, 1 + \epsilon]\} = \frac{q - 1}{4q}, \\ \lim_{\epsilon \downarrow 0} \frac{\mathbb{P}(\text{root})}{\epsilon} = \frac{d}{4}. \quad (20)$$

6 Sharp Online Polynomial Dispersion

The independent-coordinate bound gives a complete online consequence without an unspecified feature condition. Recall that losses $\ell_1, \dots, \ell_T$ are f -point-dispersed when

$$\mathbb{E} \max_{\substack{\rho, \rho' \\ \|\rho - \rho'\|_2 \leq \epsilon}} |\{t : |\ell_t(\rho) - \ell_t(\rho')| > L\|\rho - \rho'\|_2\}| \\ \leq f(T, \epsilon). \quad (21)$$

Corollary 11 (Sharp polynomial dispersion). *Let $\Theta \subset \mathbb{R}$ be a compact interval. In each of T independent rounds, suppose a $[0, 1]$ -valued loss is piecewise L -Lipschitz and its discontinuities lie among the roots of at most K monic degree- d polynomials. Within each polynomial, assume the d nonleading coefficients are independent, supported on $[-R, R]$, and have densities at most κ . Put*

$$B_d = \sqrt{2d}\kappa\{1 + R(d-1)\}.$$

Then

$$f(T, \epsilon) \leq K B_d T \epsilon + \tilde{O}_{d,K}(\sqrt{T}). \quad (22)$$

In the full-information setting, on a domain of radius D with an r_0 -interior comparator, continuous exponential weights has, for $0 < r \leq r_0$, regret

$$O\left(\sqrt{T \log(D/r)} + f(T, r) + T L r\right). \quad (23)$$

Thus $r = T^{-1/2}$ gives

$$\tilde{O}\left([1 + L + K\kappa\{\sqrt{d} + Rd^{3/2}\}]\sqrt{T}\right).$$

Conversely, if Θ contains $[1, 1 + c'/d]$, then for even $d \geq 4$, $R \geq 1$, and $\kappa R \geq 4$, there are iid rounds with $K = 1$ and piecewise-constant losses such that

$$f(T, \epsilon) \geq c\kappa R d^{3/2} T \epsilon \quad (24)$$

for all sufficiently small $\epsilon \leq c'/d$.

For a fixed interval of length ϵ , Corollary 8 and a union bound give expected crossing count at most $KB_d T \epsilon$. The polynomial path-hitting VC bound of Balcan et al. (2025) makes the interval uniform at additive $\tilde{O}_{d,K}(\sqrt{T})$ cost, which proves (22). Equation (23) is the dispersed continuous-exponential-weights theorem of Balcan et al. (2018). For the lower bound, use iid copies of Theorem 10 and set $\ell_\alpha(\rho) = \mathbf{1}\{P_\alpha(\rho) \geq 0\}$. Its construction has $P_\alpha(1) < 0 \leq P_\alpha(1+\epsilon)$ on the forcing event used in the proof, whose probability has the stated lower bound. The fixed pair $(1, 1+\epsilon)$ therefore contributes the right side of (24) in expectation. Hence the leading dispersion scale, not only the root bound, is minimax sharp.

7 Beyond Polynomials and Online Regret

For exponential features $F_i(t) = c_i e^{b_i t}$, write $u = F/\|F\|_2$ and $w_i = u_i^2$. Differentiation gives

$$u'_i = (b_i - \bar{b})u_i, \\ \|u'\|_2^2 = \text{Var}_w(b) \leq \frac{(b_{\max} - b_{\min})^2}{4}.$$

Thus (14) yields a root constant $CA(b_{\max} - b_{\min})/2$. This makes explicit the dependence suppressed in the earlier finite bound for random exponential sums (Balcan and Sharma, 2021). The same calculation applies to any feature family through its normalized spherical speed.

The calculation directly instantiates the Gaussian-kernel graph family in Balcan and Sharma (2021). Its label changes occur at roots of exponential polynomials generated by the graph objective.

Corollary 12 (Exponential-kernel graph learning). *Consider the one-parameter graph-based semi-supervised losses of Balcan and Sharma (2021). Suppose each round has at most K transition equations of the form*

$$\sum_{i=1}^N \alpha_i e^{b_i \rho} = 0, \quad b_i \in [b_{\min}, b_{\max}],$$

and the coefficient vector in each equation is supported on $[-R, R]^N$ with density at most κ . Let $A = \kappa(2R)^N$, assume rounds are independent, and put $B = b_{\max} - b_{\min}$. Then

$$f(T, \epsilon) \leq \frac{CAKB}{2} T \epsilon + \tilde{O}_{N,K}(\sqrt{T}). \quad (25)$$

In the full-information setting, continuous exponential weights consequently has

$$\tilde{O}([1 + L + AKB]\sqrt{T})$$

regret for fixed domain radius and interiority. For a Gaussian graph kernel parameterized by $\rho = \sigma^{-2}$, B is the range of the exponents induced by the squared graph distances. Thus the previously suppressed transition conditioning is an explicit instance quantity. If the coordinates of every coefficient vector are instead independent with marginal densities at most κ and support $[-R, R]$, the factor $CA/2$ in (25) improves to $\kappa R \sqrt{N/2}$. The regret is then

$$\tilde{O}([1 + L + \kappa RKB\sqrt{N}]\sqrt{T}),$$

which is polynomial in every displayed parameter.

Indeed, the uniform cube is a log-concave reference and the exponential feature calculation gives $\text{Len}_I(u) \leq B|I|/2$. A union bound over equations and the one-dimensional path-hitting VC theorem prove (25). The dispersed online-optimization theorem with $r = T^{-1/2}$ gives the regret statement. The independent-coordinate version uses Theorem 6 with $L_a = 0$ and $L_u = B/2$.

We state the online consequence in the notation of Balcan et al. (2025). Suppose each round has at most K transition hypersurfaces of fixed Pfaffian degree, function degree, and common chain length. Along every axis-aligned path segment s of length $|s|$, assume the restriction of each whitened, normalized boundary satisfies

$$\text{TV}_s(a) + \text{Len}_s(u) \leq \Gamma|s|. \quad (26)$$

Corollary 13 (Dispersion and regret). *Assume rounds are independent and every coefficient law is A -dominated by its log-concave reference. In p parameter dimensions, the sequence is f -point-dispersed with*

$$f(T, \epsilon) \leq CAKT\sqrt{p}T\epsilon + \tilde{O}(\sqrt{T}). \quad (27)$$

For losses in $[0, 1]$ that are piecewise L -Lipschitz on a domain of radius D with an r_0 -interior comparator, continuous exponential weights has regret, for $0 < r \leq r_0$,

$$O\left(\sqrt{pT \log(D/r)} + f(T, r) + T L r\right). \quad (28)$$

In particular, $r = T^{-1/2}$ gives $\tilde{O}(\sqrt{T})$ regret when the displayed structural parameters are fixed.

A path joining two points at Euclidean distance ϵ has at most p axis-aligned segments and total length at most $\sqrt{p}\epsilon$. Let $D(T, s)$ count rounds whose discontinuity set meets such a path. For a fixed path, Theorem 4, condition (26), and a union bound over K surfaces give

$$\mathbb{E}D(T, s) \leq CAKT\sqrt{p}T\epsilon.$$

The Pfaffian path-hitting VC theorem of Balcan et al. (2025) gives

$$\mathbb{E} \sup_s D(T, s) \leq \sup_s \mathbb{E}D(T, s) + \tilde{O}(\sqrt{T}),$$

where the suppressed constant depends on the fixed Pfaffian complexities, p , and K . A Lipschitz violation between two points implies a boundary crossing on their path, so (21) now gives (27). Equation (28) is the existing dispersed online-optimization theorem. Our new input is the coefficient-to-crossing bound. The same domination formulation is compatible with adaptive smoothed adversaries through existing coupling reductions (Haghtalab et al., 2024).

8 Related Work and Scope

Expected zero counts for Gaussian and general coefficient laws have a long history. The spherical-length formula of Edelman and Kostlan (1995) is the closest geometric precursor to our homogeneous flux. Exact correlation functions for independent continuous polynomial coefficients follow from coarea (Götze et al., 2018). Rice inequalities for independent random linear combinations bound expected total zeros through coordinate variation (Briquel and Bürgisser, 2020). Recent work studies real zeros of fewnomials, sparse polynomials, and Muntz polynomials (Bürgisser et al., 2019; Jindal et al., 2020; Lubinsky and Pritsker, 2025). These themes continue in concentration inequalities for Kac and general random polynomials (Can and Nguyen, 2025; Aguirre et al., 2025a), dependent trigonometric coefficients (Angst et al., 2024), and root separation and noncentered fluctuations (Michelen and Yakir, 2025a; Do et al., 2026). Pending work refines the nonuniversal constant in expected Kac root counts, expectation and variance for Weyl polynomials, and microscopic local universality (Lam and Nguyen, 2025; Aguirre et al., 2025b; Michelen and Yakir, 2025b). We read these results as complementary. The closest local-location results establish a non-Poissonian $1/n$ real-root scale near the unit circle and $1/n$ hole radii for Kac polynomials and derivatives (Michelen, 2021; Nguyen and Nguyen, 2024). They concern iid asymptotic counts, correlations, separation, or hole scales, not a finite-degree distribution-class constant for the probability of any root in every interval. Expected counts can upper bound a hit probability, but their assumptions and asymptotics differ from the local coefficient-smoothness question here.

Nonlinear small-ball theory instead controls $\mathbb{P}\{|g(X) - z| \leq \epsilon\}$ for a fixed polynomial map of random input. Recent work gives dimension-free variance, small-ball, and Fourier criteria under symmetric log-concave laws (Glazer and Mikulincer, 2026). Those results do not directly control the existential event that a coefficient-linear family has a root somewhere along a moving parameter interval.

Projection density is classical in probability. We use

the sharp consequence of Rogozin’s convolution theorem and Ball’s cube slicing theorem in the form proved by Rudelson and Vershynin (2015). Log-concavity supplies dimension-free two-dimensional envelopes (Lovász and Vempala, 2007). The novelty is their use with normalized boundary geometry to obtain the sharp polynomial root bounds and online consequences.

Our bounds expose an unavoidable conditioning parameter. Formal Pfaffian degree alone cannot control root concentration, as a rescaled linear feature already shows (Balcan et al., 2026). Spherical speed is invariant to scalar rescaling of F and measures exactly the hyperplane motion that coefficients see. It may still be large for an ill-conditioned feature family. Any application must therefore bound Γ from its instance structure.

9 Conclusion

Coefficient flux gives an exact answer for universal normalized affine boundaries and turns natural coefficient models into sharp local root bounds. It settles the requested independent-coordinate polynomial rate and proves the proposed Pfaffian normalization sufficient while explaining the remaining conditioning requirement. The independent-coordinate theorem and its online dispersion consequence are minimax sharp, and the joint-density theorem removes a full factor of degree from the prior polynomial guarantee. More broadly, the result separates two tasks in online algorithm configuration: probability controls hyperplane flux, while problem geometry controls normalized feature length.

References

- Aguirre, A., Nguyen, H. H., and Wang, J. (2025a). Concentration of the number of real roots of random polynomials. *Electronic Journal of Probability*, 30.
- Aguirre, A., Nguyen, H. H., and Wang, J. (2025b). Real roots of random weyl polynomials with general coefficients: Expectation and variance. *arXiv preprint arXiv:2511.07735*.
- Angst, J., Nguyen, O., and Poly, G. (2024). Roots of random trigonometric polynomials with general dependent coefficients. *arXiv preprint arXiv:2409.15057*.
- Balcan, M.-F., Dick, T., and Pegden, W. (2020). Semi-bandit optimization in the dispersed setting. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence*, volume 124 of *Proceedings of Machine Learning Research*, pages 909–918.
- Balcan, M.-F., Dick, T., and Vitercik, E. (2018).

- Dispersion for data-driven algorithm design, online learning, and private optimization. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science*, pages 603–614.
- Balcan, M.-F. and Long, P. M. (2013). Active and passive learning of linear separators under log-concave distributions. In *Proceedings of the 26th Annual Conference on Learning Theory*, volume 30 of *Proceedings of Machine Learning Research*, pages 288–316.
- Balcan, M. F., Nguyen, A. T., and Sharma, D. (2025). Algorithm configuration for structured pfaffian settings. *Transactions on Machine Learning Research*.
- Balcan, M.-F., Pegden, W., and Sharma, D. (2026). Invited open problem: Online optimization of piecewise-lipschitz functions with applications to data-driven algorithm design. In *Proceedings of the Thirty-Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pages 7111–7116.
- Balcan, M.-F. F. and Sharma, D. (2021). Data driven semi-supervised learning. In *Advances in Neural Information Processing Systems*, volume 34.
- Briquel, I. and Bürgisser, P. (2020). The real tau-conjecture is true on average. *Random Structures and Algorithms*, 57(2):279–303.
- Bürgisser, P., Ergür, A. A., and Tonelli-Cueto, J. (2019). On the number of real zeros of random fewnomials. *SIAM Journal on Applied Algebra and Geometry*, 3(4):721–732.
- Can, V. H. and Nguyen, O. (2025). Concentration inequalities for the number of real zeros of Kac polynomials. *Electronic Journal of Probability*, 30:1–23.
- Do, Y. Q., Nguyen, N. D. V., and O’Rourke, S. (2026). Real roots of non-centered random polynomials. *arXiv preprint arXiv:2605.26402*.
- Edelman, A. and Kostlan, E. (1995). How many zeros of a random polynomial are real? *Bulletin of the American Mathematical Society*, 32(1):1–37.
- Feliu, E. and Sadeghimanesh, A. (2022). Kac-rice formulas and the number of solutions of parametrized systems of polynomial equations. *Mathematics of Computation*, 91(338):2739–2769.
- Glazer, I. and Mikulincer, D. (2026). Anti-concentration of polynomials: L^p balls and symmetric measures. *arXiv preprint arXiv:2603.22664*.
- Götze, F., Kaliada, D., and Zaporozhets, D. (2018). Correlation functions of real zeros of random polynomials. *Journal of Mathematical Sciences*, 229(6):664–670.
- Haghtalab, N., Roughgarden, T., and Shetty, A. (2024). Smoothed analysis with adaptive adversaries. *Journal of the ACM*, 71(3):19:1–19:34.
- Jindal, G., Pandey, A., Shukla, H., and Zisopoulos, C. (2020). How many zeros of a random sparse polynomial are real? In *Proceedings of the 45th International Symposium on Symbolic and Algebraic Computation*, pages 273–280.
- Lam, P. and Nguyen, O. (2025). On the first non-universal term in random polynomial real zeros. *arXiv preprint arXiv:2509.12170*.
- Lovász, L. and Vempala, S. S. (2007). The geometry of logconcave functions and sampling algorithms. *Random Structures and Algorithms*, 30(3):307–358.
- Lubinsky, D. S. and Pritsker, I. E. (2025). Zeros of random müntz polynomials. *Journal of Mathematical Analysis and Applications*, 552(2):129799.
- Michelen, M. (2021). Real roots near the unit circle of random polynomials. *Transactions of the American Mathematical Society*, 374(6):4359–4374.
- Michelen, M. and Yakir, O. (2025a). Limit law for root separation in random polynomials. *arXiv preprint arXiv:2505.02723*.
- Michelen, M. and Yakir, O. (2025b). A simple proof of local universality for roots of Kac polynomials. *arXiv preprint arXiv:2511.21455*.
- Nguyen, H. H. and Nguyen, O. (2024). Hole radii for the Kac polynomials and derivatives. *Stochastic Processes and their Applications*, 175:104386.
- Rudelson, M. and Vershynin, R. (2015). Small ball probabilities for linear images of high-dimensional distributions. *International Mathematics Research Notices*, 2015(19):9594–9617.

A Pointers to the Supplement

The supplement proves the area formula and exact flux characterization, log-concave domination, the

independent-marginal theorem, both monomial conditioning inequalities, the matching product lower bound, the exact two-term calculation, and the online corollary.

B Radon Flux and the Area Formula

We first make the measure-theoretic convention in the main paper explicit. For orthogonal unit vectors u, v , let

$$p_u(s) = \int_{\langle u, x \rangle = s} f(x) d\mathcal{H}^{N-1}(x),$$

$$r_{u,v}(s) = \int_{\langle u, x \rangle = s} |\langle v, x \rangle| f(x) d\mathcal{H}^{N-1}(x).$$

These are Radon derivatives defined for almost every s . We use

$$\mathcal{R}_0(f) = \sup_u \|p_u\|_\infty, \quad \mathcal{R}_1(f) = \sup_{u \perp v} \|r_{u,v}\|_\infty.$$

This definition is independent of the representative of f . Equivalently,

$$\mathcal{R}_0(f) = \sup_{u, J} \frac{\mathbb{P}\{\langle u, X \rangle \in J\}}{|J|}, \quad (29)$$

$$\mathcal{R}_1(f) = \sup_{u \perp v, J} \frac{\mathbb{E}[|\langle v, X \rangle| \mathbf{1}\{\langle u, X \rangle \in J\}]}{|J|}, \quad (30)$$

where the suprema range over bounded intervals of positive length. These small-slab formulas also specify the values when a Radon derivative has an exceptional representative.

We will first use the area formula for a smooth density, then remove this temporary regularity by mollification. Let γ_ϵ be the centered Gaussian density with covariance $\epsilon^2 I$ and let $f_\epsilon = f * \gamma_\epsilon$.

Lemma 14 (Flux-preserving mollification). *For every $\epsilon > 0$,*

$$\mathcal{R}_0(f_\epsilon) \leq \mathcal{R}_0(f), \quad \mathcal{R}_1(f_\epsilon) \leq \mathcal{R}_1(f) + \epsilon \sqrt{2/\pi} \mathcal{R}_0(f).$$

Moreover, $f_\epsilon \rightarrow f$ in $L^1(\mathbb{R}^N)$.

Proof. Let X have density f , let G be an independent standard Gaussian, and fix orthogonal unit vectors u, v . Put $Y = \langle u, X \rangle$ and $Z = \langle v, X \rangle$. The projection density of $X + \epsilon G$ is the convolution of p_u with a one-dimensional Gaussian density φ_ϵ . Hence its supremum is at most $\mathcal{R}_0(f)$. Independence of the two Gaussian projections and $\mathbb{E}|Z + \epsilon \langle v, G \rangle| \leq |Z| + \epsilon \sqrt{2/\pi}$ give, pointwise in s ,

$$\begin{aligned} r_{u,v}^{(\epsilon)}(s) &\leq (r_{u,v} * \varphi_\epsilon)(s) + \epsilon \sqrt{2/\pi} (p_u * \varphi_\epsilon)(s) \\ &\leq \mathcal{R}_1(f) + \epsilon \sqrt{2/\pi} \mathcal{R}_0(f). \end{aligned}$$

Taking suprema proves the flux bounds. Gaussian kernels form an approximate identity, which proves L^1 convergence. $\square$

B.1 The swept-hyperplane inequality

Let $I = [c, d]$. Reparameterizing (a, u) by its accumulated variation gives a Lipschitz curve with the same sweep, total variation, and spherical length. Choose an orthonormal basis $e_1(c), \dots, e_{N-1}(c)$ of $u(c)^\perp$ and extend it by parallel transport

$$e'_j(t) = -\langle u'(t), e_j(t) \rangle u(t). \quad (31)$$

Differentiating inner products shows that the $e_j(t)$ remain orthonormal and perpendicular to $u(t)$. If $Q(t)$ has these columns, the map

$$\Phi(t, z) = -a(t)u(t) + Q(t)z$$

parameterizes the hyperplane $H_t = \{x : \langle u(t), x \rangle = -a(t)\}$.

Lemma 15 (Sweep Jacobian). *For almost every (t, z) ,*

$$J_\Phi(t, z) = |a'(t) + \langle u'(t), \Phi(t, z) \rangle|.$$

Consequently,

$$\begin{aligned} & \int_{\cup_{t \in I} H_t} f(x) dx \\ & \leq \int_I \int_{H_t} |a'(t) + \langle u'(t), x \rangle| f(x) d\mathcal{H}^{N-1}(x) dt. \end{aligned} \quad (32)$$

Proof. The z derivatives of Φ are the columns of Q . Equation (31) gives

$$\partial_t \Phi = -(a' + \langle u', Qz \rangle)u - au'.$$

Only its component normal to the columns of Q contributes to the N -dimensional Jacobian. Since $\langle u', \Phi \rangle = \langle u', Qz \rangle$, the displayed formula follows. Apply the area formula on $I \times B(0, M)$ and let M increase. The area formula counts image points with their multiplicity. Discarding multiplicity gives (32) by monotone convergence. $\square$

For $u'(t) \neq 0$, put $v(t) = u'(t)/\|u'(t)\|_2$. The identity $\langle u, u' \rangle = 0$ implies $v(t) \perp u(t)$. Lemma 15 and the triangle inequality give

$$\begin{aligned} & \mathbb{P}\{\exists t \in I : a(t) + \langle X, u(t) \rangle = 0\} \\ & \leq \mathcal{R}_0(f) \int_I |a'(t)| dt \\ & \quad + \mathcal{R}_1(f) \int_I \|u'(t)\|_2 dt. \end{aligned}$$

For the smooth density f_ϵ , all Radon slices in the last display are the continuous convolution versions, so their pointwise values are bounded by the fluxes. Lemma 14 therefore gives

$$\begin{aligned} \mathbb{P}_{f_\epsilon}\{X \in \mathcal{A}_I(a, u)\} & \leq \mathcal{R}_0(f) \text{TV}_I(a) + \mathcal{R}_1(f) \text{Len}_I(u) \\ & \quad + \epsilon \sqrt{2/\pi} \mathcal{R}_0(f) \text{Len}_I(u). \end{aligned}$$

The probabilities converge to the probability under f because $f_\epsilon \rightarrow f$ in L^1 . Letting ϵ decrease to zero proves the upper bound in the affine flux theorem for every density. If either flux is infinite the statement is immediate.

B.2 Sharpness of both fluxes

We next prove that neither term can be reduced for a theorem uniform over all normalized affine boundaries.

For any unit u , any $s \in \mathbb{R}$, and any $h > 0$, take $I = [0, h]$, $u(t) = u$, and $a(t) = -(s + t)$. The boundary has a root exactly when $\langle u, X \rangle \in [s, s + h]$. Its denominator is h . Taking the supremum in (29) proves

$$C_{\text{univ}}(f) \geq \mathcal{R}_0(f). \quad (33)$$

For the rotational term, fix $u \perp v$ and define

$$u_\theta = \cos(\theta)u + \sin(\theta)v, \quad a(\theta) = -s, \quad 0 \leq \theta \leq h.$$

The spherical length is h . Let $(Y, Z) = (\langle u, X \rangle, \langle v, X \rangle)$ and write the root event as

$$E_h(s) = \{\exists \theta \in [0, h] : Y \cos \theta + Z \sin \theta = s\}.$$

We prove necessity directly from the slab form, without selecting a Radon representative. Fix a bounded interval J and $M < \infty$. Let $E_h^M(s)$ be $E_h(s)$ intersected with $\{|Y| \leq M, |Z| \leq M\}$. Every offset s defines an admissible

boundary, so

$$\begin{aligned} C_{\text{univ}}(f) &\geq \sup_s \frac{\mathbb{P}(E_h(s))}{h} \\ &\geq \frac{1}{h|J|} \int_J \mathbb{P}(E_h^M(s)) ds. \end{aligned} \quad (34)$$

For fixed (Y, Z) , let

$$S_h(Y, Z) = \{Y \cos \theta + Z \sin \theta : 0 \leq \theta \leq h\}.$$

This is an interval. Fubini's theorem turns the last integral in (34) into

$$\mathbb{E}[|S_h(Y, Z) \cap J| \mathbf{1}\{|Y| \leq M, |Z| \leq M\}].$$

Except when Y is an endpoint of J , an event of probability zero, scalar differentiation at $\theta = 0$ gives

$$\lim_{h \downarrow 0} \frac{|S_h(Y, Z) \cap J|}{h} = |Z| \mathbf{1}\{Y \in J\}.$$

The ratio is at most $|Y| + |Z| \leq 2M$, since the length of the image is at most the integral of the absolute derivative. Dominated convergence in (34) yields

$$C_{\text{univ}}(f) \geq \frac{\mathbb{E}[|Z| \mathbf{1}\{Y \in J, |Y| \leq M, |Z| \leq M\}]}{|J|}.$$

Letting M increase and then taking the supremum over J proves

$$C_{\text{univ}}(f) \geq \mathcal{R}_1(f). \quad (35)$$

Together with the upper bound, (33) and (35) prove the exact identity in the affine flux theorem.

C Log-Concave Domination

Proof of the log-concave flux lemma. Every linear marginal of an isotropic log-concave law is isotropic and log-concave. A one-dimensional isotropic log-concave density has universally bounded essential supremum. This gives $\mathcal{R}_0(g) \leq C$.

For $u \perp v$, let $g_{u,v}$ be the density of $(\langle X, u \rangle, \langle X, v \rangle)$. It is an isotropic log-concave density on $\mathbb{R}^2$. Standard fixed-dimensional log-concave estimates give universal constants $C_1, c_1 > 0$ such that

$$g_{u,v}(y, z) \leq C_1 \exp(-c_1 \sqrt{y^2 + z^2}).$$

Therefore

$$\sup_y \int_{\mathbb{R}} |z| g_{u,v}(y, z) dz \leq C_1 \int_{\mathbb{R}} |z| e^{-c_1 |z|} dz \leq C.$$

The constants are independent of N, u, v . The required fixed-dimensional envelope is the $d = 2$ case of Lemma 2 in Balcan and Long (2013), based on the isotropic log-concave bounds of Lovász and Vempala (2007). $\square$

Proof of the dominated reference theorem. Write

$$X = \Sigma^{-1/2}(\alpha - m).$$

The reference law of X is isotropic and log-concave. The law induced by μ has density at most A times this reference density. Both Radon fluxes are therefore at most CA by the log-concave flux lemma. Also

$$\begin{aligned} \psi(t) + \langle \alpha, F(t) \rangle &= \psi(t) + \langle m, F(t) \rangle + \langle X, \Sigma^{1/2} F(t) \rangle \\ &= q(t) \{a(t) + \langle X, u(t) \rangle\}. \end{aligned}$$

Since $q(t) > 0$, the original and normalized boundaries have the same roots. The affine flux theorem proves the claim. $\square$

Proof of the cube and Pfaffian consequences. The uniform law on $[-R, R]^N$ has mean zero, covariance $(R^2/3)I$, and density $(2R)^{-N}$. A density bounded by κ on this cube is $A = \kappa(2R)^N$ dominated by the uniform law. Hence

$$\begin{aligned} q(t) &= \frac{R}{\sqrt{3}} \|F(t)\|_2, & u(t) &= \frac{F(t)}{\|F(t)\|_2}, \\ a(t) &= \frac{\sqrt{3}}{R} \frac{\psi(t)}{\|F(t)\|_2}. \end{aligned}$$

Substitution in the dominated reference theorem proves the cube bound. If $\psi = 0$ and $F/\|F\|_2$ is L_F -Lipschitz, its length on I is at most $L_F|I|$. This proves the normalized Pfaffian corollary. $\square$

D Independent Marginals

Proof of the independent-coefficients theorem. Let t_0 be the midpoint of I . If $h_\alpha(t) = a(t) + \langle \alpha, u(t) \rangle$ vanishes at $t \in I$, then

$$\begin{aligned} |h_\alpha(t_0)| &\leq L_a|t - t_0| + \|\alpha\|_2 L_u|t - t_0| \\ &\leq \frac{1}{2}(L_a + R\sqrt{N}L_u)|I|. \end{aligned}$$

Thus the fixed projection $\langle \alpha, u(t_0) \rangle$ lies in an interval of width at most $(L_a + R\sqrt{N}L_u)|I|$. The sharp one-dimensional case of the Rogozin–Ball comparison states that every unit projection of independent variables with marginal densities bounded by κ has density at most $\sqrt{2}\kappa$ (Rudelson and Vershynin, 2015). Multiplying density by interval width proves the claimed bound. $\square$

The dependence counterexample in the main paper is immediate from

$$r^2 + \alpha_1 r + \alpha_0 = r^2 + Zr - r^2 - rZ = 0.$$

The marginal of α_1 has density 2 on an interval of length $1/2$. The marginal of $\alpha_0 = -r^2 - rZ$ has density 8 on an interval of length $1/8$. Both supports lie in a fixed compact interval, although r is a root with probability one.

E Conditioning of the Monomial Curve

Let $F(t) = (1, t, \dots, t^{d-1})$ and $S(t) = \|F(t)\|_2^2$.

Spherical part of monomial conditioning. For $u = F/\|F\|_2$,

$$\begin{aligned} \|u'(t)\|_2 &= \frac{\|(I - u(t)u(t)^\top)F'(t)\|_2}{\|F(t)\|_2} \\ &\leq \frac{\|F'(t)\|_2}{\|F(t)\|_2}. \end{aligned}$$

If $|t| \leq 1$, then

$$\begin{aligned} \|F'(t)\|_2^2 &= \sum_{i=1}^{d-1} i^2 |t|^{2i-2} \\ &\leq (d-1)^2 \sum_{j=0}^{d-2} |t|^{2j} \leq (d-1)^2 \|F(t)\|_2^2. \end{aligned}$$

If $|t| \geq 1$, then

$$\begin{aligned} \|F'(t)\|_2^2 &\leq (d-1)^2 \sum_{i=1}^{d-1} |t|^{2i-2} \\ &\leq (d-1)^2 \frac{\|F(t)\|_2^2}{t^2} \leq (d-1)^2 \|F(t)\|_2^2. \end{aligned}$$

Hence $\|u'(t)\|_2 \leq d-1$ globally. $\square$

Affine part of monomial conditioning. It is enough by symmetry to take $t \geq 0$. On $0 \leq t \leq 1$,

$$a'_0(t) = \frac{dt^{d-1}}{\sqrt{S}} - \frac{t^d S'}{2S^{3/2}} \leq \frac{dt^{d-1}}{\sqrt{S}}.$$

Since $S(t) \geq dt^{2d-2}$ on this interval, $a'_0(t) \leq \sqrt{d}$. For $t \geq 1$, put $y = t^{-2}$ and $Q(y) = 1 + y + \dots + y^{d-1}$. Direct differentiation gives

$$a'_0(t) = \frac{\sum_{j=0}^{d-1} (j+1)y^j}{Q(y)^{3/2}}.$$

The coefficient of y^j in $Q(y)^2$ is $j+1$ for $0 \leq j \leq d-1$. Since $0 \leq y \leq 1$,

$$\sum_{j=0}^{d-1} (j+1)y^j \leq Q(y)^2.$$

It follows that $a'_0(t) \leq \sqrt{Q(y)} \leq \sqrt{d}$. The absolute derivative obeys the same bound for negative t . $\square$

Proof of the monic-polynomial corollary. For a joint density bounded by κ on the cube, the cube bound, the monomial conditioning lemma, and $A = \kappa(2R)^d$ give

$$\mathbb{P}(\text{root in } I) \leq CA \left(d-1 + \frac{\sqrt{3d}}{R} \right) |I|.$$

For independent marginals, apply the independent-coefficients theorem with $N = d$, $L_a = \sqrt{d}$, and $L_u = d-1$. This gives

$$\sqrt{2\kappa} \{ \sqrt{d} + R\sqrt{d}(d-1) \} |I|,$$

which is the independent-marginal polynomial bound. $\square$

F Matching Lower Bounds

F.1 The Gaussian monic lower bound

Let $S = P_\alpha(1)$ and $D = P'_\alpha(1)$. For iid standard Gaussian coefficients,

$$S = 1 + \sum_{i=0}^{d-1} \alpha_i, \quad D = d + \sum_{i=0}^{d-1} i\alpha_i.$$

Thus S has variance d and density

$$p_S(0) = \frac{e^{-1/(2d)}}{\sqrt{2\pi d}}.$$

The Gaussian conditioning formulas give

$$\begin{aligned} \mathbb{E}[D \mid S = 0] &= d - \frac{\sum_{i=0}^{d-1} i}{d} = \frac{d+1}{2}, \\ \text{Var}(D \mid S = 0) &= \sum_{i=0}^{d-1} i^2 - \frac{(\sum_{i=0}^{d-1} i)^2}{d} = \frac{d(d^2-1)}{12}. \end{aligned}$$

The Kac–Rice intensity at one is

$$\rho_d(1) = p_S(0) \mathbb{E}[|D| \mid S = 0].$$

For a normal variable, the expected absolute value is minimized at zero mean. Consequently,

$$\rho_d(1) \geq \frac{e^{-1/(2d)} \sqrt{d^2-1}}{\pi \sqrt{12}}.$$

It remains to pass from intensity to hit probability. Let K be a compact neighborhood of one and let $s \neq t$ lie in K . The covariance determinant of $(P_\alpha(s), P_\alpha(t))$ is

$$\det \text{Cov}(P_\alpha(s), P_\alpha(t)) = \Theta_d((s-t)^2).$$

Indeed, after dividing the second coordinate difference by $t-s$, the covariance converges to that of $(P_\alpha(s), P'_\alpha(s))$, which is nonsingular for $d \geq 2$. Hence the joint value density at zero is at most $C_d/|s-t|$.

Condition on $P_\alpha(s) = P_\alpha(t) = 0$. Rolle's theorem and Taylor's theorem give

$$|P'_\alpha(s)P'_\alpha(t)| \leq |s-t|^2 \sup_{x \in K} |P''_\alpha(x)|^2.$$

Gaussian regression, using the same nonsingular limiting value-derivative covariance, bounds the conditional expectation of the supremum on the right by a constant depending only on d and K . The two-point Kac–Rice intensity therefore satisfies

$$\rho_{2,d}(s, t) \leq C_d |s-t|.$$

If N_ϵ counts roots in $[1, 1+\epsilon]$, the one- and two-point Kac–Rice formulas now give

$$\begin{aligned} \mathbb{E}N_\epsilon &= \rho_d(1)\epsilon + o(\epsilon), \\ \mathbb{E}[N_\epsilon(N_\epsilon - 1)] &\leq C_d \int_1^{1+\epsilon} \int_1^{1+\epsilon} |s-t|, ds dt = O_d(\epsilon^3). \end{aligned}$$

Thus the difference between $\mathbb{E}N_\epsilon$ and $\mathbb{P}\{N_\epsilon \geq 1\}$ is $O_d(\epsilon^3)$. Therefore

$$\mathbb{P}\{N_{[1,1+\epsilon]} \geq 1\} = \rho_d(1)\epsilon + o(\epsilon),$$

which proves the Gaussian monic proposition.

F.2 A product law for monic polynomials

Proof of the product-law lower bound. Write $d = 2n$, set $M = R/2$, and let $\delta = 1/(2\kappa)$. For $0 \leq i < n$, define

$$m_i = -M - \frac{1}{d},$$

and for $n \leq i < d$, define

$$m_i = M - \frac{1}{d}.$$

Let ξ_i be independent and uniform on $[-\delta, \delta]$, and take $\alpha_i = m_i + \xi_i$. Each α_i has density κ on an interval of length $1/\kappa$. Since $\kappa R \geq 4$, $\delta \leq R/8$. Together with $R \geq 1$ and $d \geq 4$, this shows that all these intervals lie in $[-R, R]$.

At $t = 1$, cancellation of the centers gives

$$P_\alpha(1) = 1 + \sum_{i=0}^{d-1} m_i + \sum_{i=0}^{d-1} \xi_i = S_\xi, \quad S_\xi = \sum_{i=0}^{d-1} \xi_i.$$

The derivative at one is

$$\begin{aligned} P'_\alpha(1) &= d + \sum_{i=0}^{d-1} i(m_i + \xi_i) \\ &= \frac{Md^2}{4} + \frac{d+1}{2} + \sum_{i=0}^{d-1} i\xi_i. \end{aligned}$$

Since

$$\left| \sum_{i=0}^{d-1} i\xi_i \right| \leq \frac{\delta d^2}{2} \leq \frac{Rd^2}{16},$$

we have

$$P'_\alpha(1) \geq G, \quad G = \frac{Rd^2}{16}. \quad (36)$$

For $1 \leq t \leq 1 + c_0/d$, all coefficients have magnitude at most R , and

$$|P''_\alpha(t)| \leq C_0 R d^3$$

for a universal C_0 . Choose a universal $c' > 0$ small enough that $C_0 R d^3 (c'/d) \leq G/2$. Then for every $0 < \epsilon \leq c'/d$,

$$P'_\alpha(t) \geq G/2 \quad \text{on } [1, 1 + \epsilon]. \quad (37)$$

It remains to lower-bound the probability that S_ξ is in the crossing window. Its density q_d is symmetric and unimodal. Its standard deviation is $\sigma = \delta\sqrt{d/3}$. Chebyshev's inequality and unimodality imply

$$\frac{3}{4} \leq \mathbb{P}\{|S_\xi| \leq 2\sigma\} \leq 4\sigma q_d(0).$$

Hence

$$q_d(0) \geq \frac{3\sqrt{3}}{16\delta\sqrt{d}}. \quad (38)$$

The density is continuous for $d \geq 2$. Thus, for all sufficiently small ϵ , it is at least half the right side of (38) on $[-G\epsilon/2, 0]$. On the event

$$-G\epsilon/2 \leq S_\xi \leq -G\epsilon/4,$$

$P_\alpha(1) < 0$, while (37) gives $P_\alpha(1 + \epsilon) \geq 0$. The intermediate value theorem supplies a root. The event has probability at least

$$c \frac{1}{\delta\sqrt{d}} G\epsilon = c\kappa R d^{3/2} \epsilon.$$

This proves the theorem. The phrase ‘‘sufficiently small’’ only ensures that the crossing window lies in a neighborhood where q_d is bounded below. $\square$

F.3 Exact two-term calculation

Let α_0, α_1 be independent uniform variables on $[-1, 1]$ and $Q(t) = \alpha_0 + \alpha_1 t^d$. Put $q = (1 + \epsilon)^d$. A positive root lies in $[1, 1 + \epsilon]$ exactly when $-\alpha_0/\alpha_1 \in [1, q]$. In the quadrant $\alpha_1 > 0, \alpha_0 < 0$, write $x = \alpha_1$ and $y = -\alpha_0$. The relevant area in the unit square is

$$\int_0^{1/q} (qx - x)dx + \int_{1/q}^1 (1 - x)dx = \frac{q - 1}{2q}.$$

The opposite-sign quadrant contributes the same area. Dividing by the total square area 4 gives

$$\mathbb{P}\{Q \text{ has a root in } [1, 1 + \epsilon]\} = \frac{q - 1}{4q}.$$

Since $q = 1 + d\epsilon + o(\epsilon)$, the local coefficient is $d/4$.

G Exponential Features and Online Dispersion

For $F_i(t) = c_i e^{b_i t}$, define $u = F/\|F\|_2$ and $w_i = u_i^2$. With $\bar{b} = \sum_i w_i b_i$, differentiation gives

$$u'_i = (b_i - \bar{b})u_i, \quad \|u'\|_2^2 = \sum_i w_i (b_i - \bar{b})^2.$$

Popoviciu's variance inequality yields

$$\|u'\|_2 \leq \frac{b_{\max} - b_{\min}}{2}.$$

Substitution into the cube consequence of the dominated reference theorem proves the exponential feature claim. Finally consider the dispersion and regret corollary. An axis-aligned path between two points at Euclidean distance at most ϵ has total coordinate length at most $\sqrt{p}\epsilon$. On every segment, the dominated reference theorem and the path-conditioning assumption bound the expected number of crossed transition surfaces by

$$CAKT\sqrt{p}\epsilon.$$

Summing over T independent rounds gives the linear term in the dispersion bound. The Pfaffian path-hitting VC bound of Balcan et al. (2025) upgrades this fixed-path expectation to a uniform bound with additive $\tilde{O}(\sqrt{T})$. Substitution into the dispersed continuous exponential-weights theorem of Balcan et al. (2018) gives the stated regret bound. Setting $r = T^{-1/2}$ proves the stated rate.