Radial Phase Transitions for Gaussian Mixture NPMLEs in Fixed and Growing Dimensions

Abstract

Can a fitted Gaussian mixture distinguish a single Gaussian population from latent heterogeneity? Without regularization, the support of the mixture nonparametric maximum likelihood estimator (NPMLE) diverges even when the data come from a single Gaussian component. We ask whether fitting wider Gaussian components regularizes this overfitting. The observations are i.i.d. spherical Gaussians with covariance $(1-\varepsilon_n)I_d$, while the fitted mixture uses kernels with covariance I_d.

In dimension two, the transition occurs when $\varepsilon_n \log n$ is of constant order. The limiting profile is explicit and combines the largest point of a radial Poisson edge process with the aggregate contribution of observations just inside the edge. For every sequence $d_n \geq 3$, fixed or growing, a single exact Gamma-quantile normalization describes the transition. When the normalized coordinate tends to s, the NPMLE is unique with probability tending to one, and its support size minus one converges to $\mathrm{Pois}(e^{-s})$. Thus the one-atom probability tends to $\exp\{-e^{-s}\}$. For fixed $d \geq 3$, the Gamma-quantile normalization reduces to $\varepsilon_n = (d/2-1)(\log\log n)/\log n + O(1/\log n)$. Under a common coupling of the data, the support counts across each critical window converge jointly to a Poisson tail-count process. We also identify the fitted edge masses and, in fixed dimension, their directions.

Note. This draft was produced through an iterative process in which Mark Sellke repeatedly asked GPT-5.6 in OpenAI Codex to prove, refine, and write up further extensions.

Contents

1 Introduction

A natural way to look for latent heterogeneity is to fit a mixture and inspect the number of fitted components. For the Gaussian mixture nonparametric maximum likelihood estimator, this suggests a basic question: when the data come from one spherical Gaussian, does maximum likelihood select a single component mean? Our companion paper shows that it does not when the fitted component covariance matches the data-generating covariance [Ano26a]; in fact, the fitted support diverges in probability. Here we ask whether variance inflation can regularize this overfitting. We determine precisely how much inflation is needed in every dimension $d \geq 2$, including dimensions that grow with the sample size.

Let $d = d_n \geq 2$, let $Z_1, \ldots, Z_n$ be independent $\mathcal{N}(0, I_d)$ vectors, and observe

$$X_i = \sqrt{1 - \varepsilon_n}\, Z_i, \qquad 0 \leq \varepsilon_n < 1. \tag{1.1}$$

We fit location mixtures of $\mathcal{N}(\mu, I_d)$ densities, whose component covariance is therefore larger than the data covariance. For a probability measure G on $\mathbb{R}^d$, write

$$q_G(x) = \int_{\mathbb{R}^d} \phi_d(x - u)\, dG(u), \qquad \ell_n(G) = \sum_{i=1}^n \log q_G(X_i),$$

where ϕ_d is the $\mathcal{N}(0, I_d)$ density. A mixture NPMLE is any maximizer $\widehat{G}_n$ of ℓ_n. Because multivariate NPMLEs need not be unique [SGS25], we measure their support complexity by

$$\widehat{K}_n = \min\big\{|\operatorname{supp}(G)| : G \text{ maximizes } \ell_n\big\}. \tag{1.2}$$

Lindsay's support-reduction theorem gives a maximizer with at most n atoms, so the minimum in (1.2) is attained [Lin83]. Nonuniqueness does not affect the one-atom event: Proposition 2.1 shows that, almost surely, any one-atom maximizer is the unique NPMLE. The minimum-support convention is likewise immaterial in the critical regimes below: our proof shows that the NPMLE is unique with probability tending to one.

For each sample, the one-atom event is monotone in the variance defect. Increasing ε_n damps every likelihood perturbation away from the sample mean, so once a one-atom fit is optimal, it remains optimal under any larger defect. We can therefore study a samplewise critical value of ε_n; Lemmas 2.2 and 2.3 show that this critical value is well defined. The asymptotic law of the critical defect changes sharply between dimensions two and three: in dimension two, observations just inside the radial edge have a nonvanishing collective effect, whereas in dimensions at least three the radial extremes alone govern the transition.

1.1 Main results

We first state the exceptional two-dimensional limit. Put

$$\kappa_n = \frac{\varepsilon_n}{2(1-\varepsilon_n)}, \qquad \rho_n = \sqrt{2\log n}, \qquad \lambda_n = \kappa_n \rho_n^2. \tag{1.3}$$

For $\lambda > 0$, let x_λ be the unique real number satisfying

$$\sup_{r\in\mathbb{R}} \left\{ \Phi(-r) + \exp\left(x_\lambda - \frac{r^2}{2} \right) \right\} = e^{\lambda}, \tag{1.4}$$

where Φ and ϕ are the standard normal distribution function and density, respectively, and set

$$Q_2(\lambda) = \exp\{-e^{-x_\lambda}\}. \tag{1.5}$$

The term $\Phi(-r)$ in (1.4) is the limiting contribution of observations just inside the radial edge, while the exponential term comes from the largest radial observation. The function Q_2 extends continuously and strictly increasingly to $[0,\infty]$, with $Q_2(0) = 0$ and $Q_2(\infty) = 1$; see Lemma 3.1.

For dimensions at least three, an exact radial normalization covers all growth regimes. Let Y_n have the Gamma$(d_n/2, 1)$ distribution, with density f_n and upper tail $\overline{F}_n$. Define

$$n\overline{F}_n(b_n) = 1, \qquad a_n = \frac{\overline{F}_n(b_n)}{f_n(b_n)}, \tag{1.6}$$

and put

$$\gamma_n = (1-\varepsilon_n)^{-1}, \qquad \sigma_n^2 = 1 - \frac{1}{n}, \qquad R_n^\star = \frac{(n-1)^2}{n(n-2)}\log(n-1), \qquad s_n^\star = \frac{\gamma_n R_n^\star/\sigma_n^2 - b_n}{a_n}. \tag{1.7}$$

Thus b_n is the $1-1/n$ quantile of the radial energy $|Z|^2/2$, and a_n is its extreme-value scale. The correction $R_n^\star$ is asymptotically negligible in fixed and moderately growing dimension, but changes the limiting coordinate when $d_n \asymp n^2 \log n$.

We state the transition directly as a limit law for the support size; the one-atom probability is then the probability that $\widehat{K}_n - 1$ equals zero. Here Pois(μ) denotes the Poisson law with mean μ.

Theorem 1.1. *Under either critical-window hypothesis below, the NPMLE is unique with probability tending to one. For the model (1.1), the support-size variable in (1.2) has the following limits.*

1. *If $d = 2$ and $\lambda_n \to \lambda \in (0,\infty)$, then*

$$\widehat{K}_n - 1 \Rightarrow \mathrm{Pois}(e^{-x_\lambda}), \qquad \mathbb{P}\{\widehat{K}_n = 1\} \to Q_2(\lambda). \tag{1.8}$$

Moreover, $\widehat{K}_n = O_{\mathbb{P}}(1)$ whenever $\liminf_n \lambda_n > 0$.

2. *If $d_n \geq 3$ and $s_n^\star \to s \in \mathbb{R}$, then*

$$\widehat{K}_n - 1 \Rightarrow \mathrm{Pois}(e^{-s}), \qquad \mathbb{P}\{\widehat{K}_n = 1\} \to \exp\{-e^{-s}\}. \tag{1.9}$$

Moreover, $\widehat{K}_n = O_{\mathbb{P}}(1)$ whenever $\liminf_n s_n^\star > -\infty$, and $\widehat{K}_n = 1$ with probability tending to one if $s_n^\star \to \infty$.

For fixed $d \geq 3$, the Gamma tail expansion gives

$$b_n = \log n + \left(\frac{d}{2} - 1\right) \log\log n - \log\Gamma(d/2) + o(1), \qquad a_n \to 1.$$

The critical scaling forces $\varepsilon_n \to 0$ and $\varepsilon_n^2 \log n = o(1)$. Thus $s_n^\star \to s$ is equivalent to

$$\varepsilon_n \log n - \left(\frac{d}{2} - 1\right) \log\log n + \log\Gamma(d/2) \longrightarrow s. \tag{1.10}$$

The transition therefore occurs at scale $(\log\log n)/\log n$ in fixed dimensions $d \geq 3$, but at scale $1/\log n$ in dimension two.

The exact normalization also gives explicit scales when d_n grows. For $d_n = o(\log n)$ with $d_n \to \infty$, the critical defect satisfies

$$\varepsilon_{c,n} \sim \frac{d_n}{2\log n} \log \frac{2\log n}{d_n}, \tag{1.11}$$

and the transition window has width asymptotic to $1/\log n$. In larger dimensions, the implicit Gamma quantile (1.6) is simpler and more accurate than separate expansions for different growth rates. At scale $d_n \asymp n^2 \log n$, the $O(n^{-1})$ correction in $R_n^\star$ shifts the Gumbel coordinate by order one, and the exact formula incorporates this shift.

The two fixed-parameter limits in Theorem 1.1 describe one variance defect at a time. In fact, a single radial Poisson cloud governs the entire regularization path. The path theorem below makes precise the sense in which an extreme observation creates a fitted atom and then loses it as the fitted variance is increased. Let $\mathcal{X}$ be a Poisson point process on $\mathbb{R}$ with intensity $e^{-x}\,dx$. Couple each path through one standardized sample $Z_1, \ldots, Z_n$. Write $\widehat{K}_n(z_{1:n})$ for (1.2) computed from the data $z_{1:n}$, and define

$$\varepsilon_{n,2}(\lambda) = \frac{2\lambda}{\rho_n^2 + 2\lambda}, \qquad K_{n,2}(\lambda) = \widehat{K}_n\big(\sqrt{1 - \varepsilon_{n,2}(\lambda)}\, Z_{1:n}\big), \tag{1.12}$$

$$\varepsilon_{n,\geq 3}(s) = 1 - \frac{R_n^\star}{\sigma_n^2(b_n + a_n s)}, \qquad K_{n,\geq 3}(s) = \widehat{K}_n\big(\sqrt{1 - \varepsilon_{n,\geq 3}(s)}\, Z_{1:n}\big). \tag{1.13}$$

The parametrization in (1.13) applies to any sequence $d_n \geq 3$; both defects lie in $[0, 1)$ uniformly on the compact intervals considered below, once n is large.

Theorem 1.2. *For every compact $J \Subset (0, \infty)$, in dimension two,*

$$\{K_{n,2}(\lambda) - 1 : \lambda \in J\} \Longrightarrow \{\mathcal{X}((x_\lambda, \infty)) : \lambda \in J\} \qquad \text{in } D(J). \tag{1.14}$$

For every compact $I \subset \mathbb{R}$ and every sequence $d_n \geq 3$,

$$\{K_{n,\geq 3}(s) - 1 : s \in I\} \Longrightarrow \{\mathcal{X}((s, \infty)) : s \in I\} \qquad \text{in } D(I). \tag{1.15}$$

Here both Skorokhod spaces carry the J_1 topology. In either case, with probability tending to one, the NPMLE is unique simultaneously throughout the specified interval.

The map $\lambda \mapsto x_\lambda$ is a smooth increasing bijection; write Λ for its inverse. The limiting path is a decreasing step function: a Poisson point x disappears at $\lambda = \Lambda(x)$ in dimension two and at $s = x$ in dimensions at least three. Proposition 6.23 gives the corresponding mass profiles and spatial marks.

Samplewise monotonicity turns these probability limits into limit laws for the critical variance defect itself. For a standardized sample $z_{1:n}$, define

$$\varepsilon_*(z_{1:n}) = \inf\{\varepsilon \in [0,1) : \sqrt{1-\varepsilon}\, z_{1:n} \text{ has a one-atom NPMLE}\}.$$

In dimension two, the corresponding variable $\rho_n^2 \varepsilon_*(Z_{1:n})/\{2(1-\varepsilon_*(Z_{1:n}))\}$ has limiting distribution function Q_2. For $d_n \geq 3$, define

$$S_{*,n} = \frac{R_n^\star/[\sigma_n^2\{1-\varepsilon_*(Z_{1:n})\}] - b_n}{a_n}.$$

Then $S_{*,n}$ converges in distribution to a standard Gumbel variable.

1.2 Ideas of the proof

The first-order optimality criterion for a one-atom fit can be written as a random field. After centering the observations and rescaling displacements from the sample mean, this field becomes a sum of unit-width Gaussian bumps. The direction and radius of $Z_i - \bar{Z}_n$ determine the center and height of the ith bump. A sufficiently large observation can therefore push the field above its critical level and force an additional fitted atom.

In dimension two, observations just below the radial maximum have an order-one aggregate contribution. In a local window centered at the maximizing direction, this contribution converges to $\Phi(-r)$ at radial offset r. Angular separation leaves at most one extreme observation in the same window, which yields (1.4). In dimensions at least three, the aggregate inward contribution vanishes at the transition. The radial exceedances converge to a Poisson process with intensity $e^{-x}\,dx$, and each observation above the critical radial-energy threshold contributes one additional fitted atom.

Although the limiting law is universal for $d_n \geq 3$, proving that no other field crossing occurs requires two forms of control. In the lower-dimensional regimes, radial edge points are well separated, and spatial empirical-process estimates control the contribution of the remaining observations. When the dimension is large enough that the centered observations form an approximately regular simplex, an entropy dual replaces Euclidean covering and reduces the local optimization to an ordered mean-field Potts problem. Section 6 then converts the surviving field crossings into fitted atoms. For the path limit, we differentiate the fixed-parameter local likelihood expansions in the variance parameter. After the dimension-dependent normalization of the edge masses, each mass crosses zero with a derivative bounded away from zero. The support identity therefore continues uniformly across the critical window.

1.3 Related literature

The mixture NPMLE goes back to Robbins and Kiefer–Wolfowitz [Rob50, KW56]. Lindsay developed its likelihood geometry and finite-support reduction [Lin83]. In empirical Bayes, the fitted mixing distribution estimates the prior; representative treatments include Jiang–Zhang [JZ09], Koenker–Mizera [KM14], Dicker–Zhao [DZ16], and the multivariate theory of Soloff–Guntuboyina–Sen [SGS25].

Recent work distinguishes statistical accuracy from the combinatorial complexity of an NPMLE. Polyanskiy and Wu proved logarithmic upper bounds on the support size in one-dimensional subgaussian models and asked what happens under a single Gaussian component [PW20]. Our companion paper proves that the support diverges under every fixed finite mixing distribution [Ano26a]. Polyanskiy and Sellke study certified computation and generic support behavior for multivariate Gaussian mixtures [PS25]. A separate companion paper treats one-dimensional variance misspecification, where the critical scale is polynomial and the limit is governed by a two-sided compensated Poisson field [Ano26b].

A separate line of Gaussian-mixture NPMLE theory concerns density estimation or empirical-Bayes risk; see Zhang [Zha09], Saha–Guntuboyina [SG20], Doss et al. [DWYZ23], and more recent stability and adaptivity results [ZV26, GGMT26, CW26]. Smooth mixture NPMLEs provide a closely related form of variance inflation for estimation and inference [KS26]. In the present problem the variance inflation vanishes with n, and we study the support of the likelihood maximizer rather than its estimation risk.

Mixture homogeneity tests compare a point-mass null with an unrestricted mixture alternative; see Gu–Koenker–Volgushev [GKV18] and the references therein. Our criterion asks for more than rejection or acceptance at a likelihood-ratio threshold: it requires the unrestricted optimizer itself to collapse to one atom. Because this event is monotone in the variance inflation, each sample has a critical inflation level, analogous in spirit to the critical bandwidth in modality testing [Sil81, HY01].

2 Likelihood Geometry and Radial Normalization

Both Theorems 1.1 and 1.2 rest on two dimension-free reductions. The KKT criterion identifies one-atomicity with a supremum of a Gaussian-bump field and makes its monotonicity in the variance defect explicit. The Gamma normalization then puts the radial extremes on a common Poisson scale for every sequence $d_n \geq 2$. This section establishes the KKT and radial reductions and the exact centering identities used in all later regimes.

2.1 Notation and one-atom likelihood geometry

The deterministic part of both main theorems is the reduction from mixture likelihood to a Gaussian-bump field. This subsection derives that KKT criterion, proves monotonicity in the variance defect, and records the exact single-contact calculation used later in the simplex regime. We first fix the notation used throughout the paper.

All asymptotic notation is as $n \to \infty$. For a deterministic positive sequence r_n, the relation $U_n = O_{\mathbb{P}}(r_n)$ means that $|U_n|/r_n$ is tight, while $U_n = o_{\mathbb{P}}(r_n)$ means that $U_n/r_n \to 0$ in probability. We write $r_n \lesssim s_n$ when $r_n \leq Cs_n$ for a constant independent of n, and use $\gtrsim, \asymp, \ll$, and $\gg$ analogously. We write $U_n \xrightarrow{\mathbb{P}} U$ for convergence in probability and $U_n \Rightarrow U$ for convergence in distribution. A positive random sequence satisfies $U_n = n^{o_{\mathbb{P}}(1)}$ when, for every $\eta > 0$, $\mathbb{P}\{n^{-\eta} \leq U_n \leq n^{\eta}\} \to 1$. A subscript on asymptotic notation records allowed dependence: for example, $O_{\mathbb{P},R}(1)$ is tight for each fixed R. In iterated limits, $o_A(1)$ denotes a deterministic quantity that tends to zero as $A \to \infty$ after the $n \to \infty$ limit. Constants implicit in $O_m(\cdot)$ may depend on the fixed integer m, and other constants may change from line to line. Point processes carry the vague topology, and a standard Gumbel variable has distribution function $s \mapsto \exp\{-e^{-s}\}$.

Let $d = d_n \geq 2$, let $Z_1, \dots, Z_n$ be independent $\mathcal{N}(0, I_d)$ vectors, and observe

$$X_i = \sqrt{1 - \varepsilon_n}\, Z_i, \qquad 0 \leq \varepsilon_n < 1.$$

We fit location mixtures of $\mathcal{N}(\mu, I_d)$ densities. The only one-atom maximum-likelihood candidate is the point mass at the sample mean. Writing $\bar{Z}_n = n^{-1} \sum_i Z_i$ and $\widetilde{Z}_i = Z_i - \bar{Z}_n$, the directional-derivative criterion says that this candidate is optimal exactly when

$$E_n^{(1)} := \left\{ \sup_{t \in \mathbb{R}^d} D_n(t) \leq 1 \right\}, \qquad D_n(t) = \frac{1}{n} \sum_{i=1}^{n} \exp\left\{ t \cdot \widetilde{Z}_i - \frac{\gamma_n |t|^2}{2} \right\}, \tag{2.1}$$

where

$$\kappa_n = \frac{\varepsilon_n}{2(1-\varepsilon_n)}, \qquad \gamma_n = \frac{1}{1-\varepsilon_n} = 1 + 2\kappa_n. \tag{2.2}$$

Indeed, if a denotes displacement of a fitted component mean from the sample mean, the usual directional derivative is

$$\frac{1}{n}\sum_i \exp\left\{a \cdot (X_i - \bar{X}_n) - \frac{|a|^2}{2}\right\}.$$

The change of variables $t = \sqrt{1-\varepsilon_n}\, a$ gives (2.1). Notice that $D_n(0) = 1$, so the issue is whether the field rises above its value at the origin.

Although a multivariate mixture NPMLE need not be unique, this causes no ambiguity for the event studied here. The following dimension-free fact shows that $E_n^{(1)}$ is almost surely the event that the NPMLE is unique and one-atomic.

Proposition 2.1. *Let $n \geq 2$, and let $X_1, \ldots, X_n$ be i.i.d. from a nondegenerate spherical Gaussian distribution in $\mathbb{R}^d$. The probability that $\delta_{\bar{X}_n}$ and a different mixing measure both maximize the Gaussian-mixture likelihood is zero.*

Proof. If $\delta_{\bar{X}_n}$ and $G \neq \delta_{\bar{X}_n}$ both maximize the likelihood, strict concavity in the vector of fitted values implies $q_G(X_i) = q_{\delta_{\bar{X}_n}}(X_i)$ for every i. Since the one-atom fit is optimal, its physical-coordinate directional field

$$\mathcal{D}_{X,n}(a) = \frac{1}{n}\sum_{i=1}^n \exp\left\{a \cdot (X_i - \bar{X}_n) - \frac{|a|^2}{2}\right\}$$

satisfies $\mathcal{D}_{X,n}(a) \leq 1$. Moreover,

$$\int \mathcal{D}_{X,n}(u - \bar{X}_n)\, dG(u) = \frac{1}{n}\sum_i \frac{q_G(X_i)}{q_{\delta_{\bar{X}_n}}(X_i)} = 1.$$

Thus $\mathcal{D}_{X,n}(a) = 1$ for some $a \neq 0$.

Write $Y = (X_1 - \bar{X}_n, \ldots, X_n - \bar{X}_n) = RU$ in polar form in the subspace $\{(y_1, \ldots, y_n) : \sum_i y_i = 0\}$. Conditional on its direction U, the radius R has a density. After setting $t = Ra$, the directional-derivative field is

$$\mathcal{D}_{R,U}(t) = \frac{1}{n}\sum_i \exp\left\{t \cdot U_i - \frac{|t|^2}{2R^2}\right\}.$$

For fixed U and $t \neq 0$, this is strictly increasing in R. Consequently, for each U, at most one radius can admit both a one-atom maximizer and a nonzero contact: at any larger radius that same contact would make the directional derivative exceed one. Thus, conditional on U, the radius belongs to an at-most-singleton exceptional set, which has probability zero because $R \mid U$ has a density. Integration over U proves the claim. □

The one-atom event is monotone in the variance defect. This elementary fact allows all sharp-window limits below to be converted into laws for a samplewise critical shrinkage.

Lemma 2.2. *Fix $z_1, \ldots, z_n \in \mathbb{R}^d$, and let $x_i^{(\varepsilon)} = \sqrt{1-\varepsilon}\, z_i$ for $0 \leq \varepsilon < 1$. If the sample based on $x_1^{(\varepsilon_1)}, \ldots, x_n^{(\varepsilon_1)}$ admits a one-atom NPMLE and $\varepsilon_2 \geq \varepsilon_1$, then the sample based on $x_1^{(\varepsilon_2)}, \ldots, x_n^{(\varepsilon_2)}$ also admits a one-atom NPMLE.*

Proof. Let $\bar{z} = n^{-1} \sum_i z_i$ and

$$K_n(t) = \frac{1}{n} \sum_{i=1}^{n} \exp\left\{ t \cdot (z_i - \bar{z}) - \frac{|t|^2}{2} \right\}.$$

After the change of variables $t = a\sqrt{1-\varepsilon}$, the one-atom directional field is

$$e^{-\kappa(\varepsilon)|t|^2} K_n(t), \qquad \kappa(\varepsilon) = \frac{\varepsilon}{2(1-\varepsilon)}.$$

Because κ is increasing, the supremum of this field is nonincreasing in ε. The directional-derivative criterion proves the claim. □

The next lemma shows that the monotone one-atom set is nonempty and closed, hence has the form $[\varepsilon_*, 1)$. Its endpoint ε_* is the samplewise critical defect used in the introduction.

Lemma 2.3. *For every fixed sample $z_1, \dots, z_n \in \mathbb{R}^d$, the set of $\varepsilon \in [0, 1)$ for which $\sqrt{1-\varepsilon}\, z_1, \dots, \sqrt{1-\varepsilon}\, z_n$ admits a one-atom NPMLE is a nonempty interval of the form $[\varepsilon_*, 1)$.*

Proof. Let $y_i = z_i - \bar{z}$ and $M = \max_i |y_i|$. The empirical random variable taking the values $t \cdot y_i$ has mean zero and range of length at most $2M|t|$. Hoeffding's lemma gives

$$\log\left(\frac{1}{n} \sum_i e^{t \cdot y_i - |t|^2/2} \right) \le \frac{M^2 - 1}{2} |t|^2.$$

Hence the one-atom field $e^{-\kappa|t|^2} n^{-1} \sum_i e^{t \cdot y_i - |t|^2/2}$ is at most one for every t once $\kappa \ge (M^2 - 1)_+/2$, proving nonemptiness. Lemma 2.2 makes the set an upper interval. Finally, if ε_m belongs to the set and $\varepsilon_m \to \varepsilon < 1$, then the one-atom criterion holds at ε by pointwise convergence of its field for every t. The interval is therefore closed in $[0, 1)$ and contains its infimum. □

For the spatial analysis, it is useful to rewrite D_n as a Gaussian radial-basis field whose centers and heights are determined by the sample. This representation will allow the radial and angular contributions to be studied separately.

Lemma 2.4. *Set*

$$w = \sqrt{\gamma_n}\, t, \qquad q_i = \frac{\widetilde{Z}_i}{\sqrt{\gamma_n}}, \qquad h_i = \frac{1}{n} \exp\left\{ \frac{|q_i|^2}{2} \right\}.$$

Define the bump-coordinate field

$$B_n(w) := D_n(w/\sqrt{\gamma_n}).$$

Then, exactly,

$$B_n(w) = \sum_{i=1}^{n} h_i \exp\left\{ -\frac{|w - q_i|^2}{2} \right\}. \tag{2.3}$$

In particular, the ith bump has maximum h_i, attained at $w = q_i$, and

$$h_i > 1 \iff \frac{|\widetilde{Z}_i|^2}{2} > \gamma_n \log n. \tag{2.4}$$

Proof. Completing the square gives

$$t \cdot \widetilde{Z}_i - \frac{\gamma_n |t|^2}{2} = w \cdot q_i - \frac{|w|^2}{2} = \frac{|q_i|^2}{2} - \frac{|w - q_i|^2}{2}.$$

Summing proves (2.3); maximizing one summand proves (2.4). □

The bump representation isolates the basic mechanism. A radial extreme creates a localized bump whose height is close to one. Other observations contribute a diffuse background, and one-atomicity fails precisely when the full bump field exceeds one. At the transition the centers capable of producing a crossing have radius approximately $\sqrt{2\log n}$ in w-coordinates, even when the original dimension is much larger than $\log n$.

There is also an exact dual formulation which will be decisive in Section 5.1. Let $\Delta_n = \{p \in [0,1]^n : \sum_i p_i = 1\}$, let $\mathbf{u}_n = (1/n, \dots, 1/n)$, and write

$$K(p\|\mathbf{u}_n) = \sum_{i=1}^{n} p_i \log(np_i)$$

for relative entropy, with $0\log 0 = 0$.

Lemma 2.5. *For every centered collection $q_1, \dots, q_n \in \mathbb{R}^d$,*

$$\log \sup_{w\in\mathbb{R}^d} B_n(w) = \sup_{p\in\Delta_n} \left\{ \frac{1}{2} \left| \sum_{i=1}^{n} p_i q_i \right|^2 - K(p\|\mathbf{u}_n) \right\}. \tag{2.5}$$

The expression on the right is always nonnegative, since it vanishes at $p = \mathbf{u}_n$.

Proof. The Gibbs variational formula gives, for every w,

$$\log \left\{ \frac{1}{n} \sum_i e^{w\cdot q_i} \right\} = \sup_{p\in\Delta_n} \left\{ w \cdot \sum_i p_i q_i - K(p\|\mathbf{u}_n) \right\}.$$

Subtract $|w|^2/2$, take the supremum over w, and use $\sup_w\{w\cdot v - |w|^2/2\} = |v|^2/2$. Finally, $\sum_i q_i = 0$, so $p = \mathbf{u}_n$ gives zero. □

The point-mass choice $p = e_i$ in (2.5) gives the one-bump obstruction (2.4). A slightly diffuse competitor gives an exact and strictly stronger finite-sample threshold.

Proposition 2.6. *For $n \geq 3$, define*

$$R_n^\star = \frac{(n-1)^2}{n(n-2)} \log(n-1) = \log n - \frac{1}{n} + O\left(\frac{\log n}{n^2}\right). \tag{2.6}$$

If

$$\max_i \frac{|q_i|^2}{2} > R_n^\star,$$

then $\sup_w B_n(w) > 1$. Consequently, $E_n^{(1)}$ implies $\max_i |q_i|^2/2 \leq R_n^\star$.

Proof. Fix i, and define $p \in \Delta_n$ by

$$p_i = \frac{n-1}{n}, \qquad p_j = \frac{1}{n(n-1)} \quad (j \neq i).$$

Since $\sum_j q_j = 0$,

$$\sum_j p_j q_j = \frac{n-2}{n-1} q_i, \qquad K(p\|\mathbf{u}_n) = \frac{n-2}{n} \log(n-1).$$

The expression in braces in (2.5) is therefore positive exactly when

$$\frac{|q_i|^2}{2} > \frac{(n-1)^2}{n(n-2)} \log(n-1).$$

The expansion in (2.6) follows from Taylor expansion of $\log(n-1)$. □

2.2 Gamma extremes and empirical centering

The normalization in Theorem 1.1 is chosen so that the largest centered radial energies have an order-one Poisson limit. This subsection proves that limit uniformly over arbitrary dimension sequences and quantifies the effect of replacing Z_i by $Z_i - \bar{Z}_n$.

Put

$$L_n = \log n, \qquad k_n = \frac{d_n}{2},$$

and let Y_n have the $\Gamma(k_n, 1)$ distribution. Write f_n and $\overline{F}_n$ for its density and upper tail. Define b_n and a_n by

$$n\overline{F}_n(b_n) = 1, \qquad a_n = \frac{\overline{F}_n(b_n)}{f_n(b_n)}. \tag{2.7}$$

Thus b_n is the exact radial $1 - 1/n$ quantile and a_n is the reciprocal hazard there. Unlike the fixed-dimensional expansion of b_n, these definitions remain valid without specifying a relation between d_n and n.

We first verify that a_n is the correct extreme-value scale for every triangular array of Gamma shapes. The proof is included because uniformity in k_n is essential here.

Lemma 2.7. *For any sequence $k_n \geq 1$, the quantile b_n in (2.7) satisfies*

$$\frac{b_n - k_n}{\sqrt{k_n}} \longrightarrow \infty.$$

Moreover,

$$a_n = \frac{b_n}{b_n - k_n + 1}\{1 + o(1)\}, \tag{2.8}$$

and, uniformly for x in compact subsets of $\mathbb{R}$,

$$\frac{\overline{F}_n(b_n + a_n x)}{\overline{F}_n(b_n)} \longrightarrow e^{-x}. \tag{2.9}$$

Proof. We may argue along subsequences. If k_n remains bounded, then $b_n \to \infty$, and the first assertion is immediate. If $k_n \to \infty$, the central limit theorem gives

$$\mathbb{P}\{Y_n > k_n + M\sqrt{k_n}\} \longrightarrow 1 - \Phi(M)$$

for every fixed M, where Φ is the standard normal distribution function. Since $\overline{F}_n(b_n) = 1/n \to 0$, eventually $b_n > k_n + M\sqrt{k_n}$. Letting $M \to \infty$ proves the first claim.

For $y > k_n - 1$, changing variables $z = y + u$ in the Gamma tail gives the exact identity

$$\frac{\overline{F}_n(y)}{f_n(y)} = \int_0^\infty \exp\left\{-u + (k_n - 1)\log\left(1 + \frac{u}{y}\right)\right\} du. \tag{2.10}$$

Set $\delta_y = 1 - (k_n - 1)/y = (y - k_n + 1)/y$. The inequality $\log(1 + v) \leq v$ shows that the last integral is at most δ_y^{-1}. After the substitution $v = \delta_y u$, its exponent is

$$-v - \frac{k_n - 1}{2y^2\delta_y^2}v^2 + O\left(\frac{k_n}{y^3\delta_y^3}v^3\right)$$

on every bounded v-interval. At $y = b_n$,

$$\frac{k_n}{b_n^2 \delta_{b_n}^2} = \frac{k_n}{(b_n - k_n + 1)^2} \longrightarrow 0.$$

Dominated convergence in (2.10), using the upper envelope e^{-v}, therefore yields $a_n \delta_{b_n} \to 1$, which is (2.8).

If $y = b_n + O(a_n)$, then

$$\frac{\delta_y}{\delta_{b_n}} = 1 + o(1), \qquad \frac{\overline{F}_n(y)}{f_n(y)} = a_n\{1 + o(1)\},$$

uniformly over the indicated range. The first relation follows by differentiating δ_y and using $k_n/(b_n - k_n + 1)^2 \to 0$; the second follows because the dominated-convergence argument applied to (2.10) is uniform for $y = b_n + O(a_n)$. Since $(\log \overline{F}_n)'(y) = -f_n(y)/\overline{F}_n(y)$, integration from b_n to $b_n + a_n x$ gives (2.9). □

The triangular-array maximum therefore has the usual Gumbel law. No dimension-dependent centering expansion is needed.

Proposition 2.8. *Let $Y_{n,1}, \ldots, Y_{n,n}$ be independent $\Gamma(k_n, 1)$ variables. Then, for every $x \in \mathbb{R}$,*

$$\mathbb{P}\left\{ \max_{1 \le i \le n} \frac{Y_{n,i} - b_n}{a_n} \le x \right\} \longrightarrow \exp\{-e^{-x}\}. \tag{2.11}$$

More generally, the number of indices satisfying $Y_{n,i} > b_n + a_n x$ converges to a Poisson variable with mean e^{-x}.

Proof. By Lemma 2.7, $n\overline{F}_n(b_n + a_n x) \to e^{-x}$. The exceedance count is binomial with success probability asymptotic to e^{-x}/n, proving the Poisson limit. Its probability of being zero gives (2.11). □

Throughout the radial analysis, put

$$\sigma_n^2 = 1 - \frac{1}{n}. \tag{2.12}$$

The exact transition coordinate and the associated centered radial excesses are

$$s_n^\star = \frac{\gamma_n R_n^\star/\sigma_n^2 - b_n}{a_n}, \qquad \xi_{n,i}^\star = \frac{|Z_i - \bar{Z}_n|^2/(2\sigma_n^2) - b_n}{a_n} - s_n^\star, \qquad \delta_n^\star = \frac{\sigma_n^2 a_n}{\gamma_n}. \tag{2.13}$$

Thus, for $q_i = (Z_i - \bar{Z}_n)/\sqrt{\gamma_n}$,

$$\frac{|q_i|^2}{2} = R_n^\star + \delta_n^\star \xi_{n,i}^\star. \tag{2.14}$$

These exact quantities will be used in every theorem statement. Some proofs temporarily replace $R_n^\star$ by $\log n$; the comparison is recorded once in Lemma 6.14.

The exact Gamma normalization agrees with the moving-shell normalization in every fixed dimension. The following bridge is used when the bounded- and growing-dimensional arguments are combined.

Lemma 2.9. *Fix $d \ge 3$, put $k = d/2$, and define*

$$b_{n,d} = \log n + (k-1)\log\log n - \log\Gamma(k).$$

Then $a_n \to 1$, $b_n - b_{n,d} \to 0$, and $R_n^\star/\sigma_n^2 = \log n + o(1)$. If $s_n^\star = O(1)$, then

$$s_n^\star = \frac{\gamma_n \log n - b_n}{a_n} + o(1) = \frac{\varepsilon_n}{1 - \varepsilon_n}\log n - (k-1)\log\log n + \log\Gamma(k) + o(1). \tag{2.15}$$

Moreover, uniformly over $1 \le i \le n$,

$$\frac{|Z_i - \bar{Z}_n|^2/(2\sigma_n^2) - b_n}{a_n} = \frac{|Z_i|^2}{2} - b_{n,d} + o_{\mathbb{P}}(1). \tag{2.16}$$

Proof. The fixed-shape Gamma tail expansion gives

$$\overline{F}_n(y) = \frac{y^{k-1}e^{-y}}{\Gamma(k)}\{1 + O(y^{-1})\}, \qquad y \to \infty.$$

Substitution in $n\overline{F}_n(b_n) = 1$ yields $b_n = b_{n,d} + o(1)$, and Lemma 2.7 gives $a_n \to 1$. The expansion of $R_n^\star$ in (2.6) gives $R_n^\star/\sigma_n^2 = \log n + o(1)$. Since $s_n^\star = O(1)$ forces $\gamma_n = O(1)$, the first equality in (2.15) follows. The second uses $\gamma_n - 1 = \varepsilon_n/(1 - \varepsilon_n)$.

Finally, $\max_i |Z_i| = O_{\mathbb{P}}(\sqrt{\log n})$ and $|\bar{Z}_n| = O_{\mathbb{P}}(n^{-1/2})$. Hence

$$\max_i \left| |Z_i - \bar{Z}_n|^2 - |Z_i|^2 \right| = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{n}}\right) = o_{\mathbb{P}}(1),$$

while multiplication by $\sigma_n^{-2} = 1 + O(n^{-1})$ changes the maximum energy by $O_{\mathbb{P}}(\log n/n) = o_{\mathbb{P}}(1)$. Together with $a_n \to 1$ and $b_n - b_{n,d} \to 0$, this proves (2.16). ◻

The Karush–Kuhn–Tucker (KKT) field uses centered observations. The next condition is sufficient for empirical centering to be negligible on the Gamma extreme scale:

$$\frac{\sqrt{b_n d_n/n}}{a_n} \longrightarrow 0. \tag{2.17}$$

It is automatic throughout $d_n = o(\log n)$. In high dimension the direct norm bound behind this condition is wasteful; the leave-one-out estimate below reaches $d_n \log n = o(n^2)$.

Lemma 2.10. *Assume (2.17). Then*

$$\max_i \left| \frac{|Z_i - \bar{Z}_n|^2 - |Z_i|^2}{2a_n} \right| \longrightarrow_{\mathbb{P}} 0. \tag{2.18}$$

In particular,

$$\max_i \frac{|Z_i - \bar{Z}_n|^2/2 - b_n}{a_n} - \max_i \frac{|Z_i|^2/2 - b_n}{a_n} \longrightarrow_{\mathbb{P}} 0. \tag{2.19}$$

Consequently, the centered maximum has the Gumbel limit (2.11).

Proof. Proposition 2.8 and tightness imply $\max_i |Z_i|^2/2 = b_n + O_{\mathbb{P}}(a_n)$. Lemma 2.7 also implies $a_n = O(b_n)$, so $\max_i |Z_i| = O_{\mathbb{P}}(\sqrt{b_n})$. Since $|\bar{Z}_n| = O_{\mathbb{P}}(\sqrt{d_n/n})$,

$$\max_i \left| \frac{|Z_i - \bar{Z}_n|^2 - |Z_i|^2}{2a_n} \right| \le \frac{\max_i |Z_i|\,|\bar{Z}_n| + |\bar{Z}_n|^2/2}{a_n} = o_{\mathbb{P}}(1)$$

by (2.17). This proves (2.18); taking maxima gives (2.19). ◻

The norm bound in Lemma 2.10 controls $Z_i \cdot \bar{Z}_n$ by the product of two norms and is unnecessarily wasteful when d_n is large. A leave-one-out decomposition retains the cancellation in this inner product.

Lemma 2.11. *Uniformly over $d_n \geq 1$,*

$$\max_{1\leq i\leq n}\left|\frac{|Z_i-\bar{Z}_n|^2-|Z_i|^2}{2}\right| = O_{\mathbb{P}}\left\{\frac{d_n}{n}+\sqrt{\frac{d_nL_n}{n}}+\frac{L_n}{\sqrt{n}}\right\}. \tag{2.20}$$

Consequently, if $L_n = o(d_n)$ and $d_nL_n = o(n^2)$, then (2.19) holds.

Proof. Write

$$Z_i\cdot\bar{Z}_n = \frac{|Z_i|^2}{n}+Z_i\cdot\bar{Z}_{-i}, \qquad \bar{Z}_{-i} = \frac{1}{n}\sum_{j\neq i}Z_j.$$

The two vectors in the second inner product are independent, and $\bar{Z}_{-i}\sim\mathcal{N}(0,(n-1)n^{-2}I_{d_n})$. Its moment generating function therefore gives the subexponential bound

$$\max_i|Z_i\cdot\bar{Z}_{-i}| = O_{\mathbb{P}}\left\{\sqrt{\frac{d_nL_n}{n}}+\frac{L_n}{\sqrt{n}}\right\}.$$

The standard chi-square tail bound and a union bound also give

$$\max_i|Z_i|^2 = O_{\mathbb{P}}\{d_n+\sqrt{d_nL_n}+L_n\}, \qquad |\bar{Z}_n|^2 = O_{\mathbb{P}}\{(d_n+1)/n\}.$$

Substituting these estimates into

$$\frac{|Z_i-\bar{Z}_n|^2-|Z_i|^2}{2} = -Z_i\cdot\bar{Z}_n+\frac{|\bar{Z}_n|^2}{2}$$

proves (2.20).

In the stated range, $a_n\sim\sqrt{d_n/(4L_n)}$. Dividing the three terms in (2.20) by a_n gives, respectively,

$$O\left(\frac{\sqrt{d_nL_n}}{n}\right), \qquad O\left(\frac{L_n}{\sqrt{n}}\right), \qquad O\left(\frac{L_n^{3/2}}{\sqrt{d_n}n}\right),$$

all of which tend to zero. This proves (2.19). □

The two norm-based centering lemmas do not use the exact marginal distribution of a centered residual. In fact,

$$\frac{Z_i-\bar{Z}_n}{\sigma_n}\sim\mathcal{N}(0,I_{d_n})$$

for every i. The residuals are dependent, but their correlations are weak enough that the extreme exceedance counts still have Poisson limits. The next proposition uses this exact marginal law to reach dimensions beyond the range of Lemma 2.11.

Proposition 2.12. *Suppose*

$$L_n = o(d_n).$$

For every fixed $x\in\mathbb{R}$,

$$\mathbb{P}\left\{\max_i\frac{|Z_i-\bar{Z}_n|^2/(2\sigma_n^2)-b_n}{a_n}\leq x\right\}\longrightarrow e^{-e^{-x}}. \tag{2.21}$$

More generally, the number of indices for which

$$\frac{|Z_i-\bar{Z}_n|^2}{2\sigma_n^2} > b_n+a_nx$$

converges to $\mathrm{Pois}(e^{-x})$.

Proof. Fix $m \geq 1$. We prove, uniformly for x in compact subsets of $\mathbb{R}$, that

$$\mathbb{P}\{Y_1 > y_n, \ldots, Y_m > y_n\} = \{1 + o(1)\}\mathbb{P}\{Y_1 > y_n\}^m, \qquad y_n = b_n + a_n x, \tag{2.22}$$

where

$$V_i = \frac{Z_i - \bar{Z}_n}{\sigma_n}, \qquad Y_i = \frac{|V_i|^2}{2}.$$

The proposition will then follow by the method of factorial moments.

We first record the Gamma estimates needed below. Put $k = k_n = d_n/2$, $L = L_n = \log n$, and

$$h(\delta) = \delta - \log(1 + \delta), \qquad \delta > -1.$$

If $Y \sim \Gamma(k, 1)$ and $y = k(1 + \delta)$, where $\delta \to 0$ and $k\delta^2 \to \infty$, then

$$\mathbb{P}\{Y > y\} = \frac{1 + o(1)}{\delta\sqrt{2\pi k}} \exp\{-kh(\delta)\}. \tag{2.23}$$

Writing f_k for the density of Y, Stirling's formula gives

$$f_k\{k(1 + \delta)\} = \frac{1 + o(1)}{(1 + \delta)\sqrt{2\pi k}} \exp\{-kh(\delta)\}.$$

Moreover, the exact identity

$$\frac{\mathbb{P}\{Y > y\}}{f_k(y)} = \int_0^\infty \exp\left\{-v + (k - 1)\log\left(1 + \frac{v}{y}\right)\right\} dv$$

shows that

$$\frac{\mathbb{P}\{Y > y\}}{f_k(y)} = \frac{y}{y - k + 1}\{1 + o(1)\} = \frac{1 + \delta}{\delta}\{1 + o(1)\}. \tag{2.24}$$

For completeness, set $\eta = (y - k + 1)/y$ and substitute $s = \eta v$ in the integral. Its exponent becomes

$$-s - \frac{k - 1}{2y^2\eta^2}s^2 + O\left(\frac{k}{y^3\eta^3}s^3\right)$$

on bounded s-intervals, while $\log(1 + v/y) \leq v/y$ supplies the dominating function e^{-s}. Since $k/(y^2\eta^2) \sim (k\delta^2)^{-1} \to 0$, dominated convergence proves (2.24), and hence (2.23).

Let

$$\delta_n = \frac{b_n - k}{k}.$$

The central limit theorem and $\mathbb{P}\{Y > b_n\} = n^{-1}$ imply $\delta_n\sqrt{k} \to \infty$. The Chernoff bound

$$\mathbb{P}\{Y > k(1 + \delta)\} \leq e^{-kh(\delta)}$$

gives $kh(\delta_n) \leq L = o(k)$, and therefore $\delta_n \to 0$. Applying (2.23) at b_n yields

$$L = kh(\delta_n) + \log(\delta_n\sqrt{2\pi k}) + o(1).$$

Because $h(\delta) \sim \delta^2/2$, this implies

$$\delta_n \sim \sqrt{\frac{2L}{k}}, \qquad b_n - k \sim \sqrt{2kL}, \qquad a_n \sim \frac{1}{\delta_n} \sim \sqrt{\frac{k}{2L}}. \tag{2.25}$$

If $y_n = b_n + a_n x = k(1+\delta_{n,x})$, then, uniformly for bounded x,

$$\frac{\delta_{n,x}}{\delta_n} = 1 + o(1).$$

Using $h'(\delta) = \delta/(1+\delta)$, (2.23), and $a_n\delta_n/(1+\delta_n) \to 1$, we obtain

$$k\{h(\delta_{n,x}) - h(\delta_n)\} = x + O\left(\frac{a_n^2}{k}\right) + o(1) = x + o(1).$$

Consequently,

$$p_n(x) := \mathbb{P}\{Y_i > y_n\} = \frac{e^{-x} + o(1)}{n}. \tag{2.26}$$

We next compare the joint law of $Y_1, \ldots, Y_m$ with m independent Gamma variables. For each Euclidean coordinate, the Gaussian vector $(V_1, \ldots, V_m)$ has covariance

$$A = A_{n,m} = \frac{n}{n-1} I_m - \frac{1}{n-1}\mathbf{1}\mathbf{1}^\top = I_m + C,$$

where

$$C_{ii} = 0, \qquad C_{ij} = -\frac{1}{n-1} \quad (i \neq j), \qquad \|C\|_{\mathrm{op}} + \|C\|_{\mathrm{F}} = O_m(n^{-1}).$$

Set

$$t = t_n = 1 - \frac{k}{y_n}, \qquad \beta = \beta_n = 1 - t = \frac{k}{y_n}.$$

By (2.25),

$$t \to 0, \qquad t\sqrt{d_n} = (2 + o(1))\sqrt{L_n}, \qquad d_n t^2 = (4 + o(1))L_n. \tag{2.27}$$

Let $M^{\mathrm{dep}}_{n,m}(t)$ be the joint moment-generating function of $(Y_1, \ldots, Y_m)$ at the common argument t. The Gaussian quadratic-form formula gives

$$M^{\mathrm{dep}}_{n,m}(t) = \det(I_m - tA)^{-d_n/2}.$$

For m independent copies,

$$M^{\mathrm{ind}}_{n,m}(t) = \beta^{-md_n/2}.$$

Since $\operatorname{tr} C = 0$,

$$\begin{aligned} \log\frac{M^{\mathrm{dep}}_{n,m}(t)}{M^{\mathrm{ind}}_{n,m}(t)} &= -\frac{d_n}{2}\log\det\left(I_m - \frac{t}{\beta}C\right) \\ &= O_m\left(\frac{d_n t^2}{n^2}\right) = O_m\left(\frac{L_n}{n^2}\right) = o(1). \end{aligned} \tag{2.28}$$

Under the dependent exponential tilt, each coordinate vector is Gaussian with covariance

$$\Sigma^{\mathrm{dep}} = (A^{-1} - tI_m)^{-1} = A(I_m - tA)^{-1}.$$

Under the independent tilt the corresponding covariance is

$$\Sigma^{\mathrm{ind}} = \beta^{-1} I_m.$$

We write $\mathbb{E}_t^{\mathrm{dep}}$ and $\mathbb{E}_t^{\mathrm{ind}}$ for expectation under these two tilted laws, and use $\mathbb{E}_t^\star$ when the same formula applies to either one. Writing $f_t(a) = a/(1-ta)$ and expanding $f_t(I_m + C)$ at I_m gives

$$\Sigma^{\mathrm{dep}} = \beta^{-1} I_m + \beta^{-2} C + R_{n,m}, \qquad \|R_{n,m}\|_{\mathrm{op}} \le C_m t \|C\|_{\mathrm{op}}^2.$$

In particular,

$$(\Sigma^{\mathrm{dep}})_{ii} = \beta^{-1} + O_m(tn^{-2}), \qquad (\Sigma^{\mathrm{dep}})_{ij} = -\frac{\beta^{-2}}{n-1} + O_m(tn^{-2}) \quad (i \neq j). \tag{2.29}$$

The eigenvalues of both tilted covariance matrices therefore lie in a fixed compact subset of $(0, \infty)$.

We now prove the local limit estimate needed for the tilted overshoots. Let $f_{n,m}^{\mathrm{dep}}$ and $f_{n,m}^{\mathrm{ind}}$ denote the respective densities of

$$W_n = \frac{1}{\sqrt{d_n}} (Y_1 - y_n, \ldots, Y_m - y_n)$$

under the dependent and independent tilted laws. We claim that

$$\sup_{w \in \mathbb{R}^m} \left| f_{n,m}^\star(w) - \varphi_m(w) \right| \longrightarrow 0, \qquad \varphi_m(w) = \pi^{-m/2} e^{-|w|^2}, \tag{2.30}$$

for $\star \in \{\mathrm{dep}, \mathrm{ind}\}$.

To prove the claim, let $\Sigma^\star$ be the appropriate tilted covariance, let $D_z = \operatorname{diag}(z_1, \ldots, z_m)$, and define

$$B_z^\star = (\Sigma^\star)^{1/2} D_z (\Sigma^\star)^{1/2}.$$

The characteristic function of W_n is

$$\chi_{n,m}^\star(z) = \exp\left\{ -\frac{i y_n}{\sqrt{d_n}} \sum_{j=1}^m z_j \right\} \det\left(I_m - \frac{i}{\sqrt{d_n}} \Sigma^\star D_z \right)^{-d_n/2}. \tag{2.31}$$

For fixed z, expansion of the logarithm gives

$$\begin{aligned} \log \chi_{n,m}^\star(z) = {} & \frac{i\sqrt{d_n}}{2} \sum_{j=1}^m \left\{ (\Sigma^\star)_{jj} - \beta^{-1} \right\} z_j \\ & - \frac{1}{4} \operatorname{tr}(\Sigma^\star D_z \Sigma^\star D_z) + O_m\left(\frac{|z|^3}{\sqrt{d_n}} \right). \end{aligned}$$

The linear term vanishes identically in the independent case and is $O_m(t\sqrt{d_n}|z|/n^2) = o(1)$ in the dependent case by (2.27) and (2.29). Since $\Sigma^\star \to I_m$, it follows that

$$\chi_{n,m}^\star(z) \longrightarrow e^{-|z|^2/4}$$

for every fixed z.

We next give an integrable domination uniform in the two tilted laws. Since $B_z^\star$ is real symmetric,

$$|\chi_{n,m}^\star(z)| = \det\left(I_m + \frac{(B_z^\star)^2}{d_n} \right)^{-d_n/4}. \tag{2.32}$$

If $\mu_1, \dots, \mu_m$ are the eigenvalues of $B_z^\star$, the uniform lower eigenvalue bound for $\Sigma^\star$ gives

$$\sum_{j=1}^m \mu_j^2 = \|B_z^\star\|_{\mathrm{F}}^2 \ge c_m |z|^2.$$

Hence

$$|\chi_{n,m}^\star(z)| \le \left(1 + \frac{c_m |z|^2}{d_n}\right)^{-d_n/4}. \tag{2.33}$$

Choose $c_0 > 0$ sufficiently small. If $|z| \le c_0\sqrt{d_n}$, then all $\mu_j^2/d_n \le 1$; using $\log(1+v) \ge v/2$ in (2.32) yields

$$|\chi_{n,m}^\star(z)| \le e^{-c_m'|z|^2}. \tag{2.34}$$

On the complementary region, polar coordinates and (2.33) give

$$\int_{|z|>c_0\sqrt{d_n}} \left(1 + \frac{c_m|z|^2}{d_n}\right)^{-d_n/4} dz \\ \le C_m d_n^{m/2} (1 + c_m'')^{-d_n/8} \mathrm{B}\left(\frac{m}{2}, \frac{d_n}{8} - \frac{m}{2}\right) = o(1).$$

Here B denotes the Euler beta function. Thus the characteristic functions are uniformly integrable. Pointwise convergence, (2.34), and the last tail estimate imply

$$\int_{\mathbb{R}^m} \left|\chi_{n,m}^\star(z) - e^{-|z|^2/4}\right| dz \longrightarrow 0.$$

Fourier inversion proves (2.30). Notice that this argument treats the dependent and independent tilted densities separately; both converge uniformly to the same density φ_m, and in particular

$$f_{n,m}^{\mathrm{dep}}(0),\ f_{n,m}^{\mathrm{ind}}(0) \longrightarrow \varphi_m(0) = \pi^{-m/2}.$$

It remains to compare the tilted overshoot integrals. For $\star \in \{\mathrm{dep}, \mathrm{ind}\}$, put

$$J_{n,m}^\star = \mathbb{E}_t^\star \left[\exp\left\{-t\sum_{j=1}^m (Y_j - y_n)\right\} \mathbf{1}_{\{Y_j > y_n,\ 1\le j\le m\}}\right].$$

Writing $q_n = t\sqrt{d_n}$, the density of W_n gives

$$q_n^m J_{n,m}^\star = \int_{\mathbb{R}_+^m} e^{-\sum_j z_j} f_{n,m}^\star(z/q_n)\, dz.$$

By (2.27), $q_n \to \infty$. The uniform local limit theorem supplies a common bound on the densities, so dominated convergence yields

$$q_n^m J_{n,m}^\star \longrightarrow \varphi_m(0).$$

Therefore

$$\frac{J_{n,m}^{\mathrm{dep}}}{J_{n,m}^{\mathrm{ind}}} \longrightarrow 1.$$

The exponential-tilting identities are

$$\mathbb{P}\{Y_1 > y_n, \dots, Y_m > y_n\} = M_{n,m}^{\mathrm{dep}}(t) e^{-tmy_n} J_{n,m}^{\mathrm{dep}}$$

and

$$p_n(x)^m = M_{n,m}^{\mathrm{ind}}(t) e^{-tmy_n} J_{n,m}^{\mathrm{ind}}.$$

Together with (2.28), they prove (2.22).

Finally, let

$$N_n(x) = \sum_{i=1}^{n} \mathbf{1}_{\{Y_i > b_n + a_n x\}}.$$

For an integer-valued variable N, write $\{N\}_m = N(N-1)\cdots(N-m+1)$, and similarly $(n)_m = n(n-1)\cdots(n-m+1)$. For every fixed m, exchangeability, (2.22), and (2.26) give

$$\mathbb{E}\{N_n(x)\}_m = (n)_m \mathbb{P}\{Y_1 > y_n, \dots, Y_m > y_n\} \longrightarrow e^{-mx}$$

for every fixed m. These are the factorial moments of $\mathrm{Pois}(e^{-x})$, so $N_n(x) \Rightarrow \mathrm{Pois}(e^{-x})$. Taking the probability that $N_n(x) = 0$ proves (2.21). □

3 The Two-Dimensional Sharp Window

This section establishes the field estimates for the two-dimensional clauses of Theorems 1.1 and 1.2. At the critical scale, the KKT field near its maximizing direction contains both a single extreme radial spike and a nonvanishing contribution from observations just inside the edge. We identify their limiting profile, confine every possible crossing to a compact radial window, and prove a uniform decomposition used for support recovery in Section 6.

Throughout this section let $Z_1, \dots, Z_n$ be i.i.d. $\mathcal{N}(0, I_2)$, let $\rho_n = \sqrt{2\log n}$, set $\bar{Z}_n = n^{-1}\sum_i Z_i$, and put

$$H_n(t) = \frac{1}{n}\sum_{i=1}^{n} \exp\{t\cdot Z_i - |t|^2/2\}, \qquad D_n(t) = e^{-t\cdot\bar{Z}_n} e^{-\kappa_n |t|^2} H_n(t), \qquad \lambda_n = \kappa_n \rho_n^2.$$

The one-atom event is $E_n^{(1)} = \{\sup_t D_n(t) \le 1\}$. We write Φ and ϕ for the standard normal distribution function and density.

3.1 The limiting profile

The KKT supremum chooses its direction after seeing the sample. A fixed angular interval of width $1/\rho_n$ contains a point with bounded radial excess with probability $O(1/\rho_n)$, but the maximizing interval can be centered on such a point after the sample is observed. This selected extreme creates one random spike. The many observations whose radial energies lie just below $\log n$ produce the deterministic background.

Write

$$Z_i = R_i U_i, \qquad U_i = e^{i\Theta_i} \in S^1, \qquad \xi_i = \frac{R_i^2}{2} - \log n,$$

and let $\mathrm{ang}(\phi - \psi) \in (-\pi, \pi]$ denote the principal signed angular displacement from direction ψ to direction ϕ. We also write $\mathrm{ang}(u, v)$ for the same displacement from v to u, and set $Y_i(v) = \rho_n \,\mathrm{ang}(U_i, v)$. The global radial edge process

$$\Xi_n = \sum_{i=1}^{n} \delta_{(\xi_i, \Theta_i)}$$

converges to a Poisson point process $\Xi = \sum_j \delta_{(\xi_j,\Theta_j)}$ on $\mathbb{R}\times\mathbb{T}$, where $\mathbb{T}=\mathbb{R}/2\pi\mathbb{Z}$, with intensity

$$e^{-x}\,dx\,\frac{d\phi}{2\pi}.$$

For a fixed deterministic direction ψ, however, the locally rescaled process

$$\sum_i \delta_{(\xi_i,\,\rho_n\,\mathrm{ang}(\Theta_i-\psi))}$$

has intensity

$$\frac{1+o(1)}{2\pi\rho_n}\,e^{-x}\,dx\,dy$$

on compact x,y-sets, and hence converges to the empty process. The nontrivial local limit is obtained only after conditioning on, and recentering around, a radial edge point.

For $t=(\rho_n+s)e^{i\psi}$, one observation contributes

$$W_{n,i}(\psi,s)=\frac{1}{n}\exp\left\{(\rho_n+s)e^{i\psi}\cdot Z_i-\frac{(\rho_n+s)^2}{2}\right\}.$$

If ξ_i, s, and

$$y_i=\rho_n\,\mathrm{ang}(\Theta_i-\psi)$$

are bounded, then

$$W_{n,i}(\psi,s)=\exp\{\xi_i-s^2/2-y_i^2/2+o(1)\}.$$

Thus a radial edge point (ξ_j,Θ_j), viewed in the angular coordinate $\psi=\Theta_j+y/\rho_n$, creates the local spike

$$e^{\xi_j-s^2/2-y^2/2}.$$

The spike has $O(1)$ width in blown-up coordinates, hence $O(1/\rho_n)$ width on the original circle.

For fixed $A,C<\infty$, the $O(1)$-radial edge spikes are separated:

$$\mathbb{P}\left\{\exists i\neq j:\ \xi_i,\xi_j\geq -A,\ \rho_n|\,\mathrm{ang}(\Theta_i-\Theta_j)|\leq C\right\}\to 0,$$

because the expected number of such pairs is

$$O\left\{\left(\int_{-A}^{\infty}e^{-x}\,dx\right)^2\frac{C}{\rho_n}\right\}=o(1).$$

Consequently, after the uniform compactness estimates are proved, at most one radial extreme contributes at order one near any maximizing direction. All other observations form the deterministic background. The limiting event is therefore determined by the largest radial excess rather than by a nontrivial random process on the entire circle.

The following calculation identifies the background. Let $t=(\rho_n+s)v$, $v\in S^1$, and truncate observations at the one-dimensional projection edge in direction v:

$$H_n^{\mathrm{diff}}(v,s)=\frac{1}{n}\sum_{i=1}^{n}\exp\left\{(\rho_n+s)v\cdot Z_i-\frac{(\rho_n+s)^2}{2}\right\}\mathbf{1}_{\{v\cdot Z_i\leq\rho_n\}}$$

satisfies

$$\mathbb{E}H_n^{\mathrm{diff}}(v,s)=\Phi(-s),\qquad \operatorname{Var}H_n^{\mathrm{diff}}(v,s)\le \frac{1}{n}e^{(\rho_n+s)^2}\Phi(-\rho_n-2s)\sim \frac{e^{-s^2}}{\sqrt{2\pi}\rho_n}$$

uniformly for s in compact sets. Thus the non-extreme cloud contributes $\Phi(-s)$ to H_n, or $e^{-\lambda}\Phi(-s)$ after the damping factor in D_n.

A radial edge point with

$$\xi_i=\frac{|Z_i|^2}{2}-\log n$$

and direction U_i contributes, at $t=(\rho_n+s)U_i$,

$$e^{-\lambda}\exp\{\xi_i-s^2/2+o(1)\}.$$

Combining the two contributions, a radial excess x can produce a crossing precisely when

$$g(x):=\sup_{s\in\mathbb{R}}\{\Phi(-s)+e^{x-s^2/2}\}$$

exceeds the damped threshold e^{λ}. This gives the one-atom limit

$$Q_2(\lambda)=\exp\{-e^{-x_\lambda}\},$$

where x_λ is uniquely determined by

$$\sup_{s\in\mathbb{R}}\{\Phi(-s)+e^{x_\lambda-s^2/2}\}=e^{\lambda}.$$

Equivalently, if $a_\lambda>0$ solves

$$\Phi(a_\lambda)+\frac{\phi(a_\lambda)}{a_\lambda}=e^{\lambda},$$

then

$$x_\lambda=-\log(\sqrt{2\pi}a_\lambda),\qquad Q_2(\lambda)=\exp\{-\sqrt{2\pi}a_\lambda\}.$$

The optimizer is $s=-a_\lambda$, obtained from

$$-\phi(s)-se^{x-s^2/2}=0.$$

The map $a\mapsto\Phi(a)+\phi(a)/a$ is strictly decreasing, since its derivative is $-\phi(a)/a^2$, so a_λ is unique.

The optimized profile crosses every level above one exactly once. Its strict monotonicity later converts a fixed radial gap into a fixed KKT gap.

Lemma 3.1. *Let*

$$g(x)=\sup_{s\in\mathbb{R}}\{\Phi(-s)+e^{x-s^2/2}\}.$$

Then g is finite, continuous, and strictly increasing, with $\lim_{x\to-\infty}g(x)=1$ *and* $\lim_{x\to\infty}g(x)=\infty$. *Hence, for every $\lambda>0$, there is a unique x_λ such that $g(x_\lambda)=e^{\lambda}$, and for every $\eta>0$,*

$$g(x_\lambda-\eta)<e^{\lambda}<g(x_\lambda+\eta).$$

Proof. For every finite x, the objective exceeds 1 at some finite negative s. Indeed, with $s = -a$, Mills' ratio gives $1 - \Phi(a) \sim \phi(a)/a$, whereas $e^{x-a^2/2} = \sqrt{2\pi}e^x\phi(a)$; the second term dominates the first for sufficiently large a. On a compact x-interval, the objective converges uniformly to 1 as $s \to -\infty$ and to 0 as $s \to +\infty$. Its supremum is therefore attained in a common compact s-interval. Continuity follows from compact maximization, and strict monotonicity follows because $e^{x-s^2/2}$ is strictly increasing in x at every finite maximizer. Finally, $1 \le g(x) \le 1 + e^x$ implies $g(x) \to 1$ as $x \to -\infty$, while $g(x) \ge e^x \to \infty$ as $x \to \infty$. $\square$

The explicit parametrization also gives the endpoint behavior

$$1 - Q_2(\lambda) \sim e^{-\lambda} \quad (\lambda \to \infty), \qquad \log Q_2(\lambda) = -2\sqrt{\pi \log(1/\lambda)}\{1 + o(1)\} \quad (\lambda \downarrow 0).$$

Indeed, as $\lambda \to \infty$, the defining equation gives $a_\lambda \sim \phi(0)e^{-\lambda}$. Since $e^{-x_\lambda} = \sqrt{2\pi}a_\lambda$, this proves the first relation. As $\lambda \downarrow 0$, the two-term Mills expansion gives

$$e^\lambda - 1 = \frac{\phi(a_\lambda)}{a_\lambda^3}\{1 + O(a_\lambda^{-2})\}.$$

Thus $a_\lambda^2 \sim 2\log(1/\lambda)$, which proves the second relation. For large λ, the diffuse background is negligible and the radial Gumbel tail determines the limit. By contrast, $Q_2(\lambda) \downarrow 0$ as the shrinkage tends to zero.

3.2 Radial compactness

Radial compactness reduces all possible crossings to $|t| = \rho_n + O(1)$, without yet identifying the limiting field.

Proposition 3.2. *Assume $d = 2$ and $\lambda_n = \kappa_n\rho_n^2 \to \lambda \in (0, \infty)$, where $\rho_n = \sqrt{2\log n}$. Then*

$$\lim_{A\to\infty} \limsup_{n\to\infty} \mathbb{P}\left\{ \sup_{||t|-\rho_n|>A} D_n(t) > 1 \right\} = 0.$$

Thus, up to a probability which vanishes after $n \to \infty$ and then $A \to \infty$, every crossing has the form

$$t = (\rho_n + s)v, \qquad |s| \le A, \quad v \in S^1.$$

Write

$$H_n^{\mathrm{diff}}(r, v) = \frac{1}{n}\sum_i e^{rv\cdot Z_i - r^2/2}\mathbf{1}_{\{v\cdot Z_i \le \rho_n\}}, \qquad T_n(r, v) = \frac{1}{n}\sum_i e^{rv\cdot Z_i - r^2/2}\mathbf{1}_{\{v\cdot Z_i > \rho_n\}}.$$

Thus $H_n(rv) = H_n^{\mathrm{diff}}(r, v) + T_n(r, v)$. The split is at the one-dimensional edge in direction v.

The diffuse part cannot create a crossing below the compact window or at its upper boundary.

Lemma 3.3. *For every $\delta > 0$,*

$$\lim_{A\to\infty} \limsup_{n\to\infty} \mathbb{P}\left\{ \sup_{\rho_n/2\le r\le\rho_n - A,\ v\in S^1} H_n^{\mathrm{diff}}(r, v) > 1 + \delta \right\} = 0,$$

$$\lim_{A\to\infty} \limsup_{n\to\infty} \mathbb{P}\left\{ \sup_{v\in S^1} H_n^{\mathrm{diff}}(\rho_n + A, v) > \delta \right\} = 0.$$

Proof. Put

$$\xi_i = \frac{|Z_i|^2}{2} - \log n, \qquad \ell_n^{\mathrm{far}} = 8\log\rho_n,$$

and let

$$\mathcal{R}_n(A) = \{r : \rho_n/2 \le r \le \rho_n - A\} \cup \{\rho_n + A\}.$$

Completing the square gives the radial identity

$$\frac{1}{n}\exp\{rv\cdot z - r^2/2\} = \exp\left\{x(z) - \frac{|z-rv|^2}{2}\right\}, \qquad x(z) = \frac{|z|^2}{2} - \log n. \tag{3.1}$$

First remove the near-edge layer. Let $H_n^{\mathrm{diff},0}$ be H_n^{diff} with the additional restriction $\xi_i \le -\ell_n^{\mathrm{far}}$. Every summand retained in $H_n^{\mathrm{diff},0}$ is at most $e^{-\ell_n^{\mathrm{far}}}$ after normalization. Hence the unnormalized summand is bounded by $ne^{-\ell_n^{\mathrm{far}}}$, and its second moment is at most $ne^{-\ell_n^{\mathrm{far}}}$ times its first moment. Since

$$\mathbb{E}H_n^{\mathrm{diff},0}(r,v) \le \mathbb{E}H_n^{\mathrm{diff}}(r,v) = \Phi(\rho_n - r) \le 1 \qquad (r\in\mathcal{R}_n(A)),$$

Bernstein's inequality gives, at each fixed (r,v),

$$\mathbb{P}\{|H_n^{\mathrm{diff},0}(r,v) - \mathbb{E}H_n^{\mathrm{diff},0}(r,v)| > \eta\} \le 2\exp\{-c_\eta e^{\ell_n^{\mathrm{far}}}\}.$$

Let $\mathcal{F}_n$ be the class of normalized summands

$$z \mapsto \frac{1}{n}e^{rv\cdot z - r^2/2}\mathbf{1}_{\{v\cdot z\le\rho_n\}}\mathbf{1}_{\{x(z)\le -\ell_n^{\mathrm{far}}\}}, \qquad r\in\mathcal{R}_n(A),\ v\in S^1,$$

so that $H_n^{\mathrm{diff},0}(r,v)$ is the sum of n functions from $\mathcal{F}_n$. This is a VC-subgraph class of fixed dimension: after taking logarithms in the subgraph coordinate, its subgraphs are intersections of a fixed ball, a halfspace in z, and a halfspace in $(z,\log y)$. Standard closure properties therefore give polynomial covering numbers, uniformly in n; see [vdVW96, Sections 2.6.2 and 2.14.1]. In the present bounded-envelope setting, the Bernstein maximal inequality from that entropy bound has the form

$$\mathbb{P}\left\{\sup_{f\in\mathcal{F}_n}\left|\sum_{i=1}^n\{f(Z_i) - \mathbb{E}f(Z_i)\}\right| > \eta\right\} \le C\rho_n^C\exp\left\{-\frac{c\eta^2}{e^{-\ell_n^{\mathrm{far}}} + \eta e^{-\ell_n^{\mathrm{far}}}}\right\}.$$

Moreover, identity (3.1) bounds every member of $\mathcal{F}_n$ by $e^{-\ell_n^{\mathrm{far}}}$, and the sum of its variances by $e^{-\ell_n^{\mathrm{far}}}$. The Bernstein maximal inequality for VC-subgraph classes now upgrades the pointwise estimate to the whole class. Its entropy cost is $O(\log\rho_n)$, whereas its deviation exponent is of order $e^{\ell_n^{\mathrm{far}}} = \rho_n^8$. Hence

$$\sup_{r\in\mathcal{R}_n(A),\, v\in S^1}|H_n^{\mathrm{diff},0}(r,v) - \mathbb{E}H_n^{\mathrm{diff},0}(r,v)| \xrightarrow{\mathbb{P}} 0.$$

It remains to control the removed layer. Write $Z_i = R_iU_i$ and $Y_i(v) = \rho_n\,\mathrm{ang}(U_i,v)$. If $\xi_i > 0$ and $v\cdot Z_i \le \rho_n$, then

$$|Z_i - rv|^2 \ge |Z_i|^2 + r^2 - 2r\rho_n = 2\xi_i + (r-\rho_n)^2,$$

so every such positive-edge term is bounded by $e^{-A^2/2}$, uniformly over $r\in\mathcal{R}_n(A)$. Thus their total contribution is at most $e^{-A^2/2}N_n^+$, where $N_n^+ = \#\{i : \xi_i > 0\} = O_{\mathbb{P}}(1)$.

For $-\ell_n^{\mathrm{far}} < \xi_i \le 0$, $R_i = \rho_n + O(\ell_n^{\mathrm{far}}/\rho_n)$. Uniformly over $r\in\mathcal{R}_n(A)$, for all large n,

$$|Z_i - rv|^2 \ge cA^2 + cY_i(v)^2.$$

Therefore the negative part of the removed layer is bounded by

$$e^{-cA^2}\sup_{v\in S^1}\sum_{-\ell_n^{\text{far}}<\xi_i\le 0} e^{\xi_i}e^{-cY_i(v)^2}.$$

The last supremum is tight. Indeed, partition $(-\ell_n^{\text{far}},0]$ into the slabs $-k-1<\xi_i\le -k$, $0\le k\le\lceil\ell_n^{\text{far}}\rceil$, and cover S^1 by $O(\rho_n)$ arcs of length $O(\rho_n^{-1})$. If $N_{k\ell}$ denotes the number of points in the k-th slab and in the ℓ-th enlarged arc, then the Gaussian angular kernel has summable tails and

$$\sup_v \sum_{-\ell_n^{\text{far}}<\xi_i\le 0} e^{\xi_i}e^{-cY_i(v)^2}\le C\sum_{k\le\lceil\ell_n^{\text{far}}\rceil} e^{-k}\max_\ell N_{k\ell}.$$

Here $\mathbb{E}N_{k\ell}\le Ce^k/\rho_n$. Chernoff's bound and a union bound over the $O(\rho_n)$ arcs give, with probability tending to one,

$$\max_\ell N_{k\ell}\le C\left(\frac{e^k}{\rho_n}+k+1\right)\qquad\text{for all } k\le\lceil\ell_n^{\text{far}}\rceil.$$

Consequently the displayed shot-noise supremum is $O_{\mathbb{P}}(1)$. We have proved

$$\sup_{r\in\mathcal{R}_n(A),\,v\in S^1}\{H_n^{\text{diff}}(r,v)-H_n^{\text{diff},0}(r,v)\}\le e^{-A^2/2}O_{\mathbb{P}}(1)+e^{-cA^2}O_{\mathbb{P}}(1).$$

On the lower side, $\mathbb{E}H_n^{\text{diff},0}(r,v)\le 1$, so the first assertion follows from the concentration estimate and then $A\to\infty$. At the upper boundary, $\mathbb{E}H_n^{\text{diff},0}(\rho_n+A,v)\le\mathbb{E}H_n^{\text{diff}}(\rho_n+A,v)=\Phi(-A)$, and $\Phi(-A)\to 0$. The concentration and removed-layer bounds above therefore prove the second assertion as well. □

The tail part T_n is likewise negligible away from $|r-\rho_n|=O(1)$.

Lemma 3.4. *For every $\delta>0$,*

$$\lim_{A\to\infty}\limsup_{n\to\infty}\mathbb{P}\left\{\sup_{\substack{r\ge\rho_n/2,\ v\in S^1\\ |r-\rho_n|\ge A}} T_n(r,v)>\delta\right\}=0.$$

Proof. Put $\xi_i=|Z_i|^2/2-\log n$ and $N_n^+=\#\{i:\xi_i>0\}$. Since $|Z_i|^2/2$ is exponential in dimension two, $N_n^+\Rightarrow\mathrm{Pois}(1)$ and

$$\mathbb{P}\{\max_i\xi_i>A^2/4\}\to 1-\exp\{-e^{-A^2/4}\}.$$

Hence the event

$$\max_i\xi_i\le A^2/4,\qquad N_n^+\le e^{A^2/8}$$

has probability tending to one as first $n\to\infty$ and then $A\to\infty$. If $v\cdot Z_i>\rho_n$, then $\xi_i>0$. On the displayed event, write $R_i=|Z_i|$. For $r\le\rho_n-A$,

$$r\,v\cdot Z_i-\frac{r^2}{2}-\log n\le(\rho_n-A)R_i-\frac{(\rho_n-A)^2}{2}-\log n\le\xi_i-\frac{A^2}{2}+o(1).$$

For $r\ge\rho_n+A$, the maximum over r is attained either at $r=\rho_n+A$ or at $r=v\cdot Z_i$. Since $\xi_i\le A^2/4$ implies $R_i=\rho_n+O(A^2/\rho_n)<\rho_n+A$ for large n, the boundary case applies and gives

$$r\,v\cdot Z_i-\frac{r^2}{2}-\log n\le(\rho_n+A)R_i-\frac{(\rho_n+A)^2}{2}-\log n\le\xi_i-\frac{A^2}{2}+o(1)\le -A^2/4+o(1).$$

Therefore

$$\sup_{\substack{r \geq \rho_n/2,\ v \in S^1 \\ |r-\rho_n| \geq A}} T_n(r, v) \leq N_n^+ e^{-A^2/4+o(1)} \leq e^{-A^2/8+o(1)}$$

on the displayed event, which proves the lemma. □

An ordinary bulk estimate removes the remaining range $0 < |t| \leq \rho_n/2$.

Lemma 3.5. *Assume $d = 2$ and $\lambda_n = \kappa_n \rho_n^2 \to \lambda \in (0, \infty)$. Then*

$$\mathbb{P}\left\{ \sup_{0<|t|\leq \rho_n/2} D_n(t) > 1 \right\} \to 0.$$

Proof. We repeat the three estimates behind Proposition 4.10, proved for fixed $d \geq 3$ in Section 4.4, and verify them at the two-dimensional scale $\kappa_n \asymp \rho_n^{-2} \asymp 1/\log n$.

First consider $0 < |t| \leq c_0\kappa_n$. The sample covariance satisfies

$$\left\| \frac{1}{n} \sum_i (Z_i - \bar{Z}_n)(Z_i - \bar{Z}_n)^\top - I_2 \right\|_{\mathrm{op}} = O_{\mathbb{P}}(n^{-1/2}) = o_{\mathbb{P}}(\kappa_n).$$

The Hessian argument in Lemma 4.7 therefore applies unchanged: for a sufficiently small fixed $c_0 > 0$, with probability tending to one,

$$D_n(t) \leq 1 - c\kappa_n |t|^2 < 1 \qquad (0 < |t| \leq c_0\kappa_n).$$

For $c_0\kappa_n \leq |t| \leq 1$, use the quadratic empirical-process expansion from Lemma 4.9. The only scale conditions needed there are

$$\frac{n^{-1/4}}{\kappa_n} \to 0, \qquad \frac{|\bar{Z}_n|}{\kappa_n^2} = O_{\mathbb{P}}(n^{-1/2}\rho_n^4) \to 0.$$

They imply, uniformly on this range,

$$\frac{|H_n(t) - 1| + |t \cdot \bar{Z}_n|}{1 - e^{-\kappa_n |t|^2}} = o_{\mathbb{P}}(1).$$

For $1 \leq |t| \leq \rho_n/2$, repeat the dyadic decomposition and clipping argument from Lemma 4.9. Explicitly, partition the squared radii into dyadic intervals $[x_j, x_{j+1}]$, put

$$R_j = x_{j+1}^{1/2}, \qquad \tau_n^2 = 20\log(1/\kappa_n) + 100\log\rho_n, qquad B_j = \exp\{R_j^2/2 + R_j\tau_n\},$$

and clip each likelihood ratio at B_j. Here $|t|^2 \leq (\log n)/2$, so the variance and envelope bounds are $e^{|t|^2} \leq n^{1/2}$ and $B_j \leq n^{1/4+o(1)}$. The net entropy is still $O(\log n)$, while the Bernstein exponent is at least

$$n^{1/2-o(1)}\kappa_n^2 \gg \log n.$$

Thus there is a deterministic $\delta_n \downarrow 0$ such that, with probability tending to one,

$$|H_n(t) - 1| + |t \cdot \bar{Z}_n| \leq \delta_n g(t), \qquad g(t) = 1 - e^{-\kappa_n |t|^2},$$

uniformly on $c_0\kappa_n \leq |t| \leq \rho_n/2$. (In particular, $\rho_n|\bar{Z}_n| = o_{\mathbb{P}}(1)$.) The algebra in the proof of Proposition 4.10 then gives

$$D_n(t) \leq e^{\delta_n g(t)}\{1 - g(t)\}\{1 + \delta_n g(t)\} < 1$$

uniformly on that range for all large n. Together with the microscopic estimate, this proves the lemma. □

Proof of Proposition 3.2. Lemma 3.5 controls the bulk range:

$$\mathbb{P}\left\{\sup_{0<|t|\le\rho_n/2} D_n(t) > 1\right\} \to 0.$$

For the upper exterior, use the centered representation

$$D_n(rv) = \frac{1}{n}\sum_i \exp\left\{rv\cdot(Z_i-\bar{Z}_n) - \frac{1+2\kappa_n}{2}r^2\right\}.$$

With probability tending to one, $\max_i |Z_i-\bar{Z}_n| < \rho_n + A/2$. On this event every summand is decreasing in r for $r \ge \rho_n + A$, so the upper exterior reduces to the boundary $r = \rho_n + A$. Since $\rho_n|\bar{Z}_n| = o_{\mathbb{P}}(1)$,

$$\sup_v D_n((\rho_n+A)v) \le e^{-\lambda+o_{\mathbb{P}}(1)} \sup_v \{H_n^{\mathrm{diff}}(\rho_n+A,v) + T_n(\rho_n+A,v)\}.$$

The upper-bound parts of Lemmas 3.3 and 3.4 make the right side smaller than one with probability tending to one after $A\to\infty$.

It remains to consider $\rho_n/2 \le r \le \rho_n - A$. On this range

$$e^{-\kappa_n r^2} \le e^{-\lambda/4+o(1)}, \qquad \sup_{r,v} e^{-rv\cdot\bar{Z}_n} = e^{o_{\mathbb{P}}(1)}.$$

Lemma 3.3 gives $H_n^{\mathrm{diff}}(r,v) \le 1 + o_{\mathbb{P}}(1) + o_A(1)$ uniformly, and Lemma 3.4 gives $T_n(r,v) = o_{\mathbb{P}}(1) + o_A(1)$ uniformly. Since $e^{-\lambda/4} < 1$, the lower exterior also cannot contain a crossing once A is large. □

3.3 Compact-window decomposition

We now approximate the full field on the compact radial window. Observations below a slowly receding radial cutoff form the deterministic background; those above it produce narrow angular kernels.

Proposition 3.6. *Assume $d=2$, put $\rho_n = \sqrt{2\log n}$, and let $K\subset\mathbb{R}$ be compact. Let $\operatorname{ang}(u,v)\in(-\pi,\pi]$ denote the principal signed angular displacement from v to u. For $Z_i\neq 0$, write $Z_i = R_iU_i$, and set*

$$\xi_i = \frac{R_i^2}{2} - \log n, \qquad Y_i(v) = \rho_n \operatorname{ang}(U_i,v),$$

so that $Y_i(v)$ is the blown-up signed angular displacement from v to the observation direction U_i. If $I\subset S^1$ is a closed arc and $\ell_n^{\mathrm{edge}} = 4\log\rho_n$, then

$$\sup_{v\in I,\ s\in K}\left| H_n((\rho_n+s)v) - \Phi(-s) - \sum_{i:\xi_i\ge -\ell_n^{\mathrm{edge}}} \exp\left\{\xi_i - \frac{s^2}{2} - \frac{Y_i(v)^2}{2}\right\}\right| \xrightarrow{\mathbb{P}} 0.$$

Proof. Write $r_s = \rho_n + s$. First separate the diffuse contribution below the radial edge. For $L>0$, define

$$H_{n,L}^{\mathrm{diff}}(v,s) = \frac{1}{n}\sum_{i=1}^n \exp\left\{r_s v\cdot Z_i - \frac{r_s^2}{2}\right\}\mathbf{1}_{\{\xi_i<-L\}}.$$

We use $L = \ell_n^{\mathrm{edge}}$. If $\xi_i < -L$, then $R_i \le R_L = (\rho_n^2 - 2L)^{1/2}$, and uniformly for $s\in K$,

$$\frac{1}{n}\exp\left\{r_s v\cdot Z_i - \frac{r_s^2}{2}\right\} \le \exp\left\{-L - \frac{(r_s-R_L)^2}{2}\right\} \le C_K e^{-L}.$$

The sum of the variances is therefore at most $C_K e^{-L}$. Bernstein's inequality gives, for each fixed (v,s),

$$\mathbb{P}\{|H^{\mathrm{diff}}_{n,L}(v,s) - \mathbb{E}H^{\mathrm{diff}}_{n,L}(v,s)| > \eta\} \le 2\exp\{-c_{K,\eta}e^{L}\}.$$

The field $H^{\mathrm{diff}}_{n,L}$ is smooth in (v,s). Each summand has logarithmic derivatives bounded by $C_K\rho_n^2$ in the angular coordinate and by $C_K\rho_n$ in s. Fix $\eta > 0$, and choose an angular mesh of order ρ_n^{-2} and an s-mesh of order ρ_n^{-1}. The mesh constants can be chosen so that neighboring values of both $H^{\mathrm{diff}}_{n,L}$ and $\mathbb{E}H^{\mathrm{diff}}_{n,L}$ differ by a factor at most $1+\eta$. The grid has polynomial size in ρ_n, so a union bound upgrades the pointwise Bernstein estimate to the grid. Because $\mathbb{E}H^{\mathrm{diff}}_{n,L} \le 1$, interpolation from the grid costs only an additive $O(\eta)$; letting $\eta \downarrow 0$ gives

$$\sup_{v\in I,\ s\in K} |H^{\mathrm{diff}}_{n,\ell_n^{\mathrm{edge}}}(v,s) - \mathbb{E}H^{\mathrm{diff}}_{n,\ell_n^{\mathrm{edge}}}(v,s)| \xrightarrow{\mathbb{P}} 0,$$

because $e^{\ell_n^{\mathrm{edge}}} = \rho_n^4$ dominates the grid entropy.

It remains to compute the mean. By exponential tilting,

$$\mathbb{E}H^{\mathrm{diff}}_{n,L}(v,s) = \mathbb{P}\{|G + r_s v|^2 \le \rho_n^2 - 2L\}, \qquad G \sim \mathcal{N}(0, I_2).$$

After rotating $v = e_1$, this probability is

$$\mathbb{P}\left\{(r_s + G_1)^2 + G_2^2 \le \rho_n^2 - 2L\right\}.$$

Condition on G_2. On the event $|G_2| \le \rho_n^{1/3}$, the boundary for G_1 is

$$G_1 \le -s - \frac{\ell_n^{\mathrm{edge}} + G_2^2/2}{\rho_n} + O_K\left(\frac{(\ell_n^{\mathrm{edge}} + G_2^2 + 1)^2}{\rho_n^3}\right),$$

uniformly for $s \in K$. The quadratic inequality also has the lower boundary

$$G_1 \ge -r_s - \sqrt{\rho_n^2 - 2\ell_n^{\mathrm{edge}} - G_2^2} = -2\rho_n + O_K(1)$$

on this event, whose Gaussian probability is uniformly $o(1)$. The complement of $\{|G_2| \le \rho_n^{1/3}\}$ also has probability $o(1)$, and the displayed upper boundary converges to $-s$ uniformly on K after conditioning on G_2. Dominated convergence and the Lipschitz continuity of the Gaussian distribution function give

$$\sup_{v\in I,\ s\in K} |\mathbb{E}H^{\mathrm{diff}}_{n,\ell_n^{\mathrm{edge}}}(v,s) - \Phi(-s)| \to 0.$$

Thus the contribution from $\{\xi_i < -\ell_n^{\mathrm{edge}}\}$ is $\Phi(-s) + o_{\mathbb{P}}(1)$, uniformly on $I \times K$.

Now consider the edge contribution

$$A_n^{\mathrm{ex}}(v,s) = \sum_{i:\xi_i \ge -\ell_n^{\mathrm{edge}}} \frac{1}{n}\exp\left\{r_s v\cdot Z_i - \frac{r_s^2}{2}\right\}.$$

We compare it with the displayed atom sum. Let

$$C_n = \log\rho_n, \qquad Y_n = (\log\rho_n)^{1/4}.$$

The event $\max_i \xi_i \le C_n$ has probability tending to one, since $n\mathbb{P}\{\xi_i > C_n\} = e^{-C_n}$. Also

$$\sum_{i:-\ell_n^{\mathrm{edge}} \le \xi_i \le C_n} e^{\xi_i} = O_{\mathbb{P}}(\ell_n^{\mathrm{edge}} + C_n),$$

because ξ_i has shifted exponential intensity $e^{-x}\,dx$ in dimension two.

On the event $\max_i \xi_i \le C_n$, the exact contribution of an edge observation with $-\ell_n^{\mathrm{edge}} \le \xi_i \le C_n$ satisfies

$$\frac{1}{n}\exp\left\{r_s v\cdot Z_i - \frac{r_s^2}{2}\right\} \le C_K e^{\xi_i}\exp\{-c\min(Y_i(v)^2,\rho_n^2)\}.$$

This follows from

$$r_s v\cdot Z_i - \frac{r_s^2}{2} - \log n = \xi_i - \frac{(R_i-r_s)^2}{2} - r_s R_i\{1-\cos\operatorname{ang}(U_i,v)\},$$

and $r_s R_i \asymp \rho_n^2$ in the displayed range. Hence the total contribution of terms with $|Y_i(v)| > Y_n$, in either A_n^{ex} or the Gaussian atom sum, is

$$O_{\mathbb{P}}(\ell_n^{\mathrm{edge}} + C_n)e^{-cY_n^2} = o_{\mathbb{P}}(1)$$

uniformly in $v \in I$ and $s \in K$, by the definition of Y_n. For the remaining terms $-\ell_n^{\mathrm{edge}} \le \xi_i \le C_n$ and $|Y_i(v)| \le Y_n$, Taylor expansion gives, uniformly in i, v, s,

$$\begin{aligned}\log\left[\frac{1}{n}\exp\left\{r_s v\cdot Z_i - \frac{r_s^2}{2}\right\}\right] - \left\{\xi_i - \frac{s^2}{2} - \frac{Y_i(v)^2}{2}\right\}\\ = O_K\left(\frac{\ell_n^{\mathrm{edge}}+C_n}{\rho_n} + \frac{Y_n^2}{\rho_n} + \frac{Y_n^4}{\rho_n^2}\right) = O_K\left(\frac{\log\rho_n}{\rho_n}\right).\end{aligned}$$

The total atom weight is $O_{\mathbb{P}}(\log\rho_n)$, while the relative error is $O(\log\rho_n/\rho_n)$. Their product is $o_{\mathbb{P}}(1)$, and hence

$$\sup_{v\in I,\ s\in K}\left|A_n^{\mathrm{ex}}(v,s) - \sum_{i:\xi_i\ge -\ell_n^{\mathrm{edge}}}\exp\left\{\xi_i - \frac{s^2}{2} - \frac{Y_i(v)^2}{2}\right\}\right| \xrightarrow{\mathbb{P}} 0.$$

Combining this with $H_n = H^{\mathrm{diff}}_{n,\ell_n^{\mathrm{edge}}} + A_n^{\mathrm{ex}}$ proves the proposition. □

Proposition 3.6 uses the slowly receding cutoff $\ell_n^{\mathrm{edge}} = 4\log\rho_n$, which makes the background concentrate uniformly. For the final supremum bound, we replace it by a fixed cutoff. For every fixed $A < \infty$, the edge points with $\xi_i > -A$ are finite in the limit and have no $1/\rho_n$-scale angular collisions:

$$\mathbb{P}\left\{\exists i \ne j:\ \xi_i,\xi_j > -A,\ \rho_n|\operatorname{ang}(U_i,U_j)| \le L\right\} \le C_A L/\rho_n \to 0.$$

Edge points in the lower radial strip $-\ell_n^{\mathrm{edge}} \le \xi_i \le -A$ have negligible total contribution after $A \to \infty$, allowing a fixed cutoff.

Lemma 3.7. *Let $\ell_n^{\mathrm{edge}} = 4\log\rho_n$, and write $Y_i(v) = \rho_n\operatorname{ang}(U_i,v)$ for the principal angular displacement between U_i and v. For every $\delta > 0$,*

$$\lim_{A\to\infty}\limsup_{n\to\infty}\mathbb{P}\left\{\sup_{v\in S^1}\sum_{i:-\ell_n^{\mathrm{edge}}\le\xi_i\le -A} e^{\xi_i - Y_i(v)^2/2} > \delta\right\} = 0.$$

Proof. Put $m_n^\star = \lceil \ell_n^{\mathrm{edge}}\rceil$ and, for integers $m \ge 0$,

$$I_{m,n} = \{i : -m-1 < \xi_i \le -m\}.$$

These are unit radial-excess layers; on $I_{m,n}$, the weight e^{ξ_i} is of order e^{-m}. Since $|Z_i|^2/2$ is exponential in dimension two,

$$\mathbb{P}\{i \in I_{m,n}\} \le Ce^m/n, \qquad m \le m_n^\star,$$

and the angle Θ_i is independent and uniform on the circle.

For $q = 0, 1, \dots, \lfloor \pi\rho_n \rfloor$, let

$$N_{m,q,n} = \sup_{\phi \in \mathbb{T}} \# \left\{ i \in I_{m,n} : \rho_n |\operatorname{ang}(\Theta_i - \phi)| \le q + 1 \right\}.$$

Covering the circle by $O(\rho_n)$ arcs shows that every interval appearing in the supremum is contained in one of the covering arcs, with length enlarged by only a fixed factor. For one such enlarged arc the count is binomial with mean

$$\mu_{m,q,n} \le C(q+1)e^m/\rho_n.$$

Set

$$a_{m,q,n} = 1 + \frac{(q+1)e^m}{\rho_n} + \frac{\log \rho_n}{1 + \log_+(\rho_n/((q+1)e^m))}, \qquad \log_+ x = \max\{\log x, 0\}.$$

The binomial Chernoff bound $\mathbb{P}\{\mathrm{Bin}(N,p) \ge r\} \le (eNp/r)^r$ for $r \ge eNp$, together with the multiplicative Chernoff bound when r is a fixed multiple of Np, implies that for a sufficiently large numerical constant B,

$$\mathbb{P}\left\{ N_{m,q,n} > Ba_{m,q,n} \text{ for some } A \le m \le m_n^\star,\ 0 \le q \le \pi\rho_n \right\} \to 0$$

for each fixed A. Indeed, if $\mu_{m,q,n} < 1$, the last term in $a_{m,q,n}$ makes

$$a_{m,q,n} \log\{a_{m,q,n}/\mu_{m,q,n}\} \ge c \log \rho_n,$$

while if $\mu_{m,q,n} \ge 1$, the term $\log \rho_n$ gives the same union-bound margin unless $\mu_{m,q,n} \ge \log \rho_n$, in which case the linear mean term does. Choosing B large makes the tail summable over the $O(\rho_n^2 \log \rho_n)$ enlarged arcs, layers, and angular scales.

On this high-probability occupancy event, uniformly in v,

$$\begin{aligned}
\sum_{i: -\ell_n^{\mathrm{edge}} \le \xi_i \le -A} e^{\xi_i - Y_i(v)^2/2} &\le C \sum_{m=\lfloor A \rfloor}^{m_n^\star} e^{-m} \sum_{q=0}^{\lfloor \pi\rho_n \rfloor} e^{-q^2/2} N_{m,q,n} \\
&\le C \sum_{q=0}^{\lfloor \pi\rho_n \rfloor} e^{-q^2/2} \sum_{m=\lfloor A \rfloor}^{m_n^\star} e^{-m} a_{m,q,n}.
\end{aligned}$$

The first two terms in $a_{m,q,n}$ contribute

$$O(e^{-A}) + O(\ell_n^{\mathrm{edge}}/\rho_n).$$

For the logarithmic occupancy term, fix A and split the m-sum at $\log\{\rho_n/(q+1)\}$. This gives the deterministic bound

$$\sum_{m=\lfloor A \rfloor}^{m_n^\star} e^{-m} \frac{\log \rho_n}{1 + \log_+(\rho_n/((q+1)e^m))} \le Ce^{-A}\{1 + \log(q+2)\} + C\frac{(q+1)\log \rho_n}{\rho_n}$$

for all $n \ge n_0(A)$, uniformly in q. This fixed-A qualification is essential; the estimate is not asserted for a simultaneous choice $A = A_n$. After summing in q against $e^{-q^2/2}$, the right side is

$$O(e^{-A}) + o(1).$$

Thus the displayed supremum is at most $Ce^{-A} + o(1)$ with probability tending to one, and the lemma follows by letting $A \to \infty$. □

The moving-cutoff decomposition can therefore be replaced by the fixed-cutoff form needed for the reverse inequality.

Lemma 3.8. *For each compact $K \subset \mathbb{R}$, each $C < \infty$, and each $\delta > 0$,*

$$\lim_{A\to\infty} \limsup_{n\to\infty} \mathbb{P}\Bigg\{ \max_i \xi_i \le C,\ \sup_{\substack{v\in S^1\\ s\in K}} \Bigg| H_n((\rho_n + s)v) - \Phi(-s) - \sum_{i:-A<\xi_i\le C} e^{\xi_i - s^2/2 - Y_i(v)^2/2} \Bigg| > \delta \Bigg\} = 0.$$

Proof. Apply Proposition 3.6 on a finite closed-arc cover of S^1. On $\{\max_i \xi_i \le C\}$, its fixed-cutoff sum omits only the observations with $-\ell_n^{\mathrm{edge}} \le \xi_i \le -A$. Their absolute contribution is bounded uniformly in $s \in K$ by

$$\sup_{v\in S^1} \sum_{i:-\ell_n^{\mathrm{edge}} \le \xi_i \le -A} e^{\xi_i - Y_i(v)^2/2},$$

because $e^{-s^2/2} \le 1$. Lemma 3.7 makes this term vanish after first taking $n \to \infty$ and then $A \to \infty$, proving the fixed-cutoff form. □

4 The Spatial Edge in Dimensions $d_n \ge 3$

This section proves the spatial-exclusion estimates needed for the $d_n \ge 3$ clause of Theorem 1.1. The exact Gamma normalization gives the same radial Poisson limit in every growth regime, but the argument showing that subthreshold observations cannot create additional KKT contacts changes with dimension. We first derive the common radial coordinate, then prove the required exclusion in bounded and growing dimensions below the simplex scale. Section 5 treats the remaining high-dimensional regimes.

4.1 The radial coordinate

The exact coordinate $s_n^\star$ appears in Theorem 1.1. In the range $d_n \log n = o(n^2)$, the spatial estimates are cleaner in the following $\log n$-based coordinate:

$$s_n = \frac{\gamma_n L_n - b_n}{a_n}. \tag{4.1}$$

The corresponding critical variance defect is

$$\varepsilon_{c,n} = 1 - \frac{L_n}{b_n}, \tag{4.2}$$

and an $O(1)$ change in s_n changes ε_n on the scale

$$\Delta\varepsilon_n \asymp \frac{a_n L_n}{b_n^2}. \tag{4.3}$$

Lemma 6.14 shows that s_n agrees with the exact coordinate $s_n^\star$ whenever $d_n \log n = o(n^2)$. Define

$$\xi_{n,i} = \frac{|\widetilde{Z}_i|^2/2 - b_n}{a_n}, \qquad \delta_n = \frac{a_n}{\gamma_n}.$$

At the normalization (4.1), the bump height is exactly

$$h_i = \exp\{\delta_n(\xi_{n,i} - s_n)\}. \tag{4.4}$$

The score radius satisfies the exact identity

$$|q_i|^2 = \frac{|\widetilde{Z}_i|^2}{\gamma_n} = 2L_n + 2\delta_n(\xi_{n,i} - s_n). \tag{4.5}$$

Consequently, $|q_i|^2 = 2L_n + O_{\mathbb{P}}(\delta_n)$ for edge observations with $\xi_{n,i} = O_{\mathbb{P}}(1)$ whenever $s_n = O(1)$. Thus the top bumps live on a sphere of radius $\sqrt{2L_n}$. The dimension affects their angular geometry and the height scale δ_n, but not their leading score radius.

When $d_n \gg L_n$, however, $\delta_n \to 0$, so a radial gap of $a_n\eta$ below threshold creates only a height gap of order $\delta_n\eta$. The remainder of the field must therefore be $o_{\mathbb{P}}(\delta_n)$, not merely $o_{\mathbb{P}}(1)$.

4.2 Consequences across dimension scales

The exact definitions (2.7) are preferable in the main theorem, but their asymptotic expansions reveal how the critical defect and window width depend on dimension. In this subsection $\varepsilon_{c,n} = 1 - L_n/b_n$ denotes a leading-order proxy for the critical defect. Section 5.2 restores the finite-sample correction that becomes visible when $d_n \asymp n^2 \log n$.

Sublogarithmic dimension.

Suppose $k_n = o(L_n)$. Then $b_n/L_n \to 1$, $a_n \to 1$, and

$$b_n = L_n + (k_n - 1)\log L_n - \log\Gamma(k_n) + O\left(\frac{k_n}{L_n}\{1 + (k_n - 1)\log L_n - \log\Gamma(k_n)\}\right). \tag{4.6}$$

When in addition $k_n \to \infty$, Stirling's formula gives the simpler leading term

$$b_n - L_n \sim k_n \log\frac{L_n}{k_n}. \tag{4.7}$$

Consequently,

$$\varepsilon_{c,n} \sim \frac{k_n}{L_n}\log\frac{L_n}{k_n} = \frac{d_n}{2\log n}\log\frac{2\log n}{d_n}, \qquad \Delta\varepsilon_n \asymp \frac{1}{\log n}. \tag{4.8}$$

For fixed $d \geq 3$, (4.6) reduces to

$$b_n = \log n + \left(\frac{d}{2} - 1\right)\log\log n - \log\Gamma(d/2) + o(1),$$

recovering the fixed-dimensional normalization used in Section 4.3. If $d_n \to \infty$, then $b_n - L_n \to \infty$; this makes the diffuse edge background vanish, leaving the radial maximum as the limiting obstruction.

Justification of (4.6). For $y > k_n - 1$, (2.10) and $1 \leq \overline{F}_n(y)/f_n(y) \leq \{1 - (k_n - 1)/y\}^{-1}$ give

$$\log\overline{F}_n(y) = -y + (k_n - 1)\log y - \log\Gamma(k_n) + O(k_n/y).$$

Standard Gamma Chernoff bounds first give $b_n/L_n \to 1$. Substitute $y = b_n$, write $b_n = L_n + r_n$, and expand $\log b_n = \log L_n + O(r_n/L_n)$. The resulting equation first gives $r_n = O((k_n - 1)\log L_n - \log\Gamma(k_n) + 1)$, and a second substitution gives (4.6). Formula (4.7) follows because $L_n/k_n \to \infty$. □

Logarithmic dimension.

Suppose $k_n/L_n \to c \in (0,\infty)$. Let $x_c > 1$ be the unique solution of

$$x_c - 1 - \log x_c = \frac{1}{c}. \tag{4.9}$$

Stirling's formula and the Gamma saddlepoint estimate yield

$$\frac{b_n}{k_n} \to x_c, \qquad a_n \to \frac{x_c}{x_c - 1}, \qquad \varepsilon_{c,n} \to 1 - \frac{1}{c x_c}. \tag{4.10}$$

The transition therefore occurs at a nonzero constant variance defect, while its width in ε_n remains of order $1/\log n$.

Superlogarithmic dimension.

Suppose $k_n/L_n \to \infty$. The Gamma quantile is in its moderate-deviation range:

$$b_n = k_n + \sqrt{2k_nL_n}\{1+o(1)\}, \qquad a_n = \sqrt{\frac{k_n}{2L_n}}\{1+o(1)\}. \tag{4.11}$$

Hence

$$\varepsilon_{c,n} = 1 - \frac{L_n}{k_n}\{1+o(1)\}, \qquad \Delta\varepsilon_n \asymp \frac{\sqrt{L_n}}{k_n^{3/2}}, \qquad \delta_n \asymp \sqrt{\frac{L_n}{k_n}}. \tag{4.12}$$

The fitted variance must now be much larger than the data variance: the critical defect tends to one. The shrinking factor δ_n in (4.12) is also the precision required of the single-bump approximation.

In particular, the usual proportional-dimensional regime $d_n/n \to \tau \in (0,\infty)$ is a special case of the superlogarithmic regime, not the logarithmic regime $d_n \asymp \log n$. In that case,

$$1 - \varepsilon_{c,n} \sim \frac{2\log n}{d_n} \sim \frac{2\log n}{\tau n}, \qquad \Delta\varepsilon_n \asymp \frac{\sqrt{\log n}}{n^{3/2}}.$$

The raw squared radii are of order n, but after the variance rescaling the likelihood of one empirical bump still has to overcome its factor $1/n$; this is why the KKT threshold remains $\log n$.

For completeness, all three regimes follow from the same saddlepoint formula. If $k_n \to \infty$, put $x_n = b_n/k_n > 1$. Uniformly at the extreme quantile,

$$\overline{F}_n(k_nx_n) = \frac{1+o(1)}{(x_n-1)\sqrt{2\pi k_n}} \exp\{-k_n(x_n - 1 - \log x_n)\}. \tag{4.13}$$

Indeed, Stirling's formula gives the density at k_nx_n, while Lemma 2.7 gives the tail-to-density ratio $x_n/(x_n-1)\{1+o(1)\}$. Equating the logarithm of (4.13) to $-L_n$ proves (4.10); when $k_n/L_n \to \infty$, expanding $x - 1 - \log x = (x-1)^2/2 + O((x-1)^3)$ proves (4.11).

Finally, (2.17) is automatic when $k_n = O(L_n)$. If $k_n/L_n \to \infty$, (4.11) reduces it to

$$k_nL_n = o(n). \tag{4.14}$$

4.3 Bounded-dimensional radial patches

This subsection proves the fixed-dimensional spatial exclusion needed for the $d \geq 3$ support law in Theorem 1.1. The moving radial shell identifies the candidate KKT patches, and the uniform estimates in Subsection 4.4 exclude every other score location.

Let $Z_1, \dots, Z_n$ be i.i.d. $\mathcal{N}(0, I_d)$, $d \geq 3$, and set

$$H_n(t) = \frac{1}{n}\sum_{i=1}^{n} \exp\{t \cdot Z_i - |t|^2/2\}, \qquad t \in \mathbb{R}^d.$$

4.3.1 The moving radial shell

Let

$$Y_i = \frac{|Z_i|^2}{2}, \qquad U_i = \frac{Z_i}{|Z_i|} \in S^{d-1}, \qquad k = \frac{d}{2}.$$

Since Y_i has the $\Gamma(k, 1)$ law,

$$b_{n,d} = \log n + (k-1)\log\log n - \log\Gamma(k),$$

the fixed-dimensional specialization of the Gamma extreme-value theory in Section 2.2 gives the marked radial limit.

Lemma 4.1. *The extremal process*

$$\Xi_n = \sum_{i=1}^{n} \delta_{(Y_i - b_{n,d},\, U_i)}$$

converges to a Poisson point process on $\mathbb{R} \times S^{d-1}$ with intensity $e^{-x} dx\, \sigma_{d-1}(du)$, where σ_{d-1} is normalized surface measure. In particular, $M_n = \max_i (Y_i - b_{n,d})$ satisfies

$$\mathbb{P}\{M_n \le x\} \to \exp\{-e^{-x}\}.$$

Proof. Proposition 2.8 and Lemma 2.9 give the unmarked radial limit. Radius and direction are independent, and U_i has law σ_{d-1}, so the standard marking theorem gives the displayed point-process convergence. □

Put $\rho_n = \sqrt{2b_{n,d}}$ and

$$\lambda_n = \kappa_n \rho_n^2.$$

In the critical window $\lambda_n = O(\log\log n)$, and it is useful to write

$$s_n = \lambda_n - (k-1)\log\log n + \log\Gamma(k).$$

Completing the square in one summand of D_n, or equivalently using Lemma 2.4, shows that an observation with $Y_i = b_{n,d} + x$ has optimized log contribution $x - s_n + o_{\mathbb{P}}(1)$, uniformly for bounded x. Thus an observation is locally supercritical precisely when its radial excess x exceeds s_n; these observations index the candidate KKT crossings.

To resolve each candidate spatially, write the score and observation in normal coordinates about a direction v. Let $\exp_v : T_v S^{d-1} \to S^{d-1}$ be the Riemannian exponential map and put

$$t = \left(\rho_n + \frac{q}{\rho_n}\right) v, \qquad z = \left(\rho_n + \frac{x}{\rho_n}\right) \exp_v\left(\frac{y}{\rho_n}\right), \qquad y \in T_v S^{d-1}.$$

Then, uniformly for $x, q, |y| = O(\log\log n)$,

$$t \cdot z - \frac{|t|^2}{2} - \log n - \kappa_n |t|^2 = (k-1)\log\log n - \log\Gamma(k) - \lambda_n + x - \frac{|y|^2}{2} + O\left(\frac{(\log\log n)^4}{\rho_n^2}\right).$$

The radial score offset q cancels to first order. A radial excess x therefore raises the optimized contribution by x, whereas an angular displacement y/ρ_n lowers it by $|y|^2/2$; the variance defect subtracts λ_n. The estimates below show that the rest of the sample contributes $o_{\mathbb{P}}(1)$ uniformly on each candidate patch. Subsection 6.2 uses these patch estimates to recover the fitted support.

For the uniform exclusion argument, use the centered radial maximum

$$M_n^c = \max_i \left(\frac{|Z_i - \bar{Z}_n|^2}{2} - b_{n,d} \right).$$

It differs from $M_n = \max_i (Y_i - b_{n,d})$ by $o_{\mathbb{P}}(1)$, since

$$|M_n^c - M_n| \le \max_i |Z_i|\,|\bar{Z}_n| + \frac{|\bar{Z}_n|^2}{2} = o_{\mathbb{P}}(1).$$

4.4 Uniform moving-shell estimates

This subsection supplies the estimates used to prune subthreshold radial patches in Lemma 6.10. Bennett's inequality controls the center of each angular patch, while scaled derivative bounds and Morrey's inequality control oscillation across patches of diameter $1/\rho_n$. The following bulk estimates handle smaller score radii by a clipped empirical-process argument.

For the moving-shell bound, replace the field by an independent truncated field. On the radial-maximum event used in the comparison, every observation lies below the deterministic truncation radius, so the two fields agree.

Fix a gap $a > 0$. Set $r = |t|$,

$$T_n(a) = \log n + \lambda_n - a, \qquad R_n(a) = \{2T_n(a)\}^{1/2},$$

and write $R_a = R_n(a)$. Define

$$V_{i,t}^{(a)} = \frac{1}{n} e^{-\kappa_n r^2} \exp\left\{ t \cdot Z_i - \frac{r^2}{2} \right\} \mathbf{1}_{\{|Z_i| \le R_a\}}, \qquad S_t^{(a)} = \sum_i V_{i,t}^{(a)}.$$

Let

$$\mathcal{T}_n(B) = \left\{ t : \left| |t|^2/2 - b_{n,d} \right| \le B \log\log n \right\}.$$

The centered field is compared to this independent truncated field with a small deterministic loss in the gap. Put

$$\alpha_n = (\log\log n)^{-2}, \qquad G_n = \{\rho_n |\bar{Z}_n| \le \alpha_n\}.$$

Since $\alpha_n \sqrt{n}/\rho_n \to \infty$, $\mathbb{P}(G_n^c) \to 0$, and on G_n, for all large n,

$$\sup_{t \in \mathcal{T}_n(B)} |t \cdot \bar{Z}_n| \le 2\alpha_n.$$

If $M_n^c \le s_n - \eta$, then $|Z_i - \bar{Z}_n|^2/2 \le \log n + \lambda_n - \eta$ for all i. On G_n this implies $|Z_i|^2/2 \le \log n + \lambda_n - \eta/2$ for all large n. Consequently, uniformly over $t \in \mathcal{T}_n(B)$,

$$D_n(t) \le e^{2\alpha_n} S_t^{(\eta/2)}.$$

It is therefore enough to prove that, for fixed $a > 0$, the independent truncated field stays below $1 - o(1)$ uniformly on $\mathcal{T}_n(B)$.

The first input bounds the moments of one truncated summand. Exponential tilting turns each moment into the probability that a noncentral Gaussian lies in the truncation ball; a one-dimensional Gaussian tail estimate then gives the required decay.

Proposition 4.2. *Fix $d \geq 3$, $C_0 < \infty$, $B < \infty$, $p > 1$, and $a > 0$. Suppose the sharp-scale relation*

$$s_n = \lambda_n - \left(\frac{d}{2} - 1\right) \log\log n + \log\Gamma(d/2)$$

satisfies $|s_n| \leq C_0$. Uniformly over deterministic sequences satisfying this relation, over directions $t/|t|$, and over

$$\left|\frac{r^2}{2} - b_{n,d}\right| \leq B \log\log n,$$

one has

$$\sum_{i=1}^{n} \mathbb{E}\{V_{i,t}^{(a)}\}^p \leq C_{p,a,B,C_0,d}\rho_n^{-1} \exp\{-\lambda_n - (p-1)a + o(1)\}.$$

Proof. The exact identity is

$$\sum_{i=1}^{n} \mathbb{E}\{V_{i,t}^{(a)}\}^p = n^{1-p} e^{-p\kappa_n r^2 + (p^2-p)r^2/2} \mathbb{P}\{|G + pt| \leq R_n(a)\},$$

where $G \sim \mathcal{N}(0, I_d)$. Align $t = re_1$, write $G = We_1 + U$, and put $\Delta = pr - R_n(a)$. In the moving window, $\Delta = (p-1)\rho_n + O((\log\log n)/\rho_n)$, so $\Delta \geq c_p\rho_n$. If $|G + pt| \leq R_n(a)$, then

$$W \leq -\left(pr - \sqrt{R_n(a)^2 - |U|^2}\right).$$

Mills' bound and the inequality

$$pr - \sqrt{R_n(a)^2 - |U|^2} \geq \Delta + \frac{|U|^2}{2R_n(a)}$$

give

$$\mathbb{P}\{|G + pt| \leq R_n(a)\} \leq C_{p,d}\rho_n^{-1} \exp\left\{-\frac{(pr - R_n(a))^2}{2}\right\}.$$

Substitution gives

$$\sum_{i=1}^{n} \mathbb{E}\{V_{i,t}^{(a)}\}^p \leq C_{p,d}\rho_n^{-1} \exp\left\{-p\kappa_n r^2 + (p-1)(\lambda_n - a) - \frac{p}{2}(r - R_n(a))^2\right\}.$$

Since the sharp-scale assumption gives $\lambda_n = O(\log\log n)$, one has $\kappa_n r^2 = \lambda_n + o(1)$ uniformly in the displayed shell. This proves the stated estimate. $\square$

Under the same sharp-scale assumptions, the truncation also gives the pointwise jump bound

$$V_{i,t}^{(a)} \leq e^{-a+o(1)}$$

uniformly in the moving shell. Indeed

$$\log V_{i,t}^{(a)} \leq -\log n - \kappa_n r^2 - \frac{r^2}{2} + rR_n(a) = \lambda_n - a - \kappa_n r^2 - \frac{1}{2}(r - R_n(a))^2.$$

Thus $V_{i,t}^{(a)} \leq c_a < 1$ for large n, where c_a is deterministic.

The moment and jump bounds give a fixed-location tail strong enough for the angular union bound after the Morrey step.

Proposition 4.3. *Fix $d \geq 3$, $C_0 < \infty$, $B < \infty$, and $a > 0$, and suppose the sharp-scale relation in Proposition 4.2 satisfies $|s_n| \leq C_0$. Uniformly over deterministic sequences satisfying this relation and over the moving shell, for every deterministic $\delta_n \to 0$ with $0 \leq \delta_n \leq 1/2$ eventually,*

$$\mathbb{P}\{S_t^{(a)} \geq 1 - \delta_n\} \leq C_{a,B,C_0,d} \left(\rho_n^{-1} e^{-\lambda_n}\right)^{e^a(1-\delta_n)-o(1)}.$$

In particular, at level 1 the exponent is $e^a - o(1)$.

Proof. Let $\mu_t = \mathbb{E}S_t^{(a)}$, $v_t = \sum_i \operatorname{Var}(V_{i,t}^{(a)})$, $A_n = \rho_n^{-1}e^{-\lambda_n}$, and $x_n = 1 - \delta_n$. Proposition 4.2 with $p = 2$ gives

$$v_t \leq C\rho_n^{-1} e^{-\lambda_n - a + o(1)}, \qquad \mu_t \leq e^{-\lambda_n + o(1)} = o(1)$$

under the displayed $d \geq 3$ sharp-scale assumption. Bennett's inequality for $0 \leq V_{i,t}^{(a)} \leq c_a < 1$ gives

$$\mathbb{P}\{S_t^{(a)} \geq x_n\} \leq \exp\left[-\frac{v_t}{c_a^2} h\left(\frac{c_a(x_n - \mu_t)}{v_t}\right)\right], \qquad h(u) = (1+u)\log(1+u) - u.$$

Using the sharp jump bound $V_{i,t}^{(a)} \leq e^{-a+o(1)}$ and $\log A_n^{-1} = (d-1)\log \rho_n + O(1)$, the Bennett rate is at least

$$\{e^a(1-\delta_n) - o(1)\} \log A_n^{-1}.$$

This proves the displayed estimate. Only the second moment is needed at a fixed location; the derivative estimates used for the Morrey transfer require higher moments. □

At the sharp scale $s_n = O(1)$, equivalently $\lambda_n = (d-2)\log \rho_n + O(1)$, the quantity $\rho_n^{-1}e^{-\lambda_n}$ is $\rho_n^{-(d-1)+o(1)}$. The strict near-unit exponent $e^a(1 - o(1)) > 1$ is therefore exactly what is needed to union bound over $\rho_n^{d-1+o(1)}$ angular locations. Throughout the remaining $d \geq 3$ estimates, "at the sharp scale $s_n = O(1)$" means uniformly over deterministic sequences satisfying $|s_n| \leq C$, for each fixed $C < \infty$; constants may depend on C.

Scaled derivatives control oscillation on angular boxes at scale $1/\rho_n$, reducing the continuum to a net of size $\rho_n^{d-1+o(1)}$.

Lemma 4.4. *Fix $d \geq 3$, $p > d$, $a > 0$, and suppose $s_n = O(1)$. Let $Q \subset \mathbb{R}^{d-1}$ be a fixed cube, and let $\{u_\alpha\}$ be a bounded-overlap family of smooth charts such that $\{u_\alpha(Q/\rho_n)\}$ covers S^{d-1}, the number of charts is $O(\rho_n^{d-1})$, and all first two derivatives of the charts are bounded uniformly in α and n. For each fixed $B < \infty$, in moving-shell coordinates*

$$t = t(q,y) = \sqrt{2(b_{n,d} + q)}\, u_\alpha(y/\rho_n), \qquad |q| \leq B \log\log n, \quad y \in Q,$$

one has

$$\mathbb{E} \sum_\alpha \int_{|q| \leq B \log\log n} \int_Q |\nabla_{q,y} S_{t(q,y)}^{(a)}|^p \, dy\, dq \leq (\log\log n)^C.$$

The displayed derivative-moment bound also holds on every fixed physical-offset annulus

$$t = t(h,y) = (R_a - h)u_\alpha(y/\rho_n), \qquad |h| \leq A, \quad y \in Q,$$

with $C = C(a, A, B, d, p)$. Consequently, in either region and for every fixed $\gamma > 0$, a net of cardinality

$$\rho_n^{d-1}(\log\log n)^{C_\gamma} = \rho_n^{d-1+o(1)}$$

suffices, with probability tending to one, to approximate $S_t^{(a)}$ up to the prescribed error

$$\varepsilon_{n,\gamma} = (\log\log n)^{-\gamma}.$$

Proof. The truncation set $\{|Z_i| \le R_a\}$ is independent of t, so $S_t^{(a)}$ is C^1 in the displayed coordinates. For any coordinate $\xi \in \{q, y_1, \dots, y_{d-1}\}$ or $\xi \in \{h, y_1, \dots, y_{d-1}\}$,

$$\partial_\xi V_{i,t}^{(a)} = V_{i,t}^{(a)} M_\xi(Z_i, t), \qquad M_\xi(z,t) = \{z - (1+2\kappa_n)t\} \cdot \partial_\xi t.$$

The coordinate choices give

$$|\partial_q t| = O(\rho_n^{-1}), \qquad |\partial_h t| = 1, \qquad |\partial_{y_j} t| = O(1), \qquad t \cdot \partial_{y_j} t = 0.$$

For $m \in \{2, p\}$, exponential tilting gives the exact identity

$$\sum_i \mathbb{E}|\partial_\xi V_{i,t}^{(a)}|^m = n^{1-m} e^{-m\kappa_n r^2 + (m^2-m)r^2/2} \mathbb{E}\left[|M_\xi(G+mt,t)|^m \mathbf{1}_{\{|G+mt| \le R_a\}}\right].$$

We next bound the tilted expectation. Align $t = re_1$, write $z = G + mt$, and decompose

$$z = z_1 e_1 + z_\perp, \qquad \zeta = R_a - z_1, \qquad \Delta_m = mr - R_a.$$

In both the moving shell and fixed physical-offset annuli, $\Delta_m = (m-1)\rho_n + O(\log\log n)$, and hence $\Delta_m \ge c_m \rho_n$ for large n. We claim that, for every fixed integer L,

$$\mathbb{E}\left[(1 + \zeta_+ + |z_\perp|)^L \mathbf{1}_{\{|z| \le R_a\}}\right] \le C_{L,m} \rho_n^{-1} e^{-\Delta_m^2/2}.$$

Indeed, write $G = We_1 + U$. If $|G + mt| \le R_a$, then

$$W \le -\Delta_m - \frac{|U|^2}{2R_a},$$

because $\sqrt{R_a^2 - |U|^2} \le R_a - |U|^2/(2R_a)$. The one-dimensional Gaussian tail-moment bound

$$\mathbb{E}\{(1 + (-x - W)_+)^L \mathbf{1}_{W \le -x}\} \le C_L x^{-1} e^{-x^2/2}, \qquad x \ge 1,$$

with $x = \Delta_m + |U|^2/(2R_a)$, gives the claim after integration in U: since Δ_m/R_a stays bounded away from zero and infinity, the factor $e^{-x^2/2}$ is at most $e^{-\Delta_m^2/2} e^{-c|U|^2}$.

In these coordinates the derivative marks have the following deterministic bounds on $\{|z| \le R_a\}$:

$$\begin{aligned}
M_q(z,t) &= \{z - (1+2\kappa_n)t\} \cdot \partial_q t = O\{(1+\zeta_+)/\rho_n + \kappa_n \rho_n\}, \\
M_h(z,t) &= \{z - (1+2\kappa_n)t\} \cdot \partial_h t = O_A(1 + \zeta_+ + \kappa_n \rho_n), \\
M_{y_j}(z,t) &= \{z - (1+2\kappa_n)t\} \cdot \partial_{y_j} t = O\{1 + |z_\perp| + (\log\log n)(1+\zeta_+)/\rho_n\}.
\end{aligned}$$

For M_q, use $R_a - r = O(\log\log n/\rho_n)$ in the moving shell; for M_h, use $|h| \le A$; for M_{y_j}, use $t \cdot \partial_{y_j} t = 0$ and the uniform chart derivative bounds. Since $\kappa_n \rho_n = O(\log\log n/\rho_n) = o(1)$, the tilted tail-moment estimate from Proposition 4.2 shows that the derivative marks cost only fixed powers of $\log\log n$. Combining this with the same algebra as in Proposition 4.2 yields, uniformly in the moving shell and in fixed physical-offset annuli,

$$\sum_i \mathbb{E}|\partial_\xi V_{i,t}^{(a)}|^m \le (\log\log n)^C \rho_n^{-1} e^{-\lambda_n} \le (\log\log n)^C \rho_n^{-(d-1)}.$$

Here we used $e^{-\lambda_n} \le C\rho_n^{-(d-2)}$, which follows from $s_n = O(1)$.

Apply Rosenthal's inequality to $\partial_\xi S_t^{(a)} = \sum_i \partial_\xi V_{i,t}^{(a)}$. The centered pth-moment term is bounded by the display with $m = p$, and the variance term is bounded by the display with $m = 2$ raised to the power $p/2$, which is smaller. It remains only to bound the mean. Rotational symmetry makes the angular means zero. For radial derivatives, write

$$\mathbb{E}S_t^{(a)} = e^{-\kappa_n r^2} P_r, \qquad P_r = \mathbb{P}\{|G + re_1| \le R_a\}.$$

The derivative P_r' is uniformly bounded, because

$$P_r' = \int_{|z|\le R_a} (z_1 - r)\phi_d(z - re_1)\, dz$$

and the absolute value of this integral is at most $\mathbb{E}|G_1|$. Hence $\partial_q \mathbb{E}S_t^{(a)} = O(e^{-\lambda_n}/\rho_n)$, since $\partial_q r = 1/r = O(\rho_n^{-1})$ and $\kappa_n = O(\log\log n/\rho_n^2)$. On a fixed physical-offset annulus, $\partial_h \mathbb{E}S_t^{(a)} = O_A(e^{-\lambda_n})$. Their pth powers are also $O(\rho_n^{-(d-1)})$, because $p > d$ and $d \ge 3$. Therefore

$$\mathbb{E}|\partial_\xi S_t^{(a)}|^p \le (\log\log n)^C \rho_n^{-(d-1)}$$

uniformly in all coordinates and all displayed parameter values.

Integrating this pointwise bound over $O(\rho_n^{d-1})$ angular charts and over q-length $O(\log\log n)$, or fixed h-length $O_A(1)$, gives the displayed integral bound. By Markov, the total gradient integral is at most $(\log\log n)^{C'}$ with probability tending to one. Fix $\gamma > 0$, choose

$$K > \frac{\gamma + C'/p}{1 - d/p},$$

and, on this event, partition every coordinate patch into d-dimensional cubes. Take the side length to be

$$\ell_n = (\log\log n)^{-K}.$$

The Morrey–Poincare inequality on each cube gives

$$O\Big(\ell_n^{1-d/p}(\log\log n)^{C'/p}\Big) = O\Big((\log\log n)^{-K(1-d/p)+C'/p}\Big),$$

as a uniform bound for the oscillation of $S_t^{(a)}$ on that cube. For the displayed choice of K, this is at most $(\log\log n)^{-\gamma}$. Taking one point from each cube gives a net of size $\rho_n^{d-1}(\log\log n)^C$ in either coordinate region. ◻

The Morrey estimate is used below only to transfer fixed-location bounds to the angular continuum. Its net has size $\rho_n^{d-1+o(1)}$, which is small enough for the Bennett exponents in Proposition 4.3.

For radii below the moving shell, use physical distance from the truncated sample edge. Put

$$R_a = R_n(a), \qquad h = R_a - |t|.$$

A bounded interval of h corresponds to a radial-energy interval of length $O(\rho_n)$, so a direct energy-coordinate net would be unnecessarily large. The fixed-location estimates retain the Gaussian penalty in h.

Lemma 4.5. *Fix $d \ge 3$, $a > 0$, and $A < \infty$, and suppose $s_n = O(1)$. Uniformly over directions and $|h| \le A$, for each fixed $p > 1$,*

$$V_{i,t}^{(a)} \le e^{-a-h^2/2+o(1)}, \qquad \sum_i \mathbb{E}\{V_{i,t}^{(a)}\}^p \le C_{p,a,A,d}\,\rho_n^{-1} e^{-\lambda_n-(p-1)a-ph^2/2+o(1)}.$$

Consequently, for every fixed $\zeta \in (0,1)$,

$$\mathbb{P}\{S_t^{(a)} \ge \zeta\} \le \Big(\rho_n^{-1}e^{-\lambda_n}\Big)^{e^{a+h^2/2}\zeta - o(1)}.$$

Proof. The completing-square and Mills-ratio calculation used in Proposition 4.2 sharpens in the physical coordinate $h = R_a - r$. Uniformly for bounded h and every fixed $p > 1$,

$$V_{i,t}^{(a)} \le \exp\{-a - h^2/2 + o(1)\}, \qquad \sum_i \mathbb{E}\{V_{i,t}^{(a)}\}^p \le C\rho_n^{-1} \exp\{-\lambda_n - (p-1)a - ph^2/2 + o(1)\}.$$

The first bound is the completing-square inequality

$$\log V_{i,t}^{(a)} \le \lambda_n - a - \kappa_n r^2 - \frac{1}{2}(R_a - r)^2 = -a - h^2/2 + o(1).$$

For the second bound, exponential tilting reduces $\sum_i \mathbb{E}\{V_{i,t}^{(a)}\}^p$ to a Gaussian probability of the ball $\{|G| \le R_a\}$ under a mean shifted by pr in the t-direction. The nearest boundary distance is

$$pr - R_a = (p-1)R_a - ph.$$

Using the Mills bound as in Proposition 4.2 gives

$$\mathbb{P}\{|G + pt| \le R_a\} \le C_{p,A,d}\rho_n^{-1} \exp\left\{-\frac{(pr - R_a)^2}{2}\right\}.$$

Substituting $r = R_a - h$, $R_a^2/2 = \log n + \lambda_n - a$, and $\kappa_n r^2 = \lambda_n + o(1)$ into the exact tilted identity gives

$$\begin{aligned} &n^{1-p} e^{-p\kappa_n r^2 + (p^2-p)r^2/2} \exp\left\{-\frac{(pr - R_a)^2}{2}\right\} \\ &\quad \le \exp\{-\lambda_n - (p-1)a - ph^2/2 + o(1)\}, \end{aligned}$$

which is the displayed moment estimate for $p > 1$. Separately,

$$\mu_t = \mathbb{E}S_t^{(a)} = e^{-\kappa_n r^2}\mathbb{P}\{|G + t| \le R_a\} \le e^{-\kappa_n r^2} = e^{-\lambda_n + o(1)} = o(1).$$

The $p = 2$ case gives

$$v_t := \sum_i \operatorname{Var}(V_{i,t}^{(a)}) \le C\rho_n^{-1} e^{-\lambda_n - a - h^2 + o(1)},$$

and the jump bound is $V_{i,t}^{(a)} \le e^{-a-h^2/2+o(1)}$. Bennett's inequality gives

$$\mathbb{P}\{S_t^{(a)} \ge \zeta\} \le \left(\rho_n^{-1} e^{-\lambda_n}\right)^{e^{a+h^2/2}\zeta - o(1)}.$$

□

For growing physical offset $h = R_a - |t|$, the fixed-location tail becomes much smaller.

Lemma 4.6. *Fix $d \ge 3$, $a > 0$, $A_0 > 0$, and $\tau \in (0,1)$. At the sharp scale $s_n = O(1)$, uniformly over directions and over*

$$h = R_a - |t| \in [A_0, \rho_n/2 + 2],$$

one has

$$\mu_t := \mathbb{E}S_t^{(a)} = o(1), \qquad \max_i V_{i,t}^{(a)} \le \beta_h := \exp\{-a - h^2/4\},$$

and

$$v_t := \sum_i \operatorname{Var}(V_{i,t}^{(a)}) \le \beta_h\, \rho_n^{-(d-2)+o(1)}.$$

Consequently,

$$\mathbb{P}\{S_t^{(a)} \ge \tau\} \le \exp\{-c_{\tau,d} e^{a+h^2/4} \log \rho_n\}$$

for all large n, with $c_{\tau,d} > 0$.

Proof. Write $R = R_a$, $r = |t| = R - h$, and use $R = \rho_n + O((\log\log n)/\rho_n)$, $\lambda_n = \kappa_n\rho_n^2 = O(\log\log n)$. Since $h \geq A_0$,

$$\kappa_n(\rho_n^2 - r^2) \leq C\lambda_n h/\rho_n \leq h^2/4$$

uniformly on the displayed range, for all large n. The pointwise completing-square bound therefore gives

$$\log V_{i,t}^{(a)} \leq \lambda_n - a - \kappa_n r^2 - h^2/2 = -a - h^2/2 + \kappa_n(\rho_n^2 - r^2) \leq -a - h^2/4.$$

This proves the jump bound.

The mean satisfies

$$\mu_t = e^{-\kappa_n r^2}\mathbb{P}\{|G + t| \leq R\} \leq e^{-\kappa_n r^2} \leq e^{-\lambda_n/10} = o(1),$$

because $r \geq R - \rho_n/2 - 2 \geq 2\rho_n/5$ for large n, and $\lambda_n = (d-2)\log\rho_n + O(1)$.

It remains to bound the variance. The exact second-moment identity is

$$\sum_i \mathbb{E}\{V_{i,t}^{(a)}\}^2 = n^{-1}e^{-2\kappa_n r^2 + r^2}\mathbb{P}\{|G + 2t| \leq R\}.$$

If $h \leq R/2 - 1$, then the ball event implies a one-dimensional left-tail event of size $R - 2h$. Hence

$$\mathbb{P}\{|G + 2t| \leq R\} \leq C\exp\{-(R - 2h)^2/2\},$$

and the algebra

$$-\log n + r^2 - (R - 2h)^2/2 = \lambda_n - a - h^2$$

gives

$$\sum_i \mathbb{E}\{V_{i,t}^{(a)}\}^2 \leq C\exp\{-a - h^2 - \lambda_n + 2\kappa_n(\rho_n^2 - r^2)\} \leq Ce^{-a-h^2/2}\rho_n^{-(d-2)+o(1)}.$$

This is bounded by $\beta_h\rho_n^{-(d-2)+o(1)}$. If $h > R/2 - 1$, then $\mathbb{P}\{|G + 2t| \leq R\} \leq 1$ and

$$n^{-1}e^{-2\kappa_n r^2 + r^2} \leq \exp\{-(1/2 + o(1))\log n\},$$

because $r \leq R/2 + 1$. Since $h \leq \rho_n/2 + 2$, this is again $\leq \beta_h\rho_n^{-(d-2)+o(1)}$: here the left side is $\exp\{-(1/2 + o(1))\log n\}$, whereas β_h^{-1} is at most $\exp\{(1/8 + o(1))\log n\}$. The variance bound follows.

For large n, $\mu_t \leq \tau/2$. Bennett's inequality for centered summands bounded by β_h and total variance v_t yields

$$\mathbb{P}\{S_t^{(a)} \geq \tau\} \leq \exp\left[-\frac{v_t}{\beta_h^2}H\left(\frac{\beta_h(\tau - \mu_t)}{v_t}\right)\right], \qquad H(u) = (1 + u)\log(1 + u) - u.$$

Since $\beta_h(\tau - \mu_t)/v_t \geq c_\tau\rho_n^{d-2-o(1)}$, the exponent is at least $c_{\tau,d}\beta_h^{-1}\log\rho_n$. This is the claimed tail bound. □

The exponential decay in h^2 permits a union bound over a polynomial-size Euclidean net. The next lemmas handle the remaining bulk $|t| \leq \rho_n/2$. Very close to the origin, negative sample curvature rules out a KKT crossing without any edge analysis. The next lemma uses that curvature to prove the small-radius part of Proposition 4.10.

Lemma 4.7. *Fix $d \geq 3$. At the sharp scale $s_n = O(1)$,*

$$\mathbb{P}\left\{\sup_{0<|t|\leq c_0\kappa_n} D_n(t) > 1\right\} \to 0$$

for a sufficiently small constant $c_0 > 0$.

Proof. Write

$$D_n(t) = \frac{1}{n}\sum_{i=1}^{n} \exp\{t \cdot \widetilde{Z}_i - (1/2 + \kappa_n)|t|^2\}, \qquad \widetilde{Z}_i = Z_i - \bar{Z}_n.$$

On an event whose probability tends to one,

$$\left\| \frac{1}{n}\sum_i \widetilde{Z}_i \widetilde{Z}_i^\top - I_d \right\|_{\mathrm{op}} \le \kappa_n/4, \qquad \max_i |\widetilde{Z}_i| \le 3\rho_n, \qquad \frac{1}{n}\sum_i |\widetilde{Z}_i|^3 \le C.$$

At $t = 0$, $D_n(0) = 1$, $\nabla D_n(0) = 0$, and

$$\nabla^2 D_n(0) = \frac{1}{n}\sum_i \widetilde{Z}_i \widetilde{Z}_i^\top - (1 + 2\kappa_n) I_d \preceq -\frac{7}{4}\kappa_n I_d.$$

For $|t| \le c_0\kappa_n$, uniformly over unit vectors v,

$$\begin{aligned} &\left| v^\top \{\nabla^2 D_n(t) - \nabla^2 D_n(0)\} v \right| \\ &\qquad \le C|t| \frac{1}{n}\sum_i (1 + |\widetilde{Z}_i|^3) + C|t|^2 \le C c_0 \kappa_n. \end{aligned}$$

The first inequality follows from the explicit Hessian formula and $\max_i |t \cdot \widetilde{Z}_i| = o(1)$. Choosing c_0 small enough makes the Hessian at most $-\kappa_n I_d$ throughout the ball. Taylor's formula then gives $D_n(t) \le 1 - c\kappa_n |t|^2 < 1$ for $0 < |t| \le c_0\kappa_n$. □

Empirical centering is negligible relative to the deterministic bulk damping.

Lemma 4.8. *Fix $d \ge 3$. At the sharp scale $s_n = O(1)$, for every fixed $c_0 > 0$,*

$$\sup_{c_0\kappa_n \le |t| \le \rho_n/2} \frac{|t \cdot \bar{Z}_n|}{1 - \exp\{-\kappa_n |t|^2\}} \xrightarrow{\mathbb{P}} 0.$$

Proof. Since the supremum over directions is $|t|\,|\bar{Z}_n|$, it remains to bound

$$\sup_{c_0\kappa_n \le r \le \rho_n/2} \frac{r|\bar{Z}_n|}{1 - e^{-\kappa_n r^2}}.$$

For $r \le \kappa_n^{-1/2}$, the denominator is at least $c\kappa_n r^2$, and this part is bounded by $C|\bar{Z}_n|\kappa_n^{-2}$. For $r \ge \kappa_n^{-1/2}$, the denominator is bounded below by a positive constant, and this part is bounded by $C\rho_n|\bar{Z}_n|$. At the sharp scale $\kappa_n \asymp (\log\log n)/\log n$, while $|\bar{Z}_n| = O_{\mathbb{P}}(n^{-1/2})$. Hence both bounds are $o_{\mathbb{P}}(1)$. □

The full empirical moment-generating function is likewise negligible relative to the bulk damping.

Lemma 4.9. *Fix $d \ge 3$. At the sharp scale $s_n = O(1)$, for every fixed $c_0 > 0$,*

$$\sup_{c_0\kappa_n \le |t| \le \rho_n/2} \frac{|H_n(t) - 1| + |t \cdot \bar{Z}_n|}{1 - \exp\{-\kappa_n |t|^2\}} \xrightarrow{\mathbb{P}} 0.$$

Proof. Lemma 4.8 proves the empirical-centering part, so it remains to prove the same display with $|H_n(t)-1|$ in the numerator. Write

$$m(r) = 1 - \exp\{-\kappa_n r^2\}.$$

We first remove the tiny radii. Put $t = ru$, $u \in S^{d-1}$, and define, continuously at $r = 0$,

$$\Psi_{r,u}(z) = \frac{\exp\{ru \cdot z - r^2/2\} - 1 - ru \cdot z}{r^2}.$$

Then $\mathbb{E}\Psi_{r,u}(Z) = 0$. On the compact parameter set $[0,1] \times S^{d-1}$, the first derivatives of $\Psi_{r,u}$ are bounded by $A(z) = C_d e^{|z|}(1+|z|^3)$, which has finite moments of all orders. Take an $n^{-1/4}$-net, whose size is $O(n^{d/4})$. Rosenthal's inequality gives an $O(n^{-q/2})$ q-moment bound for any fixed $q > d$ large enough, hence an $O(n^{-q/4})$ tail at threshold $n^{-1/4}$. The union bound is $O(n^{(d-q)/4}) = o(1)$. The bound $n^{-1}\sum_i A(Z_i) = O_{\mathbb{P}}(1)$ then extends the estimate from the net to the continuum and gives

$$\sup_{0<|t|\le 1} \frac{|H_n(t) - 1 - t \cdot \bar{Z}_n|}{|t|^2} = O_{\mathbb{P}}(n^{-1/4}).$$

Since $m(r) \ge c\kappa_n r^2$ for $r \le 1$, uniformly on $c_0\kappa_n \le |t| \le 1$,

$$\frac{|H_n(t) - 1|}{m(|t|)} \le C\frac{|\bar{Z}_n|}{\kappa_n |t|} + C\frac{n^{-1/4}}{\kappa_n} = o_{\mathbb{P}}(1).$$

Here $|t| \ge c_0\kappa_n$, $|\bar{Z}_n| = O_{\mathbb{P}}(n^{-1/2})$, and $\kappa_n \asymp (\log\log n)/\log n$.

It remains to control $1 \le |t| \le \rho_n/2$. Let

$$1 = x_0 < x_1 < \cdots < x_J = b_{n,d}/2, \qquad x_{j+1} \le 2x_j,$$

be the dyadic partition in squared radius, and set

$$\mathcal{A}_j = \{t : x_j \le |t|^2 \le x_{j+1}\}, \quad R_j = x_{j+1}^{1/2}, \quad m_j = 1 - e^{-\kappa_n x_j}.$$

Since $m(|t|) \ge m_j$ on $\mathcal{A}_j$, it is enough to show that

$$\max_j \frac{1}{m_j} \sup_{t \in \mathcal{A}_j} |H_n(t) - 1| \xrightarrow{\mathbb{P}} 0.$$

Fix $\eta > 0$. Put

$$\tau_n^2 = 20\log(1/\kappa_n) + 20(d+3)\log\rho_n, \qquad B_j = \exp\{R_j^2/2 + R_j\tau_n\},$$

and write $f_t(z) = \exp\{t \cdot z - |t|^2/2\}$, $f_{t,j}^B = f_t \wedge B_j$.

The discarded lognormal tail is uniformly negligible. For $t \in \mathcal{A}_j$,

$$\mathbb{E} f_t(Z)\mathbf{1}_{\{f_t(Z) > B_j\}} \le \mathbb{P}\{G > \tau_n\} = o(m_j), \qquad G \sim \mathcal{N}(0,1).$$

For the empirical tail, use

$$\sup_{|t|\le R} f_t(z) = \begin{cases} e^{|z|^2/2}, & |z| \le R, \\ e^{R|z| - R^2/2}, & |z| > R. \end{cases}$$

Because $B_j > e^{R_j^2/2}$, the event $\sup_{|t|\le R_j} f_t(Z) > B_j$ implies $|Z| > R_j + \tau_n$. Therefore

$$\mathbb{E} \sup_{|t|\le R_j} f_t(Z)\mathbf{1}_{\{f_t(Z)>B_j\}} \le C_d(R_j + \tau_n)^{d-1} e^{-\tau_n^2/2} = o(m_j/J).$$

Markov's inequality and a union bound over j give

$$\max_j \frac{1}{m_j} \sup_{t\in\mathcal{A}_j} \frac{1}{n}\sum_{i=1}^n f_t(Z_i)\mathbf{1}_{\{f_t(Z_i)>B_j\}} \xrightarrow{\mathbb{P}} 0.$$

It remains to control the clipped process. On the event $\max_i |Z_i| \le 3\rho_n$, whose probability tends to one,

$$t \mapsto \frac{1}{n}\sum_i f^B_{t,j}(Z_i)$$

is $C\rho_n B_j$-Lipschitz on $\mathcal{A}_j$, and its expectation is no less regular: a.e. in t,

$$|\nabla_t \mathbb{E} f^B_{t,j}(Z)| \le \mathbb{E}\{|Z - t| f_t(Z)\} \le C_d.$$

Take a Euclidean net of $\mathcal{A}_j$ with mesh

$$\delta_j = \frac{\eta m_j}{16 C\rho_n B_j}.$$

Its cardinality satisfies

$$\log N_j \le C_d\{\log(\rho_n/\eta m_j) + \log B_j\} \le C_d \log n.$$

At a fixed net point, Bernstein's inequality gives

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_i\{f^B_{t,j}(Z_i) - \mathbb{E}f^B_{t,j}(Z)\}\right| > \eta m_j/4\right\} \le 2\exp\left[-c\frac{nm_j^2}{e^{R_j^2} + B_j m_j}\right].$$

Here

$$\mathrm{Var}\{f^B_{t,j}(Z)\} \le \mathbb{E}f_t(Z)^2 = e^{|t|^2} \le e^{R_j^2}, \qquad |f^B_{t,j}| \le B_j,$$

which gives the displayed Bernstein denominator. Here $R_j^2 \le b_{n,d}/2$, $B_j \le n^{1/4+o(1)}$, and $m_j \ge c\kappa_n$. Hence the exponent is at least

$$n^{1/2-o(1)}\kappa_n^2,$$

which dominates $\log N_j + \log J = O(\log n)$. A union bound over all net points and all j, followed by the Lipschitz extension from the net, proves

$$\max_j \frac{1}{m_j}\sup_{t\in\mathcal{A}_j}\left|\frac{1}{n}\sum_i\{f^B_{t,j}(Z_i) - \mathbb{E}f^B_{t,j}(Z)\}\right| \xrightarrow{\mathbb{P}} 0.$$

Combining the clipped concentration, the truncation bias, and the empirical tail bound proves the dyadic estimate and the lemma. □

The two bulk estimates exclude all crossings with $|t| \le \rho_n/2$.

Proposition 4.10. *Fix $d \ge 3$. At the sharp scale $s_n = O(1)$,*

$$\mathbb{P}\left\{\sup_{|t|\le\rho_n/2} D_n(t) > 1\right\} \to 0.$$

Proof. Lemma 4.7 handles $0 < |t| \le c_0\kappa_n$. On the event in Lemma 4.9, put $g(t) = 1 - \exp\{-\kappa_n|t|^2\}$. Uniformly on $c_0\kappa_n \le |t| \le \rho_n/2$, for a deterministic $\delta_n \downarrow 0$,

$$|H_n(t) - 1| + |t \cdot \bar{Z}_n| \le \delta_n g(t).$$

Therefore

$$D_n(t) = e^{-t\cdot\bar{Z}_n} e^{-\kappa_n|t|^2} H_n(t) \le e^{\delta_n g(t)}\{1 - g(t)\}\{1 + \delta_n g(t)\}.$$

The right side is at most $1 - cg(t)$ for all large n, uniformly in this range. Hence $D_n(t) < 1$ there with probability tending to one, and the microscopic lemma completes the proof. $\square$

4.5 Sparse edges in growing dimension

This subsection extends the spatial exclusion required by Theorem 1.1 from fixed to growing dimension below the simplex regime. We split the sample into a sparse upper radial cloud and a truncated lower cloud. Angular separation controls the upper cloud, while dimension-uniform concentration controls the lower cloud.

Fix a sequence $A_n \to \infty$ and define the upper radial cloud

$$\mathcal{I}_n(A_n) = \{i : \xi_{n,i} \ge -A_n\}. \tag{4.15}$$

The expected size of this set is $e^{A_n + o(A_n)}$ whenever $A_n = o(L_n)$. Write $B_n = B_n^{\mathrm{hi}} + B_n^{\mathrm{lo}}$ for the two sums in (2.3), split according to $\mathcal{I}_n(A_n)$.

4.5.1 Separation of the upper radial cloud

Conditional on their radii, the uncentered Gaussian directions are independent and uniform on S^{d_n-1}. If $\#\mathcal{I}_n(A_n) = \exp\{O_{\mathbb{P}}(A_n)\}$ and $A_n = o(d_n)$, spherical concentration gives

$$\max_{i\ne j\in\mathcal{I}_n(A_n)} \left|\frac{q_i}{|q_i|} \cdot \frac{q_j}{|q_j|}\right| = o_{\mathbb{P}}(1). \tag{4.16}$$

Since the upper centers have radius $\sqrt{2L_n}\{1 + o_{\mathbb{P}}(1)\}$, (4.16) implies

$$\min_{i\ne j\in\mathcal{I}_n(A_n)} |q_i - q_j|^2 = 4L_n\{1 + o_{\mathbb{P}}(1)\}.$$

Thus no location w can lie within $o(\sqrt{L_n})$ of two upper centers. The contribution of all nonnearest upper bumps is exponentially small in L_n.

Centering by $\bar{Z}_n$ perturbs every direction by $O_{\mathbb{P}}(\sqrt{d_n/n})$ in Euclidean norm. Under (2.17) and the rate ranges considered below, this does not affect (4.16).

4.5.2 The lower cloud and the inner field

At the critical normalization, every low bump has height at most

$$\max_{i\notin\mathcal{I}_n(A_n)} h_i \le \exp\{-\delta_n(A_n + s_n)\}. \tag{4.17}$$

Before empirical centering, let

$$B_n^\circ(w) = \frac{1}{n}\sum_{i=1}^{n} \exp\left\{w \cdot \frac{Z_i}{\sqrt{\gamma_n}} - \frac{|w|^2}{2}\right\}.$$

For a fixed w, its population mean is

$$\mathbb{E}B_n^\circ(w) = \exp\left\{-\frac{\varepsilon_n|w|^2}{2}\right\}. \tag{4.18}$$

Thus the deterministic background is already negligible near $|w| = \sqrt{2L_n}$ whenever $d_n \to \infty$, because then $\varepsilon_n L_n \to \infty$ at the critical scale.

The proof below establishes the following low-cloud estimate. Its precision $o_{\mathbb{P}}(\delta_n)$ is essential in superlogarithmic dimension: for a deterministic radius $r_n = o(\sqrt{L_n})$,

$$\sup_{|w|\geq r_n} B_n^{\mathrm{lo}}(w) = o_{\mathbb{P}}(\delta_n), \tag{4.19}$$

provided A_n is chosen so that

$$A_n = o(L_n \wedge d_n), \qquad \delta_n A_n - \log(d_n L_n) \longrightarrow \infty. \tag{4.20}$$

For fixed w, (4.17), (4.18), and Bernstein's inequality give much more than (4.19). The work is to transfer that bound over w. A direct Euclidean net of the relevant ball has logarithmic size $O(d_n \log L_n)$; the second condition in (4.20) is designed to make the bounded-jump exponent dominate this entropy. Radial peeling avoids an unnecessarily fine net near the origin.

The inner region contains the deterministic equality $B_n(0) = 1$, so it cannot satisfy (4.19). The corresponding target is strict concavity away from zero:

$$\mathbb{P}\left\{\sup_{0<|w|\leq r_n} B_n(w) < 1\right\} \longrightarrow 1. \tag{4.21}$$

At the origin,

$$\nabla B_n(0) = 0, \qquad \nabla^2 B_n(0) = \frac{1}{n}\sum_i \frac{\widetilde{Z}_i\widetilde{Z}_i^\top}{\gamma_n} - I_{d_n}.$$

Its expectation is $-\varepsilon_n I_{d_n} + o(1)$. Standard covariance concentration therefore gives a negative Hessian whenever $\sqrt{d_n/n} = o(\varepsilon_n)$. The remaining task is to propagate this local curvature through the interval up to r_n using a clipped exponential process. This is easier than the outer-edge estimate because the population gap in (4.18) grows monotonically with $|w|$.

Equations (4.19) and (4.21) give the field exclusion used later in Proposition 6.18. On the event $\max_i \xi_{n,i} \leq s_n - \eta$, every upper bump has height at most

$$e^{-\delta_n\eta},$$

and hence lies below one by at least $c_\eta\delta_n$ whenever $0 < \delta_n \leq 1$. At the center of any upper bump, the separation estimate makes all other upper bumps $o_{\mathbb{P}}(\delta_n)$, and (4.19) makes the low-cloud contribution $o_{\mathbb{P}}(\delta_n)$. Hence the outer field is below one, while (4.21) handles the inner field.

4.5.3 The direct truncation range

The truncation and angular-separation conditions can be imposed simultaneously in the range

$$\log\log n \ll d_n, \qquad d_n\{\log d_n\}^2 = o\{(\log n)^3\}, \qquad d_n \log n = o(n). \tag{4.22}$$

Indeed, in the worst part of this range, $\delta_n \asymp \sqrt{L_n/d_n}$. Conditions (4.20) admit a choice of A_n precisely when

$$\delta_n^{-1} \log(d_n L_n) = o(L_n \wedge d_n),$$

which is implied by (4.22). The lower bound $d_n \gg \log\log n$ ensures $A_n = o(d_n)$, so the upper radial points are uniformly separated. Smaller dimensions are handled below by constant-mesh estimates. At the upper end, where δ_n is too small for (4.20), Section 5.1 replaces spatial truncation by the entropy dual.

4.5.4 Slowly growing dimension

We first treat growing dimensions below $(\log L_n)^2$. A constant-mesh argument is enough; its key input is an exact semiconvexity property of every Gaussian-bump field. Fixed dimensions use the moving-shell estimates in Section 4.3.

Lemma 4.11. *Let*

$$F(w) = \sum_{j=1}^{m} c_j \exp\left\{w \cdot q_j - \frac{|w|^2}{2}\right\}, \qquad c_j > 0.$$

Then

$$\nabla^2 \log F(w) = \mathrm{Cov}_{p_w}(q_j) - I \succeq -I, \tag{4.23}$$

where p_w is proportional to $c_j e^{w\cdot q_j - |w|^2/2}$. Every global maximizer w_* belongs to the convex hull of $q_1, \ldots, q_m$, and

$$F(w) \geq F(w_*) e^{-|w-w_*|^2/2} \qquad (w \in \mathbb{R}^d). \tag{4.24}$$

Proof. Differentiating $\log F$ gives (4.23). At a maximizer, $0 = \nabla \log F(w_*) = \sum_j p_{w_*}(j) q_j - w_*$, which proves the convex-hull assertion. Integrating the lower Hessian bound along the segment from w_* to w proves (4.24). □

Fix $\alpha > 0$, put

$$Q_i = \frac{Z_i}{\sqrt{\gamma_n}}, \qquad R_{n,\alpha} = \sqrt{2(L_n - \alpha)},$$

and define the independent truncated field

$$S_{n,\alpha}(w) = \frac{1}{n} \sum_{i=1}^{n} e^{w\cdot Q_i - |w|^2/2} \mathbf{1}_{\{|Q_i| \leq R_{n,\alpha}\}}.$$

The next lemma gives the only outer-field estimate needed below. It uses only bounded summands, their population mean, and a constant Euclidean net.

Lemma 4.12. *Suppose*

$$d_n \longrightarrow \infty, \qquad d_n \leq (\log L_n)^2, \qquad s_n = O(1).$$

For every fixed $\alpha > 0$, choose a fixed $\zeta > 0$ with $\zeta^2/2 < \alpha$, let $\mathcal{G}_n$ be a ζ-net of $B(0, R_{n,\alpha})$, and fix

$$0 < \rho < e^{-\zeta^2/2} - e^{-\alpha}.$$

Then

$$\mathbb{P}\left\{ \max_{\substack{w \in \mathcal{G}_n \\ |w| \geq R_{n,\alpha}/2 - \zeta}} S_{n,\alpha}(w) \geq e^{-\zeta^2/2} - \rho \right\} \longrightarrow 0. \tag{4.25}$$

Proof. For fixed w, every summand of $S_{n,\alpha}(w)$ is at most

$$M_n(w) = \exp\left\{-\alpha - \frac{(R_{n,\alpha} - |w|)^2}{2}\right\}, \tag{4.26}$$

while

$$\mathbb{E}S_{n,\alpha}(w) \le e^{-\varepsilon_n |w|^2/2}. \tag{4.27}$$

If independent nonnegative summands have sum of means μ, are bounded by M, and $t > \mu$, the exponential-Chernoff bound is

$$\mathbb{P}\left\{\sum_j X_j \ge t\right\} \le \exp\left[-\frac{1}{M}\left\{t \log\frac{t}{\mu} - t + \mu\right\}\right]. \tag{4.28}$$

Apply this with $t = e^{-\zeta^2/2} - \rho$. Uniformly for $R_{n,\alpha}/2 - \zeta \le |w| \le R_{n,\alpha}$,

$$-\log \mathbb{P}\{S_{n,\alpha}(w) \ge t\} \ge (1 - o(1)) e^{\alpha} t\, \varepsilon_n L_n.$$

Indeed, the infimum of $e^{(R_{n,\alpha}-r)^2/2} r^2$ over this interval is $(1 - o(1)) R_{n,\alpha}^2$.

In the present range, $k_n = d_n/2 \to \infty$ and

$$\varepsilon_n L_n = (1 + o(1)) k_n \log\frac{L_n}{k_n} = (1 - o(1)) k_n \log L_n. \tag{4.29}$$

On the other hand,

$$\log|\mathcal{G}_n| \le d_n \log\frac{C R_{n,\alpha}}{\zeta} = k_n \log L_n + O(d_n).$$

Since $e^{\alpha} t > 1$, the union bound proves (4.25). □

The population gap in (4.18) controls the entire inner half-ball once the microscopic curvature estimate is combined with a relative empirical-process bound.

Lemma 4.13. *Suppose* $3 \le d_n \le (\log L_n)^2$ *and* $s_n = O(1)$. *Then*

$$\mathbb{P}\left\{\sup_{0<|w|\le R_{n,\alpha}/2} B_n(w) > 1\right\} \longrightarrow 0. \tag{4.30}$$

Proof. At the origin, $B_n(0) = 1$, $\nabla B_n(0) = 0$, and Gaussian sample-covariance concentration gives

$$\nabla^2 B_n(0) = \frac{1}{n}\sum_i \frac{\widetilde{Z}_i \widetilde{Z}_i^\top}{\gamma_n} - I \preceq -\frac{\varepsilon_n}{2} I$$

with probability tending to one. Since $\max_i |\widetilde{Z}_i| = O_{\mathbb{P}}(\sqrt{L_n})$, Taylor's theorem makes the inequality strict on $0 < |w| \le r_{0,n} := \varepsilon_n/(16 L_n^{3/2})$.

For completeness, we give the empirical-process estimate used beyond this microscopic ball. Before centering, put

$$H_n(w) = \frac{1}{n}\sum_i \exp\left\{w \cdot Q_i - \frac{|w|^2}{2\gamma_n}\right\}, \qquad \mathbb{E}H_n(w) = 1,$$

and $g_n(r) = 1 - e^{-\varepsilon_n r^2/2}$. On $r_{0,n} \le |w| \le R_{n,\alpha}/2$,

$$\sup_w \frac{|H_n(w) - 1|}{g_n(|w|)} \longrightarrow_{\mathbb{P}} 0. \tag{4.31}$$

To verify this, split the range into squared-radius annuli $\mathcal{A}_j = \{r_j \le |w| \le R_j\}$ with $R_j^2 \le 2r_j^2$, and put $g_j = g_n(r_j)$. There are $O(\log L_n)$ annuli. On an annulus with outer radius $R = R_j$, clip the lognormal summands at

$$B_R = \exp\{R^2/(2\gamma_n) + RT_n\}, \qquad T_n^2 = C\{d_n \log L_n + \log g_n(r_{0,n})^{-1}\}.$$

We first control both discarded tails. For a fixed w, exponential tilting gives

$$\sup_{|w|\le R} \mathbb{E}\Big[e^{w\cdot Q-|w|^2/(2\gamma_n)}\mathbf{1}_{\{e^{w\cdot Q-|w|^2/(2\gamma_n)}>B_R\}}\Big] \le e^{-cT_n^2}.$$

Indeed, tilt Q from $\mathcal{N}(0,\gamma_n^{-1}I)$ to $\mathcal{N}(w/\gamma_n,\gamma_n^{-1}I)$; the threshold is then at least $\sqrt{\gamma_n}T_n$ standard deviations in the w-direction. For the empirical envelope, use

$$\sup_{|w|\le R} e^{w\cdot q-|w|^2/(2\gamma_n)} = \begin{cases} e^{\gamma_n|q|^2/2}, & \gamma_n|q|\le R,\\ e^{R|q|-R^2/(2\gamma_n)}, & \gamma_n|q|>R.\end{cases}$$

Thus exceeding B_R forces $|q| > R/\gamma_n + T_n$. Completing the square against the $\mathcal{N}(0,\gamma_n^{-1}I)$ density gives

$$\mathbb{E}\left[\sup_{|w|\le R} e^{w\cdot Q-|w|^2/(2\gamma_n)}\mathbf{1}_{\{\sup_{|v|\le R} e^{v\cdot Q-|v|^2/(2\gamma_n)}>B_R\}}\right] \le Ce^{Cd_n\log L_n}\int_{T_n}^{\infty} e^{-cu^2}\,du.$$

Taking the constant in T_n^2 sufficiently large makes this $o(g_j/\log L_n)$. Markov's inequality and a union bound over the annuli control the empirical discarded tails at the same order.

It remains to control the clipped fields. Fix $\rho>0$. On the event $\max_i|Q_i|\le C\sqrt{L_n}$, each is $C\sqrt{L_n}B_R$-Lipschitz. Use a net on $\mathcal{A}_j$ of mesh

$$\frac{\rho g_j}{C\sqrt{L_n}B_R}.$$

Its logarithmic cardinality is $O(d_nL_n)$. Because $R^2\le L_n/2+O(1)$, one has $B_R\le n^{1/4+o(1)}$, while Bernstein's inequality at deviation $\rho g_j/4$ has exponent at least

$$\frac{cng_n(r)^2}{e^{R^2/\gamma_n}+B_Rg_n(r)} \ge n^{1/2-o(1)}g_n(r_{0,n})^2.$$

This dominates d_nL_n. The union bound over all net points and all annuli shows that the probability that the left side of (4.31) exceeds ρ tends to zero. Since $\rho>0$ was arbitrary, (4.31) follows.

Now

$$B_n(w) = e^{-w\cdot\bar{Z}_n/\sqrt{\gamma_n}}e^{-\varepsilon_n|w|^2/2}H_n(w).$$

Moreover,

$$\sup_{r_{0,n}\le r\le R_{n,\alpha}/2}\frac{r|\bar{Z}_n|/\sqrt{\gamma_n}}{g_n(r)} \le C\frac{|\bar{Z}_n|}{\varepsilon_n r_{0,n}} = o_{\mathbb{P}}(1),$$

because $d_n\le(\log L_n)^2$, $|\bar{Z}_n| = O_{\mathbb{P}}(\sqrt{d_n/n})$, and $\varepsilon_n\ge c\log L_n/L_n$. Thus (4.31) preserves a fixed fraction of the population gap $g_n(|w|)$, proving (4.30). □

4.5.5 Moderately growing dimension

We now prove (4.19) and (4.21) under (4.22). The argument deliberately uses a deeper truncation than (4.20): this makes every remaining low-cloud summand so small that elementary Bernstein bounds survive a Euclidean net.

Put

$$\mathfrak{h}_n = \log \frac{d_n L_n}{\delta_n \varepsilon_n}.$$

The expansions in Section 4.2 and (4.22) imply

$$\mathfrak{h}_n \asymp \log L_n, \qquad \frac{\varepsilon_n L_n}{\mathfrak{h}_n} \longrightarrow \infty, \qquad \log \frac{L_n^3}{\varepsilon_n^3} = O(\log L_n). \tag{4.32}$$

For $d_n = o(L_n)$, these statements follow from $\delta_n = 1 + o(1)$ and $\varepsilon_n L_n \sim (d_n/2)\log(2L_n/d_n)$; in particular,

$$\mathfrak{h}_n = 2\log L_n - \log\log(2L_n/d_n) + O(1).$$

When $d_n \asymp L_n$, both ε_n and δ_n stay bounded away from zero and infinity. When $d_n/L_n \to \infty$, $\varepsilon_n \to 1$ and $\delta_n \asymp \sqrt{L_n/d_n}$. The last two conditions in (4.22) then give all three relations in (4.32). They also permit deterministic sequences $\omega_n \to \infty$ and

$$A_n = \frac{\omega_n \mathfrak{h}_n}{\delta_n} = o(L_n \wedge d_n). \tag{4.33}$$

Indeed, this is immediate when $d_n = O(L_n)$; when $d_n/L_n \to \infty$, it is equivalent, up to logarithmic factors, to $d_n(\log d_n)^2 = o(L_n^3)$. Finally set

$$r_n^2 = \frac{40\mathfrak{h}_n}{\varepsilon_n} = o(L_n). \tag{4.34}$$

We shall also use a uniform count for the upper radial cloud. It is a moderate-deviation extension of Lemma 2.7.

Lemma 4.14. *Under (4.22), if $A_n \to \infty$ and $A_n = o(L_n \wedge d_n)$, then*

$$n\mathbb{P}\{Y_n \geq b_n - a_n A_n\} = \exp\{(1 + o(1))A_n\}.$$

Consequently, the number of indices with $(|Z_i|^2/2 - b_n)/a_n \geq -A_n$ is at most e^{2A_n} with probability tending to one.

Proof. On $[b_n - a_n A_n, b_n]$, the reciprocal Gamma hazard equals $a_n\{1 + o(1)\}$ uniformly. Indeed, (2.10) gives this once $a_n A_n = o(b_n - k_n + 1)$. For $k_n = O(L_n)$, that relation follows from $A_n = o(d_n \wedge L_n)$; for $k_n/L_n \to \infty$, it follows from $a_n \asymp \sqrt{k_n/L_n}$ and $A_n = o(L_n)$. Integrating the hazard gives the displayed tail ratio. The count is binomial with mean $e^{(1+o(1))A_n}$, so Markov's inequality proves the second claim. □

The next proposition proves the low-cloud estimate with more precision than is needed for the Gumbel law. The stronger factor L_n^{-10} leaves enough slack to differentiate the field up to three times in the support-count argument.

Proposition 4.15. *Suppose (4.22) holds and $s_n = O(1)$. Choose A_n, r_n as in (4.33)–(4.34). Then*

$$\sup_{|w| \geq r_n} B_n^{\mathrm{lo}}(w) = O_{\mathbb{P}}(\delta_n L_n^{-10}). \tag{4.35}$$

The same conclusion holds after differentiating the bump field up to three times, with the right side replaced by $o_{\mathbb{P}}(\delta_n L_n^{-7})$.

Proof. We first work before empirical centering. Put $\xi^\circ_{n,i} = (|Z_i|^2/2 - b_n)/a_n$ and define

$$S^{\mathrm{lo}}_n(w) = \frac{1}{n}\sum_{i=1}^n \exp\left\{w\cdot\frac{Z_i}{\sqrt{\gamma_n}} - \frac{|w|^2}{2}\right\}\mathbf{1}_{\{\xi^\circ_{n,i}<-A_n+1\}}.$$

Completing the square and using $\gamma_n L_n = b_n + a_n s_n$ show that every summand is bounded by

$$\mathfrak{j}_n = \exp\{-\delta_n(A_n - 1 + s_n)\} = \exp\{-(1+o(1))\omega_n \mathfrak{h}_n\}. \tag{4.36}$$

For fixed w,

$$\mathbb{E}S^{\mathrm{lo}}_n(w) \le \exp\{-\varepsilon_n |w|^2/2\}. \tag{4.37}$$

Since a nonnegative summand bounded by $\mathfrak{j}_n$ has variance at most $\mathfrak{j}_n$ times its mean, Bernstein's inequality gives

$$\mathbb{P}\{S^{\mathrm{lo}}_n(w) - \mathbb{E}S^{\mathrm{lo}}_n(w) > t\} \le \exp\left\{-c\frac{t^2}{\mathfrak{j}_n(\mathbb{E}S^{\mathrm{lo}}_n(w) + t)}\right\}. \tag{4.38}$$

With probability tending to one, all score centers have norm at most $C\sqrt{L_n}$. Every local maximum of the low bump field lies in their convex hull, so it is enough to consider $|w| \le C\sqrt{L_n}$. On this ball the field is multiplicatively $C\sqrt{L_n}$-Lipschitz:

$$S^{\mathrm{lo}}_n(w') \ge e^{-C\sqrt{L_n}|w-w'|}S^{\mathrm{lo}}_n(w). \tag{4.39}$$

Take a net of mesh $c/\sqrt{L_n}$; its logarithmic cardinality is at most $Cd_n \log L_n$. For $|w| \ge r_n$, (4.37) is at most $e^{-20\mathfrak{h}_n}$. Apply (4.38) with $t = \delta_n L_n^{-10}$. Equations (4.32) and (4.36) give

$$\frac{t}{\mathfrak{j}_n} \gg d_n \log L_n, \qquad \frac{t^2}{\mathfrak{j}_n e^{-20\mathfrak{h}_n}} \gg d_n \log L_n.$$

The union bound and (4.39) prove the uncentered version of (4.35).

The proof of Lemma 2.10 gives

$$\max_i |\xi_{n,i} - \xi^\circ_{n,i}| = o_{\mathbb{P}}(1).$$

Moreover,

$$B_n(w) = \exp\left\{-w\cdot\frac{\bar{Z}_n}{\sqrt{\gamma_n}}\right\}B^\circ_n(w), \qquad \sup_{|w|\le C\sqrt{L_n}}\left|w\cdot\frac{\bar{Z}_n}{\sqrt{\gamma_n}}\right| = O_{\mathbb{P}}(L_n/\sqrt{n}) = o_{\mathbb{P}}(\delta_n).$$

Replacing $A_n - 1$ by A_n therefore proves the centered statement. For a multi-index β of order $j \le 3$, differentiation gives

$$\left|\partial^\beta_w e^{-|w-q_i|^2/2}\right| \le C_j(1 + |w-q_i|^j)e^{-|w-q_i|^2/2}.$$

On a fixed enlargement of the convex hull, both $|w|$ and $\max_i |q_i|$ are $O_{\mathbb{P}}(\sqrt{L_n})$. Repeating the net-and-Bernstein argument used for the order-zero field with this polynomial envelope costs at most $L_n^{j/2}$; the base estimate $O_{\mathbb{P}}(\delta_n L_n^{-10})$ is therefore $o_{\mathbb{P}}(\delta_n L_n^{-7})$ for every $j \le 3$. Outside that enlargement, the polynomially weighted Gaussian envelope is radially decreasing toward the convex hull once the enlargement radius exceeds $\sqrt{3}$. Its supremum is consequently attained on the controlled boundary, proving the derivative estimate globally. □

The inner field requires a different argument because it equals one at the origin. Sample-covariance curvature handles a microscopic ball, and the bounded-jump estimate (4.38) handles the remaining annuli.

Proposition 4.16. *Under the assumptions of Proposition 4.15,*

$$\mathbb{P}\left\{\sup_{0<|w|\le r_n} B_n(w) < 1\right\} \longrightarrow 1. \tag{4.40}$$

Proof. At the origin,

$$\nabla B_n(0) = 0, \qquad \nabla^2 B_n(0) = \frac{1}{n}\sum_i q_i q_i^\top - I.$$

The Gaussian sample-covariance bound and (4.22) imply

$$\left\|\frac{1}{n}\sum_i q_i q_i^\top - \frac{1}{\gamma_n} I\right\|_{\mathrm{op}} = o_{\mathbb{P}}(\varepsilon_n), \qquad \nabla^2 B_n(0) \preceq -\frac{\varepsilon_n}{2} I \tag{4.41}$$

with probability tending to one. On the same event $\max_i |q_i| = O(\sqrt{L_n})$, so the Hessian is $O(L_n^{3/2})$-Lipschitz near zero. It follows that

$$B_n(w) < 1 \quad \text{for} \quad 0 < |w| \le r_{0,n} := \frac{\varepsilon_n}{16 L_n^{3/2}}. \tag{4.42}$$

It remains to consider $r_{0,n} \le |w| \le r_n$. Put $g(r) = 1 - e^{-\varepsilon_n r^2/2}$. In this range

$$g(r) \ge c\varepsilon_n r_{0,n}^2 \ge c\frac{\varepsilon_n^3}{L_n^3} =: g_{0,n}.$$

The scale relations above give $\log g_{0,n}^{-1} = O(\log L_n)$. Cover each dyadic radial annulus by a net of mesh $cg(r)/\sqrt{L_n}$. The total logarithmic cardinality is $O(d_n \log L_n)$, and (4.39) shows that a crossing of level one would produce a net point at level at least $1 - g(r)/8$.

For the low cloud, apply (4.38) with $t = g(r)/4$. Uniformly over all annuli, its exponent is bounded below by

$$c\frac{g_{0,n}^2}{\mathfrak{j}_n} \gg d_n \log L_n.$$

Its mean is at most $1 - g(r)$ by (4.37). Lemma 4.14, with a one-unit buffer, shows that the upper cloud contains at most e^{3A_n} points with probability tending to one. On $|w| \le r_n = o(\sqrt{L_n})$, each corresponding bump is at most $e^{-L_n + o(L_n)}$, so their total contribution is $e^{-L_n+o(L_n)} = o(g_{0,n})$. Finally, empirical centering changes the logarithm of the field by at most

$$|w|\frac{|\bar{Z}_n|}{\sqrt{\gamma_n}}, \qquad \sup_{r \ge r_{0,n}} \frac{|\bar{Z}_n|/\sqrt{\gamma_n}}{\varepsilon_n r} = o_{\mathbb{P}}(1),$$

where the last relation follows from $|\bar{Z}_n|/\sqrt{\gamma_n} = O_{\mathbb{P}}(\sqrt{L_n/n})$ and

$$\frac{L_n^2}{\sqrt{n}\,\varepsilon_n^2} = o(1),$$

which follows from (4.22). The union bound now excludes every crossing outside the microscopic ball. Together with (4.42), this proves (4.40). □

Together with the upper-cloud separation established at the start of this subsection, Propositions 4.15 and 4.16 exclude all subthreshold contacts in the direct-truncation range.

4.5.6 The logarithmic crossover

The slowly growing and direct-truncation arguments cover the sparse-edge regimes except the crossover scale $d_n \asymp L_n^3/(\log L_n)^2$; the larger entropy regime is handled in Section 5. Here a direct entropy comparison is unnecessarily delicate. A spatial argument is more effective: distinct sample centers are separated by order $\sqrt{L_n}$, whereas a neighborhood of radius $\sqrt{\log L_n}$ already makes every nonnearest Gaussian bump negligible.

Lemma 4.17. *Fix $0 < c_- < c_+ < \infty$, put $\ell_n = \log L_n$, and suppose*

$$c_- \frac{L_n^3}{\ell_n^2} \le d_n \le c_+ \frac{L_n^3}{\ell_n^2}, \qquad s_n = O(1).$$

Then

$$\gamma_n \asymp \frac{L_n^2}{\ell_n^2}, \qquad \delta_n := \frac{a_n}{\gamma_n} \asymp \frac{\ell_n}{L_n}, \qquad \varepsilon_n \longrightarrow 1. \tag{4.43}$$

Moreover, with probability tending to one,

$$\max_i \frac{|Z_i|}{\sqrt{\gamma_n}} \le C\sqrt{L_n}, \qquad \min_{i \ne j} |q_i - q_j|^2 \ge 3L_n, \qquad \frac{|\bar{Z}_n|}{\sqrt{\gamma_n}} = O_{\mathbb{P}}\left(\sqrt{\frac{L_n}{n}}\right). \tag{4.44}$$

Proof. The scale relations follow from (4.11) and (4.12). Also $|\bar{Z}_n| = O_{\mathbb{P}}(\sqrt{d_n/n})$, which gives the last estimate in (4.44). A Gamma Chernoff bound and a union bound over i give the first estimate there. Finally,

$$q_i - q_j = \frac{Z_i - Z_j}{\sqrt{\gamma_n}}, \qquad \frac{|Z_i - Z_j|^2}{2} \sim \chi^2_{d_n}.$$

Since $d_n/\gamma_n \sim 2L_n$, the lower-tail Chernoff bound for $\chi^2_{d_n}$ is at most e^{-cd_n} at the threshold $|q_i - q_j|^2 = 3L_n$. This remains summable over the fewer than n^2 pairs because $d_n/L_n \to \infty$. □

We next use this separation together with truncation of the independent, uncentered field. The truncation is only needed away from the sample centers, where each individual bump is already small.

Proposition 4.18. *Under the assumptions of Lemma 4.17, for every fixed $\eta > 0$,*

$$\mathbb{P}\left\{\max_i \xi_{n,i} \le s_n - \eta, \qquad \sup_{w \in \mathbb{R}^{d_n}} B_n(w) > 1\right\} \longrightarrow 0.$$

Proof. Work on the high-probability events in Lemma 4.17 and on $\{\max_i \xi_{n,i} \le s_n - \eta\}$. Then

$$\max_i \frac{|q_i|^2}{2} \le L_n - g_n, \qquad g_n = \eta \frac{a_n}{\gamma_n} \asymp_\eta \frac{\ell_n}{L_n}.$$

Put

$$Q_i = \frac{Z_i}{\sqrt{\gamma_n}}, \qquad c_n = \frac{\bar{Z}_n}{\sqrt{\gamma_n}}, \qquad S_n(w) = \frac{1}{n}\sum_i e^{w \cdot Q_i - |w|^2/2}.$$

Then $q_i = Q_i - c_n$ and, exactly,

$$B_n(w) = e^{-w \cdot c_n} S_n(w), \qquad \mathbb{E}S_n(w) = e^{-\varepsilon_n |w|^2/2}. \tag{4.45}$$

Lemma 4.17 gives

$$\sup_{|w|\le\sqrt{2L_n}} |w\cdot c_n| = O_{\mathbb{P}}(L_n/\sqrt{n}) = o_{\mathbb{P}}(g_n). \tag{4.46}$$

In particular,

$$|c_n| = O_{\mathbb{P}}\Big(\sqrt{L_n/n}\Big) = o_{\mathbb{P}}(L_n^{-3/2}),$$

which will make centering negligible even in the smallest annulus below.

We first exclude neighborhoods of the sample centers. Write $R_n = \sqrt{8\ell_n}$. The bump representation and the radial buffer give

$$h_i = \frac{1}{n}e^{|q_i|^2/2} \le e^{-g_n} \le 1 - \frac{g_n}{2}$$

for all large n. If $|w - q_i| \le R_n$, pairwise separation gives

$$B_n(w) \le e^{-g_n} + n\exp\left\{-\frac{1}{2}(\sqrt{3L_n} - R_n)^2\right\} = 1 - \frac{g_n}{2} + e^{-L_n/2+o(L_n)} < 1.$$

Thus a crossing cannot occur within distance R_n of any q_i.

We next exclude a fixed ball around the origin. Standard Gaussian sample-covariance concentration and $d_n/n \to 0$ imply

$$\frac{1}{n}\sum_i q_i q_i^\top \preceq \frac{1}{4}I_{d_n}$$

with probability tending to one. Hence $\nabla^2 B_n(0) \preceq -3I_{d_n}/4$. On the radial-buffer event, $\max_i |q_i| \le \sqrt{2L_n}$, so the operator norm of the third derivative of B_n is at most $CL_n^{3/2}$ throughout a ball of radius $c_0 L_n^{-3/2}$, where $c_0 > 0$ is a sufficiently small universal constant. Taylor's theorem therefore gives

$$B_n(w) \le 1 - \frac{|w|^2}{4}, \qquad 0 < |w| \le r_{0,n} := c_0 L_n^{-3/2}. \tag{4.47}$$

It remains to pass from this microscopic ball to a fixed radius. On the event $\max_i |Q_i| \le C\sqrt{L_n}$,

$$\sup_{|w|\le 3} |\nabla \log S_n(w)| \le C\sqrt{L_n}. \tag{4.48}$$

Partition $\{r_{0,n} \le |w| \le 3\}$ into the $O(\ell_n)$ annuli $\mathcal{A}_j = \{r_j \le |w| \le 2r_j\}$, where $r_j = 2^j r_{0,n}$, truncating the last annulus at radius 3. For each j, take a deterministic net of $\mathcal{A}_j$ with mesh

$$\rho_j = c_1 \frac{r_j^2}{\sqrt{L_n}},$$

where $c_1 > 0$ is a sufficiently small constant. The logarithm of the total number of net points is $O(d_n\ell_n) = O(L_n^3/\ell_n)$.

For a fixed net point v, define the independent variables

$$Y_i(v) = \min\left\{\frac{1}{n}e^{v\cdot Q_i - |v|^2/2}, U_n\right\}, \qquad U_n = n^{-1}e^{C_1\sqrt{L_n}},$$

with C_1 chosen so that truncation is inactive on $\{\max_i |Q_i| \le C\sqrt{L_n}\}$ for $|v| \le 3$. Since $\varepsilon_n \ge 1/2$ for all large n,

$$\sum_i \mathbb{E}Y_i(v) \le e^{-\varepsilon_n|v|^2/2} \le 1 - c_2|v|^2, \qquad |v| \le 3,$$

for a numerical $c_2 > 0$, while $\sum_i \operatorname{Var} Y_i(v) \le U_n$. Bernstein's inequality gives

$$\mathbb{P}\left\{\sum_i Y_i(v) > e^{-\varepsilon_n |v|^2/2} + c_3 r_j^2\right\} \le \exp\{-c r_j^4 / U_n\}.$$

The exponent is at least $n^{1-o(1)} L_n^{-6}$, uniformly in j, and therefore dominates the total net entropy. Thus, with probability tending to one, $S_n(v) \le 1 - c_4 r_j^2$ at every such net point.

If $B_n(w) \ge 1$ at some $w \in \mathcal{A}_j$, then (4.45) and $|c_n| = o_{\mathbb{P}}(r_{0,n})$ give $S_n(w) \ge 1 - o(r_j^2)$. At the nearest net point v, (4.48) and the choice of c_1 then give $S_n(v) > 1 - c_4 r_j^2$, a contradiction. In combination with (4.47), this proves

$$\sup_{0<|w|\le 3} B_n(w) < 1 \tag{4.49}$$

with probability tending to one on the radial-buffer event.

We finally exclude the part of the field outside the fixed ball and the center neighborhoods. Let $\mathcal{G}_n$ be a deterministic $1/4$-net of $B(0, \sqrt{2L_n})$. Then

$$\log |\mathcal{G}_n| \le d_n \log(C\sqrt{L_n}) = O(L_n^3/\ell_n).$$

For $v \in \mathcal{G}_n$ with $|v| \ge 11/4$, set

$$\widetilde{Y}_i(v) = \min\left\{\frac{1}{n} e^{v\cdot Q_i - |v|^2/2}, \mathfrak{b}_n\right\}, \qquad \mathfrak{b}_n = 2L_n^{-15/4}.$$

These variables are independent, and

$$\sum_i \mathbb{E}\widetilde{Y}_i(v) \le e^{-\varepsilon_n |v|^2/2} < 0.05$$

for all large n. Since $\sum_i \operatorname{Var} \widetilde{Y}_i(v) \le 0.05\mathfrak{b}_n$, Bernstein's inequality and a union bound give

$$\mathbb{P}\left\{\max_{\substack{v\in\mathcal{G}_n \\ |v|\ge 11/4}} \sum_i \widetilde{Y}_i(v) > 0.9\right\} \le \exp\{O(L_n^3/\ell_n) - cL_n^{15/4}\} = o(1). \tag{4.50}$$

Suppose now that $\sup_w B_n(w) > 1$, and let w_* be a global maximizer. Lemma 4.11 places w_* in the convex hull of the q_i's, hence in $B(0, \sqrt{2L_n})$. The preceding two parts of the proof imply

$$|w_*| > 3, \qquad \min_i |w_* - q_i| > R_n.$$

Let $v \in \mathcal{G}_n$ satisfy $|v - w_*| \le 1/4$. The semiconvex estimate (4.24) gives

$$B_n(v) \ge e^{-1/32} B_n(w_*) > e^{-1/32}.$$

Moreover, $|v| \ge 11/4$ and $\min_i |v - q_i| \ge R_n - 1/4$. The exact identity

$$\frac{1}{n} e^{v\cdot Q_i - |v|^2/2} = e^{v\cdot c_n} h_i e^{-|v-q_i|^2/2},$$

together with the radial-buffer bound on h_i and (4.46), shows that every uncentered summand at v is at most

$$e^{|v\cdot c_n|} e^{-(R_n - 1/4)^2/2} \le \mathfrak{b}_n.$$

Thus truncation is inactive and, by (4.45),

$$\sum_i \widetilde{Y}_i(v) = S_n(v) = e^{v\cdot c_n} B_n(v) > 0.9$$

for all large n, contradicting (4.50). □

The exclusion propositions in this section cover, respectively, slowly growing dimensions, the direct truncation range, and the remaining crossover band. They are combined with the entropy estimate in Proposition 6.18; no separate regime-by-regime limit theorem is needed.

5 Entropy and the Simplex Regime

This section completes the spatial-exclusion part of the $d_n \geq 3$ analysis in Theorem 1.1. Once Euclidean covering becomes inefficient, an entropy dual reduces the KKT optimization to the nearly regular simplex formed by the centered sample. The resulting estimates both rule out subthreshold contacts and identify the finite-sample correction in the exact Gumbel coordinate $s_n^\star$.

5.1 Entropy reduction above the sparse-edge range

The spatial truncation argument is no longer efficient once the dimension is much larger than $(\log n)^3$. We use the entropy dual under

$$\frac{d_n\{\log(d_n/L_n)\}^2}{L_n^3} \longrightarrow \infty, \qquad d_n L_n = o(n^2). \tag{5.1}$$

This range includes $d_n = \lfloor (\log n)^4 \rfloor$ and $d_n \asymp n$. The entropy dual (2.5) provides a different route: away from the vertices of the probability simplex, entropy beats the quadratic likelihood gain by a quantitative margin; near a vertex, the corresponding radial observation controls the optimization.

5.1.1 Deterministic entropy inequalities

We first record the entropy estimate used to separate the diffuse and near-vertex regions of the simplex. Write $H(p) = -\sum_i p_i \log p_i$, so that $K(p\|\mathbf{u}_n) = L_n - H(p)$, and let

$$h_{\mathrm{bin}}(y) = -y \log y - (1-y)\log(1-y)$$

denote binary entropy, with the usual continuous values at zero and one.

Lemma 5.1. *There are universal constants $c_0 > 0$ and $n_0 < \infty$ such that, for all $n \geq n_0$ and $p \in \Delta_n$,*

$$K(p\|\mathbf{u}_n) \geq \frac{n-1}{n-2}\log(n-1)|p - \mathbf{u}_n|_2^2 \geq L_n |p - \mathbf{u}_n|_2^2. \tag{5.2}$$

Moreover, if $\max_i p_i \leq 1 - L_n^{-2}$, *then*

$$K(p\|\mathbf{u}_n) - L_n|p - \mathbf{u}_n|_2^2 \geq \frac{c_0}{L_n}|p - \mathbf{u}_n|_1^2. \tag{5.3}$$

Proof. Write

$$\mathfrak{c}_n = \frac{n-1}{n-2}\log(n-1), \qquad \beta_n = 2\mathfrak{c}_n,$$

and first consider

$$G_n(p) = K(p\|\mathbf{u}_n) - \mathfrak{c}_n|p - \mathbf{u}_n|_2^2.$$

We identify the global minima of G_n. A minimum cannot lie on the boundary of the simplex: if $p_j = 0 < p_i$, transferring mass t from coordinate i to coordinate j changes the entropy part by $t \log t + O(t)$, whereas the quadratic part changes by $O(t)$. The total change is therefore negative for all sufficiently small $t > 0$.

At an interior minimum, the Lagrange equations are

$$\log p_i - \beta_n p_i = \text{constant}. \tag{5.4}$$

The function $x \mapsto \log x - \beta_n x$ increases up to $1/\beta_n$ and decreases thereafter, so the coordinates of a stationary point take at most two values. If these values are $a > b$, then $b < 1/\beta_n < a$. Moreover, the larger value has multiplicity one: if two coordinates were equal to a, the second variation in the direction that transfers mass between them would be $2(1/a - \beta_n) < 0$. Thus every nonuniform local minimum has the form

$$p(a) = \left(a, \frac{1-a}{n-1}, \dots, \frac{1-a}{n-1}\right), \qquad \frac{1}{n} < a < 1.$$

Put $x = (n-1)a/(1-a)$. Direct differentiation gives

$$\frac{d}{da} G_n(p(a)) = g_n(x), \qquad g_n(x) = \log x - \beta_n \frac{x-1}{x+n-1}.$$

The three zeros

$$x = 1, \qquad x = n-1, \qquad x = (n-1)^2$$

are immediate from the definition of β_n. They are the only zeros: indeed,

$$g_n'(x) = \frac{1}{x} - \frac{n\beta_n}{(x+n-1)^2}$$

has at most two zeros. Also $g_n'(1) > 0$ and $g_n'((n-1)^2) > 0$, because $\beta_n < n$. For $n = 3$ the second inequality is immediate; for $n \geq 4$ it follows from $\log m < m/2 - 1/(2m)$, $m = n-1$: this holds at $m = 3$, and the derivative of the right side minus the left side is $(m-1)^2/(2m^2) > 0$. Hence g_n is positive on $(1, n-1)$, negative on $(n-1, (n-1)^2)$, and positive on $((n-1)^2, \infty)$. Now $G_n(\mathbf{u}_n) = 0$, while at $x = (n-1)^2$,

$$p(a) = \left(\frac{n-1}{n}, \frac{1}{n(n-1)}, \dots, \frac{1}{n(n-1)}\right) \quad \text{and} \quad G_n(p(a)) = 0.$$

The preceding sign pattern therefore gives $G_n(p(a)) \geq 0$ for every $a \in [1/n, 1]$. Since every global minimum of G_n is among the points just considered, this proves $K(p\|\mathbf{u}_n) \geq \mathfrak{c}_n |p - \mathbf{u}_n|_2^2$. Finally, $\mathfrak{c}_n \geq \log n = L_n$ for $n \geq 3$. For example, after writing $m = n-1$, this is equivalent to $r(m) := m \log m - (m-1)\log(m+1) \geq 0$. Here $r(2) > 0$, and

$$r'(m) = \log \frac{m}{m+1} + \frac{2}{m+1} > 0 \qquad (m \geq 2),$$

because $-\log(1-t) < t/(1-t) < 2t$ for $t = 1/(m+1)$. This proves (5.2).

We next prove the quantitative gap from the vertices. In the remainder of the proof take $L = L_n \geq 16$ and set

$$F_n(p) = K(p\|\mathbf{u}_n) - L|p - \mathbf{u}_n|_2^2, \qquad T(p) = |p - \mathbf{u}_n|_1.$$

We shall prove

$$F_n(p) \geq \frac{1}{32L} T(p)^2 \qquad \text{if} \qquad \max_i p_i \leq 1 - L^{-2}. \tag{5.5}$$

We first dispose of diffuse vectors. The elementary scalar inequality

$$x \log x - x + 1 \geq \frac{(x-1)^2}{2(x+1)}, \qquad x \geq 0,$$

follows by differentiation. Applied coordinatewise, it gives

$$K(p\|\mathbf{u}_n) \geq \frac{1}{2}A(p), \qquad A(p) = \sum_i \frac{(p_i - 1/n)^2}{p_i + 1/n}.$$

If $\max_i p_i \leq (8L)^{-1}$, then $p_i + 1/n \leq (4L)^{-1}$ for all i, once $L \geq 16$. Consequently,

$$L|p - \mathbf{u}_n|_2^2 \leq \frac{1}{4}A(p), \qquad T(p)^2 \leq A(p)\sum_i (p_i + 1/n) = 2A(p),$$

and hence $F_n(p) \geq T(p)^2/8$, which is stronger than (5.5).

It remains to consider $m = \max_i p_i \geq (8L)^{-1}$. Relabel so that $p_1 = m$, put $y = 1 - m$, and write

$$p = (m, yq_1, \dots, yq_{n-1}), \qquad q \in \Delta_{n-1}.$$

Let $\bar{p}_m = (m, y/(n-1), \dots, y/(n-1))$, and define

$$f_n(m) = F_n(\bar{p}_m), \qquad t_0 = |\bar{p}_m - \mathbf{u}_n|_1 = 2(m - 1/n).$$

The entropy chain rule and the orthogonal decomposition of the squared Euclidean distance give the exact identity

$$F_n(p) = f_n(m) + y\left\{K(q\|\mathbf{u}_{n-1}) - Ly|q - \mathbf{u}_{n-1}|_2^2\right\}. \tag{5.6}$$

The first part of the lemma in dimension $n-1$ gives

$$K(q\|\mathbf{u}_{n-1}) \geq C_{n-1}|q - \mathbf{u}_{n-1}|_2^2, \qquad C_{n-1} = \frac{n-2}{n-3}\log(n-2) \geq L.$$

The last inequality holds for $n \geq 4$: with $x = n-2$, it is equivalent to $x \log x \geq (x-1)\log(x+2)$; the derivative of the difference is $\log(1-t) + 3t/2$, where $t = 2/(x+2) \in (0, 1/2]$. This derivative is nonnegative because $-\log(1-t)/t$ is increasing and its value at $t = 1/2$ is $2\log 2 < 3/2$. Pinsker's inequality then yields

$$K(q\|\mathbf{u}_{n-1}) - Ly|q - \mathbf{u}_{n-1}|_2^2 \geq \left(1 - \frac{Ly}{C_{n-1}}\right)K(q\|\mathbf{u}_{n-1}) \geq \frac{m}{2}|q - \mathbf{u}_{n-1}|_1^2. \tag{5.7}$$

Here Pinsker's inequality with natural logarithms follows by grouping the positive and negative coordinate deviations and applying the binary bound whose second derivative is at least four.

It remains only to bound f_n. Put

$$x = \frac{(n-1)m}{1-m}, \qquad \widetilde{g}_n(x) = \log x - 2L\frac{x-1}{x+n-1}.$$

Then $f_n'(m) = \widetilde{g}_n(x)$, while

$$\widetilde{g}_n'(x) = \frac{1}{x} - \frac{2nL}{(x+n-1)^2}$$

has at most two zeros. Thus $\widetilde{g}_n$ has at most three zeros on $(0, \infty)$, one of which is $x = 1$. Moreover, $\widetilde{g}_n'(1) = 1 - 2L/n > 0$. At the left endpoint $m = (8L)^{-1}$, we have $x_0 = (n-1)/(8L-1)$ and, since

$$2L\frac{x_0 - 1}{x_0 + n - 1} = \frac{n - 8L}{4(n-1)} < \frac{1}{4},$$

$$\widetilde{g}_n(x_0) > L - \log(16L) - \frac{1}{4} > 0.$$

At the right endpoint $M = 1 - L^{-2}$, where $x = (n-1)(L^2-1)$,

$$\begin{aligned}\widetilde{g}_n(x) &= -L + 2\log L + \log(1-1/n) + \log(1-L^{-2}) + \frac{2n}{(n-1)L}\\ &\le -L + 2\log L + \frac{4}{L} < 0.\end{aligned}$$

Thus the derivative of f_n is positive at the left endpoint and negative at the right endpoint. If f_n had an interior local minimum, its derivative would have to cross first from positive to negative, then from negative to positive at that minimum, and finally back to negative before M. Together with the zero at $x = 1$, this would give at least four zeros of $\widetilde{g}_n$, a contradiction. Therefore the minimum of f_n on this interval occurs at an endpoint.

At $m = (8L)^{-1}$, using $-(1-m)\log(1-m) \le m$ and $|\bar{p}_m - \mathbf{u}_n|_2^2 \le 2m^2$, we obtain

$$f_n(m) \ge m\{L - \log(8L) - 1\} - 2Lm^2 = m\{L - \log(8L) - 5/4\} \ge \frac{1}{16}.$$

At $m = M$, put $y = L^{-2}$. Since $|\bar{p}_m - \mathbf{u}_n|_2^2 \le m^2$ and $h_{\mathrm{bin}}(y) \le y\log(e/y)$,

$$\begin{aligned}f_n(M) &\ge Ly(1-y) - h_{\mathrm{bin}}(y)\\ &\ge \frac{1}{L} - \frac{1}{L^3} - \frac{1+2\log L}{L^2} \ge \frac{1}{2L}.\end{aligned}$$

All numerical inequalities in the last three displays hold for $L \ge 16$. It follows that, throughout $[(8L)^{-1}, M]$,

$$f_n(m) \ge \frac{1}{2L} \ge \frac{1}{8L}t_0^2. \tag{5.8}$$

Combining (5.6), (5.7), and (5.8), and using $m \ge (8L)^{-1}$, gives

$$F_n(p) \ge \frac{t_0^2}{8L} + \frac{y^2}{16L}|q - \mathbf{u}_{n-1}|_1^2.$$

Finally, $T(p) \le t_0 + y|q - \mathbf{u}_{n-1}|_1$, so the last display implies (5.5). Thus (5.3) holds with $c_0 = 1/32$ and, for example, $n_0 = \lceil e^{16} \rceil$. □

Remark 5.2. The sharp constant in (5.2) is the mean-field Potts transition coefficient. The proof also identifies all equality cases: the uniform distribution and the n permutations of $((n-1)/n, 1/[n(n-1)], \dots, 1/[n(n-1)])$.

The next proposition converts the entropy estimate into a deterministic one-atom criterion. It is tailored to a radial gap much smaller than the largest off-diagonal inner product.

Proposition 5.3. *Let $q_1, \dots, q_n \in \mathbb{R}^d$ satisfy $\sum_i q_i = 0$. Put*

$$\mu_n^{\mathrm{off}} = \max_{i \ne j} |q_i \cdot q_j|.$$

Suppose, for some $0 < g_n \le 1$,

$$\max_i \frac{|q_i|^2}{2} \le L_n - g_n, \qquad \mu_n^{\mathrm{off}} = o(L_n^{-1}), \qquad ng_n \longrightarrow \infty. \tag{5.9}$$

Then, for all sufficiently large n,

$$\sup_{w \in \mathbb{R}^d} B_n(w) \le 1.$$

Proof. Fix $p \in \Delta_n$. Since $\sum_i q_i = 0$, writing $r = p - \mathbf{u}_n$ gives

$$\frac{1}{2}\left|\sum_i p_i q_i\right|^2 = \frac{1}{2}\left|\sum_i r_i q_i\right|^2 \le L_n |r|_2^2 + \frac{\mu_n^{\mathrm{off}}}{2}|r|_1^2.$$

If $\max_i p_i \le 1 - L_n^{-2}$, Lemma 5.1 and $\mu_n^{\mathrm{off}} = o(L_n^{-1})$ show that this is at most $K(p\|\mathbf{u}_n)$.

It remains to consider a distribution with $\max_i p_i = 1 - y > 1 - L_n^{-2}$. Let i be its largest coordinate and write

$$\sum_j p_j q_j = (1-y)q_i + y\bar{q},$$

where $\bar{q}$ is a convex combination of the remaining q_j's. Convexity of the squared norm and (5.9) give

$$\frac{|\bar{q}|^2}{2} \le L_n - g_n, \qquad q_i \cdot \bar{q} \le \mu_n^{\mathrm{off}}.$$

Consequently,

$$\frac{1}{2}\left|\sum_j p_j q_j\right|^2 \le (L_n - g_n)\{(1-y)^2 + y^2\} + \mu_n^{\mathrm{off}} y(1-y). \tag{5.10}$$

On the other hand,

$$H(p) \le h_{\mathrm{bin}}(y) + y\log(n-1) \le h_{\mathrm{bin}}(y) + yL_n.$$

Subtracting (5.10) from $K(p\|\mathbf{u}_n) = L_n - H(p)$, and using $y \le L_n^{-2}$, gives

$$\begin{aligned} K(p\|\mathbf{u}_n) - \frac{1}{2}\left|\sum_j p_j q_j\right|^2 &\ge g_n + \{2(L_n - g_n) - \mu_n^{\mathrm{off}}\}y(1-y) - h_{\mathrm{bin}}(y) - yL_n \\ &\ge g_n + y\{L_n - 4 + \log y\} \end{aligned}$$

for all large n, because $h_{\mathrm{bin}}(y) \le y\log(e/y)$. The minimum of $y \mapsto y(L_n - 4 + \log y)$ over $y \ge 0$ is $-e^3/n$. The last display is therefore positive by $ng_n \to \infty$. The entropy dual in Lemma 2.5 completes the proof. ◻

In very high dimension, the absolute off-diagonal bound in Proposition 5.3 discards the regular-simplex geometry created by centering. The next criterion retains it. The quantity C_n measures the deviation of each row of the Gram matrix from the identity $q_i \cdot \{(n-1)^{-1}\sum_{j\ne i} q_j\} = -|q_i|^2/(n-1)$.

Proposition 5.4. *Let $q_1, \ldots, q_n \in \mathbb{R}^d$ satisfy $\sum_i q_i = 0$, and put*

$$G_n = (q_i \cdot q_j)_{i,j\le n}, \qquad \mu_n^{\mathrm{off}} = \max_{i\ne j}|q_i \cdot q_j|, \qquad C_n = \max_{i\ne j}\left|q_i \cdot q_j + \frac{|q_i|^2}{n-1}\right|.$$

Suppose, for some $g_n > 0$, that

$$\max_i \frac{|q_i|^2}{2} \le R_n^\star - g_n, \qquad \mu_n^{\mathrm{off}} = o(L_n^{-1}), \qquad \|G_n\|_{\mathrm{op}} = O(L_n), \qquad C_n = o(g_n L_n^2).$$

Then, for all sufficiently large n,

$$\sup_{w\in\mathbb{R}^d} B_n(w) \le 1.$$

Proof. Fix $p \in \Delta_n$. If $\max_i p_i \le 1 - L_n^{-2}$, the argument away from the vertices in Proposition 5.3 applies: since $R_n^\star < L_n$, Lemma 5.1 and $\mu_n^{\mathrm{off}} = o(L_n^{-1})$ give

$$\frac{1}{2}\left|\sum_i p_i q_i\right|^2 \le K(p\|\mathbf{u}_n).$$

Suppose instead that $p_i = 1 - y > 1 - L_n^{-2}$. Write $p_j = y r_j$ for $j \ne i$, where $r = (r_j)_{j\ne i} \in \Delta_{n-1}$, and let $\mathbf{u}_{n-1}$ denote the uniform distribution on these $n-1$ coordinates. Set

$$c_y = 1 - \frac{ny}{n-1}, \qquad v_r = \sum_{j\ne i}\left(r_j - \frac{1}{n-1}\right) q_j.$$

Since $\sum_{j\ne i} q_j = -q_i$,

$$\sum_j p_j q_j = c_y q_i + y v_r.$$

The relative entropy separates as

$$K(p\|\mathbf{u}_n) = k_n(y) + yK(r\|\mathbf{u}_{n-1}),$$

where

$$k_n(y) = (1-y)\log\{n(1-y)\} + y\log\frac{ny}{n-1}.$$

Apply the sharp inequality (5.2) to the vector $(1-y, y/(n-1), \dots, y/(n-1))$. Since its squared distance from $\mathbf{u}_n$ is $(n-1)c_y^2/n$, we obtain

$$k_n(y) \ge R_n^\star c_y^2. \tag{5.11}$$

The definition of C_n gives

$$|q_i \cdot v_r| = \left|\sum_{j\ne i} r_j\left(q_i\cdot q_j + \frac{|q_i|^2}{n-1}\right)\right| \le C_n.$$

Moreover,

$$|v_r|^2 \le \|G_n\|_{\mathrm{op}}|r - \mathbf{u}_{n-1}|_2^2 \le \frac{\|G_n\|_{\mathrm{op}}}{\log(n-1)} K(r\|\mathbf{u}_{n-1}),$$

where the last inequality is (5.2) with $n-1$ in place of n. Because $y \le L_n^{-2}$ and $\|G_n\|_{\mathrm{op}} = O(L_n)$, it follows that

$$K(r\|\mathbf{u}_{n-1}) - \frac{y}{2}|v_r|^2 \ge 0$$

for all large n. Combining this with (5.11) and $|q_i|^2/2 \le R_n^\star - g_n$ gives

$$\begin{aligned} K(p\|\mathbf{u}_n) - \frac{1}{2}\left|\sum_j p_j q_j\right|^2 &\ge g_n c_y^2 - y|c_y|C_n \\ &\ge g_n\{1 - O(L_n^{-2})\} - \frac{C_n}{L_n^2} > 0. \end{aligned}$$

The entropy dual in Lemma 2.5 completes the proof. ∎

The local-simplex criterion applies when d_nL_n/n^2 is bounded below. When $d_nL_n = o(n^2)$, the following criterion is more convenient: it separates diffuse probability vectors, controlled spectrally, from their heavy coordinates, controlled by entropy and μ_n^{off}.

Proposition 5.5. *Let $q_1, \dots, q_n \in \mathbb{R}^d$ satisfy $\sum_i q_i = 0$, and put*

$$G = (q_i \cdot q_j)_{i,j\le n}, \qquad \upsilon_n = \frac{\|G\|_{\mathrm{op}}}{2n}, \qquad \mu_n^{\mathrm{off}} = \max_{i\neq j} |q_i \cdot q_j|.$$

Suppose $\max_i |q_i|^2/2 \le L_n - g_n$*, where $0 < g_n \le 1$. Assume that there are deterministic $a_n^\circ \to \infty$ and $y_n \downarrow 0$, with $T_n = e^{a_n^\circ}$, such that*

$$a_n^\circ = o(L_n), \qquad \frac{\upsilon_n T_n}{a_n^\circ} = o(1), \qquad \frac{\mu_n^{\mathrm{off}} + L_nT_n/n}{a_n^\circ} = o(1), \qquad \frac{L_n y_n}{a_n^\circ} \longrightarrow \infty, \tag{5.12}$$

and

$$e^{2L_ny_n+\mu_n^{\mathrm{off}}} = o(ng_n). \tag{5.13}$$

Then, for all sufficiently large n,

$$\sup_{w\in\mathbb{R}^d} B_n(w) \le 1.$$

Proof. Fix $p \in \Delta_n$, write $f_i = np_i$, and let $\phi(x) = x\log x - x + 1$. Then

$$K(p\|\mathbf{u}_n) = \frac{1}{n}\sum_i \phi(f_i).$$

Set

$$\ell_i = f_i \wedge T_n, \qquad h_i = (f_i - T_n)_+, \qquad \bar\ell = \frac{1}{n}\sum_i \ell_i.$$

Since $\sum_i q_i = 0$,

$$\sum_i p_iq_i = \frac{1}{n}\sum_i(\ell_i - \bar\ell)q_i + \frac{1}{n}\sum_i h_iq_i =: v_0 + v_1.$$

The elementary inequality

$$(x-1)^2 \le C\frac{T_n}{a_n^\circ}\phi(x), \qquad 0 \le x \le T_n,$$

and the definition of υ_n give, with $K_0 = n^{-1}\sum_i \phi(\ell_i)$,

$$\frac{|v_0|^2}{2} \le C\frac{\upsilon_nT_n}{a_n^\circ}K_0 = o(K_0). \tag{5.14}$$

Put $b_i = h_i/n$ and $z = \sum_i b_i$. The diagonal and off-diagonal parts of the Gram matrix give

$$\frac{|v_1|^2}{2} \le L_n\sum_i b_i^2 + \frac{\mu_n^{\mathrm{off}}}{2}z^2, \qquad |v_0\cdot v_1| \le \left(2\mu_n^{\mathrm{off}} + \frac{2L_nT_n}{n}\right)z. \tag{5.15}$$

Suppose first that $\max_i p_i \le 1 - y_n$. Uniformly for $T_n \le x \le n(1-y_n)$, one-variable calculus gives

$$\phi(x) - \phi(T_n) \ge \frac{L_n}{n}(x-T_n)^2 + \frac{a_n^\circ}{4}(x-T_n). \tag{5.16}$$

Indeed, after subtracting the right side, the derivative first increases and then decreases, and is positive at T_n. The minimum is therefore at an endpoint. At $x_n = n(1-y_n)$, the difference between the two sides is

$$\begin{aligned} & x_n\left\{L_n y_n + \log(1-y_n) - 1 - \frac{a_n^\circ}{4}\right\} \\ & \quad + 2L_n(1-y_n)T_n - \frac{3a_n^\circ T_n}{4} + T_n - \frac{L_n T_n^2}{n}. \end{aligned}$$

The first line is positive and of order $nL_n y_n$, while the absolute value of the second line is $o(nL_n y_n)$ by (5.12). Thus the upper-endpoint value is positive. Summing (5.16) yields

$$K(p\|\mathbf{u}_n) - K_0 \geq L_n \sum_i b_i^2 + \frac{a_n^\circ}{4} z.$$

Equations (5.12), (5.14), and (5.15) now show that $|v_0 + v_1|^2/2 \leq K(p\|\mathbf{u}_n)$: indeed, $z \leq 1$, and the last two conditions in (5.12) make the off-diagonal and cross terms in (5.15) at most $a_n^\circ z/4$, while (5.14) is at most K_0.

It remains to consider $p_i = 1 - y > 1 - y_n$ for some i. If $\bar{q}$ is the barycenter of the remaining q_j's under their normalized weights, convexity gives

$$\frac{1}{2}|(1-y)q_i + y\bar{q}|^2 \leq (L_n - g_n)\{(1-y)^2 + y^2\} + \mu_n^{\mathrm{off}} y(1-y).$$

On the other hand, $K(p\|\mathbf{u}_n) \geq (1-y)L_n - h_{\mathrm{bin}}(y)$. Their difference is at least

$$\frac{g_n}{2} + y\{L_n - 2L_n y - \mu_n^{\mathrm{off}} + \log y - 1\}.$$

The negative part of the second term is at most $\exp\{-L_n + 2L_n y_n + \mu_n^{\mathrm{off}} + 1\} = o(g_n)$ by (5.13). Thus the difference is positive. The entropy dual in Lemma 2.5 completes the proof. $\square$

5.1.2 The random Gram matrix

Under (5.1), the next lemma verifies the off-diagonal and spectral Gram assumptions of Proposition 5.5.

Lemma 5.6. *Suppose the first condition in* (5.1) *holds and*

$$\frac{\gamma_n L_n}{\sigma_n^2 b_n} \longrightarrow 1.$$

For

$$q_i = \frac{Z_i - \bar{Z}_n}{\sqrt{\gamma_n}}, \qquad \mu_n^{\mathrm{off}} = \max_{i \neq j} |q_i \cdot q_j|, \qquad C_n = \max_{i\neq j}\left|q_i \cdot q_j + \frac{|q_i|^2}{n-1}\right|, \qquad \upsilon_n = \frac{\|(q_i \cdot q_j)_{i,j\leq n}\|_{\mathrm{op}}}{2n},$$

one has

$$\mu_n^{\mathrm{off}} = O_{\mathbb{P}}\left\{\frac{L_n}{n} + \frac{L_n^{3/2}}{\sqrt{d_n}} + \frac{L_n^2}{d_n}\right\}, \qquad \upsilon_n = O_{\mathbb{P}}\left(\frac{L_n}{d_n} + \frac{L_n}{n}\right). \tag{5.17}$$

Moreover,

$$\delta_n^\star := \frac{\sigma_n^2 a_n}{\gamma_n} = \sqrt{\frac{L_n}{d_n}}\{1 + o(1)\}. \tag{5.18}$$

In addition,

$$C_n = O_{\mathbb{P}}\left(\frac{L_n^{3/2}}{\sqrt{d_n}}\right). \tag{5.19}$$

If $d_n L_n/n^2$ is bounded below by a positive constant, then

$$\mu_n^{\mathrm{off}} = o_{\mathbb{P}}(L_n^{-1}), \qquad \|(q_i \cdot q_j)_{i,j\le n}\|_{\mathrm{op}} = O_{\mathbb{P}}(L_n). \tag{5.20}$$

Proof. Section 4.2, with $k_n = d_n/2 \gg L_n$, gives

$$b_n \sim k_n, \qquad a_n \sim \sqrt{\frac{k_n}{2L_n}}, \qquad \gamma_n \sim \frac{\sigma_n^2 b_n}{L_n} \sim \frac{d_n}{2L_n}.$$

This proves (5.18).

For the Gram bound, let $P_n = I_n - \mathbf{1}\mathbf{1}^\top/n$. For every coordinate ℓ, the vector

$$(\widetilde{Z}_{1\ell}, \ldots, \widetilde{Z}_{n\ell})$$

is centered Gaussian with covariance P_n, and these vectors are independent over ℓ. Hence, for $i \neq j$,

$$\mathbb{E}(\widetilde{Z}_i \cdot \widetilde{Z}_j) = -\frac{d_n}{n}.$$

Products of two standard Gaussian variables have uniformly bounded subexponential norm. Bernstein's inequality and a union bound over fewer than n^2 pairs therefore give

$$\max_{i\neq j}\left|\widetilde{Z}_i \cdot \widetilde{Z}_j + \frac{d_n}{n}\right| = O_{\mathbb{P}}\{\sqrt{d_n L_n} + L_n\}.$$

Dividing by $\gamma_n \asymp d_n/L_n$ yields

$$\mu_n^{\mathrm{off}} = O_{\mathbb{P}}\left\{\frac{L_n}{n} + \frac{L_n^{3/2}}{\sqrt{d_n}} + \frac{L_n^2}{d_n}\right\},$$

as claimed. The nonzero eigenvalues of the Gram matrix agree with those of $\gamma_n^{-1}\widetilde{Z}^\top\widetilde{Z}$. Hence the standard Gaussian operator-norm bound gives

$$\upsilon_n \le \frac{C}{\gamma_n}\left(1 + \sqrt{\frac{d_n}{n}}\right)^2 = O_{\mathbb{P}}\left(\frac{L_n}{d_n} + \frac{L_n}{n}\right).$$

For the centered row bound, Gaussian regression gives, conditionally on $\widetilde{Z}_i$,

$$\widetilde{Z}_j + \frac{\widetilde{Z}_i}{n-1} \sim \mathcal{N}\left(0, \frac{n-2}{n-1} I_{d_n}\right) \qquad (j \neq i).$$

Thus

$$\widetilde{Z}_i \cdot \left(\widetilde{Z}_j + \frac{\widetilde{Z}_i}{n-1}\right)$$

is conditionally centered Gaussian with variance at most $|\widetilde{Z}_i|^2$. Since $\max_i |\widetilde{Z}_i|^2 = O_{\mathbb{P}}(d_n)$ in the present regime, a conditional Gaussian tail bound and a union bound over the $n(n-1)$ pairs give

$$\max_{i\neq j}\left|\widetilde{Z}_i \cdot \left(\widetilde{Z}_j + \frac{\widetilde{Z}_i}{n-1}\right)\right| = O_{\mathbb{P}}(\sqrt{d_n L_n}).$$

Division by $\gamma_n \asymp d_n/L_n$ proves (5.19).

Finally, suppose $d_n L_n/n^2$ is bounded below. Since

$$\mu_n^{\text{off}} \le C_n + \frac{\max_i |q_i|^2}{n-1}$$

and $\max_i |q_i|^2 = O_{\mathbb{P}}(L_n)$, (5.19) gives $\mu_n^{\text{off}} = o_{\mathbb{P}}(L_n^{-1})$. Applying the standard Gaussian operator-norm bound once more gives

$$\|(q_i \cdot q_j)_{i,j\le n}\|_{\text{op}} = O_{\mathbb{P}}\left\{\frac{(\sqrt{d_n}+\sqrt{n})^2}{\gamma_n}\right\} = O_{\mathbb{P}}(L_n).$$

This proves (5.20). □

5.2 Exact centering and the high-dimensional correction

The entropy reduction in this section uses $\log n$ as the radial threshold. When $d_n \asymp n^2 \log n$, empirical centering and the regular-simplex geometry shift that threshold by an amount visible in the Gumbel coordinate. The exact coordinate $s_n^\star$ in (2.13) includes both effects.

The following calculation translates the exact coordinate into a critical variance defect and records when this finite-sample correction is visible.

Lemma 5.7. *Define*

$$\varepsilon_{c,n}^\star = 1 - \frac{R_n^\star}{\sigma_n^2 b_n}, \qquad w_n^\star = \frac{a_n R_n^\star}{\sigma_n^2 b_n^2}. \tag{5.21}$$

If $x_n = (\varepsilon_n - \varepsilon_{c,n}^\star)/w_n^\star = O(1)$, then

$$s_n^\star = \frac{x_n}{1 - a_n x_n/b_n} = x_n + o(1).$$

Let

$$s_n^{\text{cen}} = \frac{\gamma_n \log n/\sigma_n^2 - b_n}{a_n}.$$

If $s_n^\star = O(1)$, then

$$s_n^\star = s_n^{\text{cen}} + o(1) \quad \textit{when } d_n = o(n^2 \log n),$$

whereas, if $d_n/(n^2 \log n) \to \tau \in (0,\infty)$,

$$s_n^\star = s_n^{\text{cen}} - \sqrt{\tau} + o(1). \tag{5.22}$$

Proof. The first identity follows by substituting $\varepsilon_n = \varepsilon_{c,n}^\star + x_n w_n^\star$ into $\gamma_n = (1-\varepsilon_n)^{-1}$; Lemma 2.7 gives $a_n/b_n \to 0$. Also

$$s_n^{\text{cen}} - s_n^\star = \frac{\gamma_n(\log n - R_n^\star)}{\sigma_n^2 a_n}.$$

Under $s_n^\star = O(1)$, the expansion $R_n^\star = \log n - n^{-1} + O(\log n/n^2)$, together with the Gamma asymptotics in Section 4.2, gives

$$s_n^{\text{cen}} - s_n^\star = \frac{\sqrt{d_n}}{n\sqrt{\log n}} + o(1)$$

in the superlogarithmic range; the error is smaller when $d_n = O(\log n)$. The two assertions follow. □

6 From Radial Edge Points to NPMLE Atoms

This section completes the proofs of Theorems 1.1 and 1.2. Sections 3–5 locate every possible noncentral KKT contact; the propositions below convert those field crossings into an exact support identity. In both fixed and growing dimensions, we show that each superthreshold radial observation creates exactly one fitted atom and that quantitative gaps away from the edge exclude all others. Subsection 6.4 combines these identities with the radial Poisson limit, and Subsection 6.5 proves that the support identities hold uniformly along the variance path.

6.1 Fixed-dimensional likelihood localization

Fixed-dimensional likelihood localization is the first step in converting the field estimates into the exact support identity of Theorem 1.1. We apply the one-atom geometry around a fit that already carries a small amount of mass away from its central atom. The resulting comparison shows that edge mass cannot create new KKT contacts where the central directional field has a quantitative gap below one.

It is convenient to work in score coordinates. Replace a component location u by $t = \sqrt{1-\varepsilon_n}\,u$; this bijection preserves support size. We reuse G for the transformed mixing measure. Recall that $\gamma_n = (1-\varepsilon_n)^{-1} = 1 + 2\kappa_n$. If the central atom is at τ, the likelihood ratio of the atom at $\tau + a$ to the atom at τ, evaluated at Z_i, is

$$q_{n,\tau,a}(i) = \exp\left\{a \cdot (Z_i - \gamma_n \tau) - \frac{\gamma_n |a|^2}{2}\right\}.$$

For a perturbation $G = (1-W)\delta_\tau + G^{\text{out}}$, write

$$m_G(i) = 1 - W + \int q_{n,\tau,u-\tau}(i)\, G^{\text{out}}(du)$$

for its likelihood ratio to δ_τ, and define

$$D_{n,\tau}(a) = \frac{1}{n}\sum_{i=1}^{n} q_{n,\tau,a}(i), \qquad \Psi_G(\tau + a) = \frac{1}{n}\sum_{i=1}^{n} \frac{q_{n,\tau,a}(i)}{m_G(i)}.$$

The KKT conditions are $\Psi_G \le 1$, with equality at every support point.

The deterministic no-propagation estimate below separates the total mass of the edge atoms from the geometry of their locations.

Lemma 6.1. *For every measure $G = (1-W)\delta_\tau + G^{\text{out}}$ defined above,*

$$\Psi_G(\tau + a) \le \frac{D_{n,\tau}(a)}{1-W}, \qquad a \in \mathbb{R}^d.$$

Consequently, if $D_{n,\tau} \le 1 - \eta$ on a set A and $W \le \eta/2$, then G has no support point in $\tau + A$.

Proof. Every likelihood ratio is nonnegative, so $m_G(i) \ge 1 - W$. Substitution in the definition of Ψ_G gives the first inequality, and $(1-\eta)/(1-\eta/2) < 1$ gives the second. □

An NPMLE places almost all its mass near the true score origin. The following coarse estimate isolates a central component before the dimension-specific edge analysis. For a mixing measure G in score coordinates, put

$$m_G(z) = \int \exp\left\{t \cdot z - \frac{\gamma_n |t|^2}{2}\right\} G(dt), \qquad \mathfrak{D}_n(G) = \int \{1 - e^{-\kappa_n |t|^2}\}\, G(dt).$$

If P is standard Gaussian measure, then $Pm_G = 1 - \mathfrak{D}_n(G)$, and hence

$$P \log m_G \le \log Pm_G = \log\{1 - \mathfrak{D}_n(G)\} \le -\mathfrak{D}_n(G). \tag{6.1}$$

The population loss yields a uniform empirical bound whose coarse polylogarithmic precision is already sharper than the critical KKT gaps used in the fixed-dimensional support argument.

Lemma 6.2. *For each fixed d, there is $C_d < \infty$ such that, uniformly over $0 \le \varepsilon_n < 1$,*

$$\mathfrak{D}_n(\widehat{G}_n) = O_{\mathbb{P}}\left\{\frac{(\log n)^{C_d}}{n}\right\}$$

for every NPMLE $\widehat{G}_n$.

Proof. It suffices to optimize over mixing measures supported in $\mathrm{conv}\{Z_1/\gamma_n, \dots, Z_n/\gamma_n\}$. Indeed, if p is the Euclidean projection of t onto this convex hull, then

$$(t-p)\cdot Z_i - \frac{\gamma_n}{2}(|t|^2 - |p|^2) = (t-p)\cdot(Z_i - \gamma_n p) - \frac{\gamma_n}{2}|t-p|^2 \le 0,$$

so moving t to p weakly increases every fitted density. With probability tending to one this hull lies in $B(0, C\sqrt{\log n})$.

For $\mathcal{M}(R) = \{m_G : \mathrm{supp}(G) \subset B(0,R)\}$, let $N_{[]}(\eta, \mathcal{M}(R), L^1(P))$ be the least number of brackets $[l,u] = \{f : l \le f \le u\}$, with $P(u-l) \le \eta$, needed to cover $\mathcal{M}(R)$. We use the compact Gaussian-mixture entropy bound from [Ano26a, Section 2]; see also [GVDV01, Section 3]:

$$\log N_{[]}\{\eta, \mathcal{M}(R), L^1(P)\} \le C_d(1+R)^d\{1 + \log(1/\eta)\}^{d+1}. \tag{6.2}$$

The constant is uniform in κ_n, since

$$m_G(z)\phi_d(z) = \int e^{-\kappa_n |t|^2}\phi_d(z-t)\, G(dt),$$

and an $L^1(P)$ bracket for m_G is an ordinary L^1 bracket for a Gaussian location mixture of total mass at most one.

Write $\mathbb{P}_n f = n^{-1}\sum_{i=1}^n f(Z_i)$. Fix a dyadic shell $\delta < \mathfrak{D}_n(G) \le 2\delta$, and cover it by upper brackets u of $L^1(P)$-width $\delta/4$. If $m_G \le u$, then $Pu \le 1 - 3\delta/4$, and Markov's inequality gives

$$\mathbb{P}\{\mathbb{P}_n \log m_G \ge -\delta/4,\ m_G \le u\} \le e^{n\delta/4}(Pu)^n \le e^{-n\delta/2}.$$

For $R = O(\{\log n\}^{1/2})$, the logarithm of the number of brackets is at most $(\log n)^{C_d}$ whenever $\delta \ge n^{-2}$. A union bound over brackets and dyadic shells therefore shows that every measure with $\mathfrak{D}_n(G) > C(\log n)^{C_d}/n$ has negative likelihood gain relative to δ_0, with probability tending to one. Such a measure cannot be an NPMLE, proving the claim. ◻

The defect bound does not rule out several nearby atoms. The local collapse lemma shows that every maximizer carries all central mass on one atom, with the remaining edge mass smaller than the KKT gaps used later.

Lemma 6.3. *Assume either $d = 2$ and $\lambda_n \to \lambda \in (0, \infty)$, or fixed $d \ge 3$ with $s_n^\star$ converging in $\mathbb{R}$. For a sufficiently small fixed $r_0 > 0$, every NPMLE can, with probability tending to one, be written*

$$\widehat{G}_n = (1 - W_n)\delta_{\tau_n} + G_n^{\mathrm{out}}, \qquad \operatorname{supp}(G_n^{\mathrm{out}}) \subset B(0, r_0)^c, \tag{6.3}$$

where

$$W_n + |\tau_n|^2 = O_{\mathbb{P}}\left\{\frac{(\log n)^{C_d}}{n\kappa_n}\right\}.$$

In all these regimes, $W_n = o_{\mathbb{P}}(\kappa_n)$.

Proof. Lemma 6.2 and $1 - e^{-\kappa_n |t|^2} \ge c\kappa_n \min\{|t|^2, 1\}$ imply the displayed bounds for the mass outside $B(0, r_0)$ and for the second moment of the restriction to that ball. Normalize the central restriction to a probability measure G^{cen}, and let τ and V be its mean and covariance. Cauchy–Schwarz gives the asserted bound for $|\tau|^2$.

We show that replacing G^{cen} by δ_τ strictly increases the likelihood unless $V = 0$. Keep the outside part fixed and interpolate

$$G_u = (1 - W)\{(1 - u)\delta_\tau + uG^{\mathrm{cen}}\} + G^{\mathrm{out}}, \qquad 0 \le u \le 1.$$

By concavity it is enough to show that the derivative at $u = 0$ is negative. With $h = t - \tau$, put

$$g_i(h) = \exp\left\{h \cdot (Z_i - \gamma_n\tau) - \frac{\gamma_n |h|^2}{2}\right\}, \qquad Q_i^{\mathrm{out}} = \int q_{n,\tau,t-\tau}(i)\, G^{\mathrm{out}}(dt).$$

The likelihood ratio under G_0 is $m_i^0 = 1 - W + Q_i^{\mathrm{out}}$. Define $\omega_i = (1 - W)/m_i^0$. The derivative is

$$\sum_{i=1}^n \omega_i \int \{g_i(h) - 1\}\, G^{\mathrm{cen}}(dt). \tag{6.4}$$

First replace ω_i by one. Let $\widehat{F}_\tau(h) = n^{-1}\sum_i g_i(h)$. Uniformly for $|\tau| = o_{\mathbb{P}}(\sqrt{\kappa_n})$ and $|h| \le 2r_0$, a compact finite-dimensional empirical-process bound after two differentiations gives

$$\sup_{|h| \le 2r_0} \|\nabla^2 \widehat{F}_\tau(h) - \nabla^2 F_\tau(h)\|_{\mathrm{op}} = O_{\mathbb{P}}(n^{-1/2}) = o_{\mathbb{P}}(\kappa_n).$$

Here the differentiated class has an integrable envelope of the form $C(1 + |Z|^2)e^{2r_0|Z|}$, and the population average is

$$F_\tau(h) = \exp\{-\kappa_n |h|^2 - \gamma_n \tau \cdot h\}.$$

Indeed,

$$\nabla^2 F_\tau(h) = F_\tau(h)\left[(2\kappa_n h + \gamma_n\tau)(2\kappa_n h + \gamma_n\tau)^\top - 2\kappa_n I_d\right].$$

The bound on $|\tau|^2$, together with $\kappa_n \gtrsim 1/\log n$, gives $|\tau|^2 = o_{\mathbb{P}}(\kappa_n)$. Thus, after decreasing the fixed r_0 if needed, $\nabla^2 \widehat{F}_\tau(h) \preceq -c\kappa_n I_d$ throughout the ball, with probability tending to one. Taylor's theorem and $\int h\, G^{\mathrm{cen}}(dt) = 0$ now show that the unweighted version of (6.4) is at most $-cn\kappa_n \operatorname{tr} V$.

It remains to restore the weights. Let m_i^1 be the likelihood ratio under the original NPMLE G_1, and let

$$r_i^1 = \frac{\int q_{n,\tau,t-\tau}(i)\, G^{\mathrm{out}}(dt)}{m_i^1}$$

be its posterior responsibility for the outside component. Integrating the componentwise KKT equality over G^{out} gives $\sum_i r_i^1 = nW$.

We next compare the two sets of weights explicitly. Put $\bar{g}_i = \int g_i(h)\, G^{\text{cen}}(dt)$, so that $m_i^1 = (1-W)\bar{g}_i + Q_i^{\text{out}}$. On the event $R_n := \max_i |Z_i| \le C\sqrt{\log n}$,

$$e^{-K_n} \le \bar{g}_i \le e^{K_n}, \qquad K_n = C_{r_0}(1+R_n),$$

uniformly in i. Consequently $m_i^1/m_i^0 \le e^{K_n}$, and hence

$$1-\omega_i = \frac{Q_i^{\text{out}}}{m_i^0} \le e^{K_n} r_i^1, \qquad \sum_i (1-\omega_i) \le e^{K_n} nW.$$

For each i, another second-order expansion, again using $\int h\, G^{\text{cen}}(dt) = 0$, gives

$$\left| \int \{g_i(h) - 1\}\, G^{\text{cen}}(dt) \right| \le C_{r_0}(1+|Z_i|^2) e^{C_{r_0}|Z_i|} \operatorname{tr} V.$$

It follows that restoring the weights changes (6.4) by at most

$$C_{r_0} nW(1+\log n) e^{C_{r_0}\sqrt{\log n}} \operatorname{tr} V.$$

Relative to $n\kappa_n \operatorname{tr} V$, the last expression is

$$O_{\mathbb{P}}\left\{ \frac{(\log n)^{C_d+1} e^{C_{r_0}\sqrt{\log n}}}{n\kappa_n^2} \right\} = o_{\mathbb{P}}(1),$$

since $\kappa_n \gtrsim 1/\log n$ in both stated critical windows. The derivative is therefore negative whenever $V \neq 0$, contradicting the maximality of G_1. Hence $V = 0$, which proves (6.3). The same rate and $\kappa_n \gtrsim 1/\log n$ give $W = o_{\mathbb{P}}(\kappa_n)$. □

6.2 Fixed-dimensional support recovery

This subsection proves the fixed-dimensional support identity used in both parts of Theorem 1.1. Exposed radial points are separated by much more than the $1/\rho_n$ width of a likelihood patch, so their local optimizations decouple: a superthreshold patch contains one fitted atom and a subthreshold patch contains none. Here $\rho_n = \sqrt{2\log n}$ for $d = 2$ and $\rho_n = \sqrt{2b_{n,d}}$ for fixed $d \ge 3$, while τ denotes the central score location from Lemma 6.3. Put $\Delta_n^{\text{anc}} = \gamma_n \tau - \bar{Z}_n$. The exact anchor identity is

$$D_{n,\tau}(a) = \frac{1}{n}\sum_i q_{n,\tau,a}(i) = e^{-a\cdot\Delta_n^{\text{anc}}} D_{n,*}(a), \qquad D_{n,*}(a) = \frac{1}{n}\sum_i e^{a\cdot(Z_i-\bar{Z}_n)-\gamma_n|a|^2/2}. \tag{6.5}$$

The central location is tied to the sample mean by an exact barycenter identity. For a score-coordinate NPMLE G, let

$$\Pi_i(dt) = \frac{\exp\{t\cdot Z_i - \gamma_n|t|^2/2\}\, G(dt)}{m_G(Z_i)}$$

be the posterior mixing distribution for observation i. KKT equality G-almost everywhere gives

$$\frac{1}{n}\sum_{i=1}^n \Pi_i(dt) = G(dt).$$

On the other hand, differentiating the likelihood when every score location is translated by the same vector gives

$$0 = \frac{1}{n}\sum_{i=1}^{n}\left\{Z_i - \gamma_n \int t\,\Pi_i(dt)\right\}.$$

Consequently,

$$\bar{Z}_n = \gamma_n \int t\,G(dt). \tag{6.6}$$

For the decomposition in Lemma 6.3, this implies

$$\Delta_n^{\mathrm{anc}} = \gamma_n \int (\tau - t)\,G_n^{\mathrm{out}}(dt).$$

Support reduction places G in $\operatorname{conv}\{Z_1/\gamma_n, \dots, Z_n/\gamma_n\}$. Since $\max_i |Z_i| = O_{\mathbb{P}}(\rho_n)$ in fixed dimension, the rate in Lemma 6.3 therefore gives

$$\rho_n|\Delta_n^{\mathrm{anc}}| = O_{\mathbb{P}}(W_n\rho_n^2) = o_{\mathbb{P}}(\kappa_n) \tag{6.7}$$

in both fixed-dimensional critical windows. For the last equality, use

$$\rho_n^2 = O(\log n), \qquad W_n = O_{\mathbb{P}}\left(\frac{(\log n)^{C_d}}{n\kappa_n}\right), \qquad \kappa_n \gtrsim \frac{1}{\log n}.$$

We use one symbol for the dimension-specific radial excess:

$$x_{n,i} = \begin{cases} |Z_i|^2/2 - \log n, & d = 2, \\ |Z_i - \bar{Z}_n|^2/2 - b_{n,d}, & d \geq 3. \end{cases} \tag{6.8}$$

In fixed dimension, replacing the centered radius in the second line by $|Z_i|$ changes all excesses by $o_{\mathbb{P}}(1)$, uniformly in i.

We now fix the local coordinate maps used below. In $d = 2$, write $U_j = Z_j/|Z_j|$, let R_θ denote planar rotation through angle θ, and set

$$a_{n,j}^{(2)}(r,z) = (\rho_n + r)\mathsf{R}_{z/\rho_n}U_j, \qquad (r,z) \in \mathbb{R}^2. \tag{6.9}$$

For fixed $d \geq 3$, set

$$a_{n,j}^{(\geq 3)}(h) = a_{n,j}^* + h, \qquad a_{n,j}^* = \frac{Z_j - \gamma_n\tau}{\gamma_n}, \qquad h \in \mathbb{R}^d. \tag{6.10}$$

If ν_j is a measure in the corresponding local coordinates, its pushforward under $a_{n,j}^{(d)}$ is a measure on score displacements. Below we write

$$\nu = \sum_j (a_{n,j}^{(d)})_\# \nu_j \tag{6.11}$$

for the combined displacement measure.

For a finite positive measure ν on displacement space, with $m = \nu(\mathbb{R}^d) \leq n$, consider the anchored mixture

$$G_{n,\tau,\nu} = \left(1 - \frac{m}{n}\right)\delta_\tau + \frac{1}{n}(\tau + \cdot)_\# \nu.$$

Its exact likelihood gain over δ_τ is

$$\mathcal{L}_{n,\tau}(\nu) = \sum_{i=1}^{n}\log\left\{1 + F_{n,i}(\nu) - \frac{m}{n}\right\}, \qquad F_{n,i}(\nu) = \frac{1}{n}\int q_{n,\tau,a}(i)\,\nu(da). \tag{6.12}$$

Thus $m = nW$ is the natural rescaled edge mass.

6.2.1 Local likelihood programs

The field limits in Sections 3 and 4.3 determine the likelihood program associated with one radial extreme. Optimizing this program shows that an active patch contains one atom rather than a split mixing measure.

Suppose first that $d \geq 3$, along a sequence on which $s_n \longrightarrow s$. For an observation of centered radial excess x, completing the square at the displacement $a = a_{n,i}^{(\geq 3)}(h)$ gives

$$\frac{1}{n} q_{n,\tau,a}(i) \longrightarrow A_x(h) := e^{x-s-|h|^2/2}.$$

If a finite positive measure ν of rescaled mass is placed in this patch, its limiting gain is

$$\mathcal{J}_x^{(\geq 3)}(\nu) = \log\left\{1 + \int A_x(h)\,\nu(dh)\right\} - \nu(\mathbb{R}^d). \tag{6.13}$$

For fixed total mass, the integral is uniquely maximized by putting all mass at $h = 0$. The remaining scalar problem is

$$\max_{y \geq 0} \{\log(1 + e^{x-s} y) - y\},$$

whose optimizer is $y_x^* = (1 - e^{-(x-s)})_+$. Hence the patch is active exactly when $x > s$, and every nonzero limiting optimizer is the one-atom measure $y_x^* \delta_0$.

In dimension two, use radial score offset r and blown-up angular displacement z, and write

$$B_\lambda(r) = e^{-\lambda}\Phi(-r), \qquad C_{\lambda,x}(r,z) = e^{-\lambda} e^{x-r^2/2-z^2/2}.$$

The corresponding local program is

$$\mathcal{J}_x^{(2)}(\nu) = \log\left\{1 + \int C_{\lambda,x}\,d\nu\right\} - \int \{1 - B_\lambda(r)\}\,d\nu(r,z). \tag{6.14}$$

Put

$$R_x(r,z) = \frac{C_{\lambda,x}(r,z)}{1 - B_\lambda(r)}.$$

If $u = \int C_{\lambda,x}\,d\nu$, then

$$\int (1 - B_\lambda)\,d\nu \geq \frac{u}{\sup_{r,z} R_x(r,z)},$$

with equality only when ν is supported on the maximizers of R_x. The angular maximizer is uniquely $z = 0$. The radial maximizer is uniquely $r = -a_\lambda$, where

$$e^\lambda = \Phi(a_\lambda) + \frac{\phi(a_\lambda)}{a_\lambda},$$

and the calculation in Section 3 gives $\sup R_x = e^{x-x_\lambda}$. Thus the patch is active exactly when $x > x_\lambda$, and every nonzero limiting optimizer is a single atom at $(-a_\lambda, 0)$. Its rescaled mass is

$$q_x^* = \frac{1 - e^{-(x-x_\lambda)}}{1 - e^{-\lambda}\Phi(a_\lambda)}.$$

For fixed $d \geq 3$, the contact argument requires differentiated control near a random radial extreme. The next lemma supplies that control without conditioning on a sample-centered exceedance event. It first fixes the raw extreme, so the punctured sample remains independent, and then sums over the possible distinguished observations by the Palm formula.

Lemma 6.4. *Fix $d \geq 3$ and suppose $s_n = O(1)$. Put*

$$\widetilde{x}_{n,i} = \frac{|Z_i|^2}{2} - b_{n,d}, \qquad \mathcal{E}_n(C) = \{j : |\widetilde{x}_{n,j}| \leq C\}.$$

For fixed $A, C, R < \infty$, define, for $j \in \mathcal{E}_n(C)$,

$$\mathcal{R}_{n,j,A}(h) = \frac{1}{n} \sum_{i \neq j} \exp\left\{ \left(\frac{Z_j}{\gamma_n} + h \right) \cdot Z_i - \frac{\gamma_n}{2} \left| \frac{Z_j}{\gamma_n} + h \right|^2 \right\} \mathbf{1}_{\{\widetilde{x}_{n,i} \leq -A\}}, \qquad h \in B_R.$$

Then

$$\max_{j \in \mathcal{E}_n(C)} \|\mathcal{R}_{n,j,A}\|_{C^2(B_R)} \xrightarrow{\mathbb{P}} 0. \tag{6.15}$$

Moreover, if $K_{n,j,i}(h)$ denotes the exponential summand in the definition of $\mathcal{R}_{n,j,A}$, then

$$\max_{j \in \mathcal{E}_n(C)} \max_{|\beta| \leq 2} \sup_{h \in B_R} \frac{1}{n} \sum_{i \neq j} |\partial_h^\beta K_{n,j,i}(h)| \mathbf{1}_{\{\widetilde{x}_{n,i} \leq -A\}} \xrightarrow{\mathbb{P}} 0. \tag{6.16}$$

The same conclusions hold with the candidate condition $|x_{n,j}| \leq C$ and lower-cloud cutoff $x_{n,i} \leq -A$, where $x_{n,i}$ is the sample-centered excess in (6.8).

Proof. Let $R_{n,A}^2/2 = b_{n,d} - A$, choose $p > d$, and fix a deterministic z with $\left||z|^2/2 - b_{n,d}\right| \leq C$. For $t_z(h) = z/\gamma_n + h$, set

$$V_{n,z,i}(h) = \frac{1}{n} e^{t_z(h) \cdot Z_i - \gamma_n |t_z(h)|^2/2} \mathbf{1}_{\{|Z_i| \leq R_{n,A}\}}.$$

We first establish, uniformly over these z, $h \in B_R$, $q \in \{2, p\}$, and $|\beta| \leq 3$,

$$\sum_{i=2}^{n} \mathbb{E} |\partial_h^\beta V_{n,z,i}(h)|^q \leq C_{A,C,R,p,d} \rho_n^{-1} e^{-\lambda_n}. \tag{6.17}$$

Indeed, differentiation inserts a polynomial of degree at most $|\beta|$ in $Z_i - \gamma_n t_z(h)$. Exponential tilting therefore gives

$$\begin{aligned} &\sum_{i=2}^{n} \mathbb{E} |\partial_h^\beta V_{n,z,i}(h)|^q \\ &\qquad \leq C n^{1-q} e^{-q\kappa_n |t_z(h)|^2 + (q^2 - q)|t_z(h)|^2/2} \mathbb{E}\left[\{1 + |G + (q - \gamma_n) t_z(h)|\}^{q|\beta|} \mathbf{1}_{\{|G + q t_z(h)| \leq R_{n,A}\}} \right]. \end{aligned}$$

Aligning $t_z(h) = r e_1$, conditioning on $G_\perp$, and applying Mills' bound to G_1 show that the last expectation is at most

$$C_{A,C,R,p,d} \rho_n^{-1} \exp\{-(qr - R_{n,A})^2/2\}.$$

Here the polynomial causes only a constant loss: on the truncation event its longitudinal argument is $R_{n,A} - \gamma_n r + O(1/\rho_n)$, which is bounded for $|r - |z|/\gamma_n| \leq R$, while the transverse Gaussian moments are fixed. Equivalently, this weighted Mills estimate follows by integrating $(1 + |u|)^m e^{-u^2/2}$ beyond the same one-dimensional boundary.

The exponent outside ρ_n^{-1} equals

$$\begin{aligned} &(1 - q) \log n - q\kappa_n r^2 + \frac{q^2 - q}{2} r^2 - \frac{1}{2}(qr - R_{n,A})^2 \\ &\qquad = (q - 1)\{b_{n,d} - \log n - A\} - q\kappa_n r^2 - \frac{q}{2}(r - R_{n,A})^2. \end{aligned}$$

Since $b_{n,d} - \log n = \lambda_n - s_n$, $\kappa_n r^2 = \lambda_n + o(1)$, and $s_n = O(1)$, it is at most $-\lambda_n + O_{A,C,R,p,d}(1)$. This proves (6.17). The same tilting calculation with $q = 1$ gives

$$\sup_{z,h,|\beta|\le 3} \left| \sum_{i=2}^{n} \mathbb{E}\partial_h^\beta V_{n,z,i}(h) \right| \le C_{A,C,R,d} e^{-\lambda_n} = o(1). \tag{6.18}$$

The same $q = 1$ calculation applies to the absolute values of the derivatives. Using the derivatives of order three and the fundamental theorem of calculus on a fixed mesh of B_R gives

$$\sup_z \mathbb{E}\left[\max_{|\beta|\le 2} \sup_{h\in B_R} \frac{1}{n} \sum_{i=2}^{n} |\partial_h^\beta K_{n,z,i}(h)| \mathbf{1}_{\{|Z_i|\le R_{n,A}\}} \right] \le C_{A,C,R,d} e^{-\lambda_n} = o(1).$$

The Palm union bound below and Markov's inequality therefore also prove (6.16).

Apply Rosenthal's inequality to the centered derivatives and integrate over B_R. Equations (6.17) and (6.18) yield, uniformly in z,

$$\mathbb{E}\left\| \sum_{i=2}^{n} V_{n,z,i} \right\|_{W^{3,p}(B_R)}^p = o(1).$$

Because $p > d$, the embedding $W^{3,p}(B_R) \hookrightarrow C^2(B_R)$ and Markov's inequality imply

$$\sup_{z:\,||z|^2/2 - b_{n,d}|\le C} \mathbb{P}\left\{ \left\| \sum_{i=2}^{n} V_{n,z,i} \right\|_{C^2(B_R)} > \delta \right\} = o(1)$$

for every $\delta > 0$. Independence now gives the Palm union bound

$$\begin{aligned} &\mathbb{P}\left\{ \max_{j\in\mathcal{E}_n(C)} \|\mathcal{R}_{n,j,A}\|_{C^2(B_R)} > \delta \right\} \\ &\quad \le n \int_{\{||z|^2/2 - b_{n,d}|\le C\}} \mathbb{P}\left\{ \left\| \sum_{i=2}^{n} V_{n,z,i} \right\|_{C^2(B_R)} > \delta \right\} P(dz) = o(1), \end{aligned}$$

because the Gamma tail estimate gives $nP\{||Z|^2/2 - b_{n,d}| \le C\} = O_C(1)$.

Finally,

$$\max_i \left| \frac{|Z_i - \bar{Z}_n|^2 - |Z_i|^2}{2} \right| = o_{\mathbb{P}}(1).$$

Fix $\delta > 0$. Apart from an event of probability tending to zero, $|x_{n,j}| \le C$ implies $|\widetilde{x}_{n,j}| \le C + \delta$, while $x_{n,i} \le -A$ implies $\widetilde{x}_{n,i} \le -(A - \delta)$. Applying the raw result with candidate bound $C + \delta$ and lower cutoff $A - \delta$ proves the sample-centered assertion directly. □

6.2.2 Uniform patch stability

We need the local approximations uniformly over the random edge patches and all candidate mixing measures. As in Sections 4.4 and 3, the lower radial cutoff is removed after the finite-sample limit.

The candidate patches and the observational cutoff play different roles. The candidate set is fixed by the activation threshold and remains unchanged as the cutoff is removed. The cutoff is used only to divide the data into a finite exposed cloud and a lower cloud. The following estimate controls the lower cloud uniformly over every mixing measure on the candidate patches.

Lemma 6.5. *Fix $C, R, M, J < \infty$. Let $\mathcal{I}_n$ be any set of at most J candidate observations with $|x_{n,j}| \le C$, and let ν_j be supported in B_R for $d \ge 3$, or in $[-R, R]^2$ for $d = 2$, with $m = \sum_{j \in \mathcal{I}_n} \nu_j(\mathbb{R}^d) \le M$. Form ν by (6.11) and write $F_{n,i} = F_{n,i}(\nu)$. Suppose $s_n = O(1)$ for fixed $d \ge 3$, or $\lambda_n \to \lambda \in (0, \infty)$ for $d = 2$.*

There is $A_0 = A_0(C, R, M) > C + 1$ such that, for each fixed $A \ge A_0$, on the event $\max_i x_{n,i} \le C$ and the usual edge-separation event,

$$\max_{i: x_{n,i} \le -A} F_{n,i} \le C_{C,R} M e^{-A}, \qquad \sum_{i: x_{n,i} \le -A} F_{n,i}^2 \le C_{C,R} M^2 e^{-A} + o_{\mathbb{P},A}(1), \tag{6.19}$$

apart from an event of probability tending to zero. For every candidate patch, one also has the derivative-envelope bound

$$\max_{|\beta| \le 2} \sup_u \frac{1}{n} \sum_{i: x_{n,i} \le -A} \left| \partial_u^\beta q_{n,\tau,a_{n,j}(u)}(i) \right| = O_{\mathbb{P},A_0,C,R}(1), \tag{6.20}$$

uniformly for $A \ge A_0$. Uniformly over the same measures,

$$\sum_{i: x_{n,i} \le -A} F_{n,i} = \begin{cases} o_{\mathbb{P},A}(1), & d \ge 3, \\ \displaystyle\sum_{j \in \mathcal{I}_n} \int B_{\lambda_n}(r)\, \nu_j(dr, dz) + O_{\mathbb{P},C,R,J}(M e^{-A}) + o_{\mathbb{P},A}(1), & d = 2, \end{cases} \tag{6.21}$$

where

$$B_{\lambda_n}(r) = e^{-\lambda_n} \Phi(-r).$$

The observations with $-A < x_{n,i} \le C$ that do not belong to $\mathcal{I}_n$ contribute $o_{\mathbb{P},A}(1)$, uniformly over the candidate patches and measures.

Proof. Completing the square gives, for every score displacement a,

$$\frac{1}{n} q_{n,\tau,a}(i) = \exp\left\{ \frac{|Z_i - \gamma_n \tau|^2}{2\gamma_n} - \log n - \frac{\gamma_n}{2} \left| a - \frac{Z_i - \gamma_n \tau}{\gamma_n} \right|^2 \right\}. \tag{6.22}$$

For $d \ge 3$, the anchor estimate (6.7) and $b_{n,d} - \log n = \lambda_n - s_n$ make the first exponent at most $-A - s_n + o_{\mathbb{P}}(1)$ when $x_{n,i} \le -A$. For $d = 2$, the radial identity in Lemma 6.8 bounds the exponent by $-A + O_{C,R}(1)$. Integrating (6.22) over a measure of mass at most M proves the first inequality in (6.19).

For fixed $d \ge 3$, Lemma 6.4, followed by the translation from Z_j/γ_n to $(Z_j - \gamma_n \tau)/\gamma_n$, gives

$$\max_{j \in \mathcal{I}_n} \sup_u \frac{1}{n} \sum_{i: x_{n,i} \le -A} q_{n,\tau,a_{n,j}^{(\ge 3)}(u)}(i) = o_{\mathbb{P},A}(1).$$

The translation is legitimate in C^2 because $|\tau| = o_{\mathbb{P}}(1)$ and the Palm lemma controls a slightly larger compact set. The translated Palm field is multiplied by $e^{-\gamma_n a \cdot \tau}$; uniformly on these patches its logarithm is $O_{\mathbb{P}}(\rho_n |\tau|) = o_{\mathbb{P}}(1)$ by Lemma 6.3. Positivity and Fubini now give the first line of (6.21). Equation (6.16) gives (6.20) in this case. The bound is uniform for $A \ge A_0$, because the absolute derivative sum is nonincreasing as the lower cutoff $-A$ is moved inward.

For $d = 2$, use the fixed-cutoff decomposition Lemma 3.8, the quantitative absorption bound in the proof of Lemma 3.7, and the anchor multiplier

$$e^{-a \cdot \Delta_n^{\text{anc}} - a \cdot \bar{Z}_n} = e^{-\gamma_n a \cdot \tau} = 1 + o_{\mathbb{P}}(1)$$

uniformly on the candidate patches. Together, these three estimates show that the lower field in a patch with radial coordinate r is $B_{\lambda_n}(r) + O_{\mathbb{P},C,R}(e^{-A}) + o_{\mathbb{P},A}(1)$, uniformly in its direction. Integrating over ν proves the second line of (6.21). The radial identity bounds the absolute value of each derivative by a fixed polynomial times the Gaussian angular kernel in Lemma 6.8. For the deep part $\xi_i < -\ell_n^{\mathrm{edge}}$, repeat the boxwise Rosenthal–Morrey argument in Lemma 6.8 with a polynomial majorant for the absolute derivatives; its mean is uniformly bounded by exponential tilting. For the strip $-\ell_n^{\mathrm{edge}} \le \xi_i \le -A_0$, the slab-and-arc estimate in that lemma, with the summable weight $(1+q^4)e^{-cq^2}$, is $O_{\mathbb{P},C,R}(e^{-A_0})$. Their sum proves (6.20) for $d = 2$. Since the absolute sum over $\{x_{n,i} \le -A\}$ decreases with A, the $O_{\mathbb{P},C,R}(e^{-A_0})$ bound holds simultaneously for all $A \ge A_0$.

The linear bound is $O_{\mathbb{P}}(M)$ in either dimension. Therefore positivity and the first inequality in (6.19) give

$$\sum_{i:x_{n,i}\le -A} F_{n,i}^2 \le \left(\max_{i:x_{n,i}\le -A} F_{n,i}\right) \sum_{i:x_{n,i}\le -A} F_{n,i} \le C_{C,R} M^2 e^{-A} + o_{\mathbb{P},A}(1),$$

which is the second inequality in (6.19).

Finally, for fixed A, C, the exposed cloud $\{-A < x_{n,i} \le C\}$ has tight cardinality. Its distinct directions are separated by $\ell_n^{\mathrm{sep}}/\rho_n$, with $\ell_n^{\mathrm{sep}} \to \infty$, apart from an event of probability tending to zero. Completing the square and differentiating on fixed patches then show that the response of every nonassociated exposed observation is $O(e^{-c(\ell_n^{\mathrm{sep}})^2})$. Summing over the tight exposed cloud proves the last assertion. □

Lemma 6.5 makes the lower-cloud contribution uniform over all mixing measures on finitely many candidate patches. The next proposition uses its linear, quadratic, and derivative bounds to reduce the exact likelihood to a finite collection of local programs.

Proposition 6.6. *Fix $C, R, M, J < \infty$, and let $\mathcal{I}_n$ be any set of at most J candidate observations with $|x_{n,j}| \le C$. Work on the event $\max_i x_{n,i} \le C$ and the edge-separation event. For $d \ge 3$, let ν_j be supported in $B_R \subset \mathbb{R}^d$; for $d = 2$, let it be supported in $[-R, R]^2$. Form the displacement measure using only $j \in \mathcal{I}_n$, assume $\sum_{j\in\mathcal{I}_n} \nu_j(\mathbb{R}^d) \le M$, and suppose the anchor satisfies (6.7). Assume that $s_n = O(1)$ for fixed $d \ge 3$, or that $\lambda_n \to \lambda \in (0,\infty)$ for $d = 2$, and define*

$$\mathcal{J}_n^{(\ge 3)}(\nu) = \sum_{j\in\mathcal{I}_n} \log\left\{1 + \int e^{x_{n,j}-s_n-|h|^2/2}\,\nu_j(dh)\right\} - \sum_{j\in\mathcal{I}_n} \nu_j(\mathbb{R}^d),$$

$$B_{\lambda_n}(r) = e^{-\lambda_n}\Phi(-r), \qquad C_{n,x}(r,z) = e^{-\lambda_n} e^{x-r^2/2-z^2/2},$$

$$\mathcal{J}_n^{(2)}(\nu) = \sum_{j\in\mathcal{I}_n} \log\left\{1 + \int C_{n,x_{n,j}}\,d\nu_j\right\} - \sum_{j\in\mathcal{I}_n} \int \{1 - B_{\lambda_n}(r)\}\,d\nu_j.$$

There is $A_0 = A_0(C,R,M) < \infty$ such that, for every fixed $A \ge A_0$, uniformly over these measures,

$$\sup_\nu |\mathcal{L}_{n,\tau}(\nu) - \mathcal{J}_n^{(d)}(\nu)| = O_{\mathbb{P},C,R,J}\{(M+M^2)e^{-A}\} + o_{\mathbb{P},A}(1), \tag{6.23}$$

where the random coefficient in $O_{\mathbb{P},C,R,J}(1)$ is tight uniformly for $A \ge A_0$. In particular, for every $\delta > 0$,

$$\lim_{A\to\infty} \limsup_{n\to\infty} \mathbb{P}\left\{\sup_\nu |\mathcal{L}_{n,\tau}(\nu) - \mathcal{J}_n^{(d)}(\nu)| > \delta\right\} = 0. \tag{6.24}$$

The candidate space and $\mathcal{J}_n^{(d)}$ do not depend on A. After $n \to \infty$ and then $A \to \infty$, the limiting functional is the sum of the independent local programs (6.13) or (6.14) over the candidate patches.

Proof. Put $m = \nu(\mathbb{R}^d)$, $v_i = F_{n,i}(\nu) - m/n$, and split the data into the candidate observations, the noncandidate exposed cloud $\{-A < x_{n,i} \le C\} \setminus \mathcal{I}_n$, and the lower cloud $\{x_{n,i} \le -A\}$. For a candidate observation j, completing the square gives, uniformly on its associated patch,

$$\frac{1}{n} q_{n,\tau,a^{(\ge 3)}_{n,j}(h)}(j) = e^{x_{n,j} - s_n - |h|^2/2 + o(1)}$$

in $d \ge 3$, and gives $C_{n,x_{n,j}}(r,z) + o(1)$ in $d = 2$. The separation event makes every cross-patch response $o(1)$. Hence the candidate observations contribute the logarithmic terms in $\mathcal{J}_n^{(\ge 3)}$ or $\mathcal{J}_n^{(2)}$, uniformly over ν, up to $o_{\mathbb{P},A}(1)$. The final assertion of Lemma 6.5 shows that the noncandidate exposed cloud contributes $o_{\mathbb{P},A}(1)$.

Choose A_0 so that $C_{C,R} M e^{-A_0} \le 1/4$, and then take $n \ge 4M$. On the lower cloud, the elementary inequality

$$|\log(1+v) - v| \le 2v^2, \qquad v \ge -M/n, \quad |v| \le 1/2,$$

and Lemma 6.5 give

$$\sum_{i : x_{n,i} \le -A} \{\log(1 + v_i) - v_i\} = O_{\mathbb{P},C,R,J}(M^2 e^{-A}) + o_{\mathbb{P},A}(1).$$

Since the exposed cloud has $O_{\mathbb{P},A}(1)$ observations,

$$\sum_{i : x_{n,i} \le -A} v_i = \sum_{i : x_{n,i} \le -A} F_{n,i} - m + o_{\mathbb{P},A}(1).$$

Equation (6.21) therefore supplies the penalty $-m$ in $d \ge 3$, and the penalty $-\sum_j \int \{1 - B_{\lambda_n}(r)\}\, d\nu_j$ in $d = 2$. Lemma 6.5 already includes the anchor error. Combining the candidate observations, the noncandidate exposed cloud, and the lower cloud proves (6.23). □

An elementary coercivity estimate bounds the edge mass without assuming a fixed number of fitted atoms; it depends only on the candidate data points, not on the mixing complexity within their patches.

Proposition 6.7. *Restrict (6.12) to J edge patches, and let $\mathcal{I}_n$ be their associated observations. Suppose throughout the patches that*

$$\frac{1}{n} \sum_{i \notin \mathcal{I}_n} q_{n,\tau,a}(i) \le b < 1, \qquad \max_{i \in \mathcal{I}_n} \frac{1}{n} q_{n,\tau,a}(i) \le C_0.$$

If $m = nW$ is the total rescaled edge mass, then, for all large n,

$$\mathcal{L}_{n,\tau}(\nu) \le -c_b m + J \log(1 + C_0 m), \qquad c_b = (1-b)/3.$$

Consequently, every maximizer satisfies $m \le 4C_0 J^2 / c_b^2$.

Proof. Fubini's theorem gives $\sum_{i \notin \mathcal{I}_n} F_{n,i}(\nu) \le bm$. Jensen's inequality then implies

$$\sum_{i \notin \mathcal{I}_n} \log\{1 - m/n + F_{n,i}(\nu)\} \le (n - J) \log\left\{1 - \frac{m}{n} + \frac{bm}{n - J}\right\} \le -c_b m.$$

At each retained observation the likelihood ratio is at most $1 + C_0 m$, which proves the displayed bound. A maximizer has nonnegative gain, and $\log(1+x) \le 2\sqrt{x}$ then yields $m \le 4C_0 J^2 / c_b^2$. □

The two-dimensional contact argument requires differentiated, rather than only uniform, control of the inward-edge background. The next lemma upgrades the fixed-cutoff decomposition to the C^2 form needed below.

Lemma 6.8. *Assume $d = 2$ and $\lambda_n \to \lambda \in (0, \infty)$, and put $\ell_n^{\mathrm{edge}} = 4 \log \rho_n$. For $v \in S^1$, define*

$$a_{n,v}(r, z) = (\rho_n + r)\mathsf{R}_{z/\rho_n} v$$

and, for fixed $R < \infty$, let

$$\mathcal{B}_{n,v}(r, z) = \frac{1}{n} e^{-\kappa_n |a_{n,v}(r,z)|^2} \sum_{i:\, \xi_i < -\ell_n^{\mathrm{edge}}} e^{a_{n,v}(r,z)\cdot Z_i - |a_{n,v}(r,z)|^2/2},$$

$$\mathcal{A}_{n,v,A}(r, z) = \frac{1}{n} e^{-\kappa_n |a_{n,v}(r,z)|^2} \sum_{i:\, -\ell_n^{\mathrm{edge}} \le \xi_i \le -A} e^{a_{n,v}(r,z)\cdot Z_i - |a_{n,v}(r,z)|^2/2}.$$

Then

$$\sup_{v \in S^1} \|\mathcal{B}_{n,v} - e^{-\lambda}\Phi(-r)\|_{C^2([-R,R]^2)} \xrightarrow{\mathbb{P}} 0, \tag{6.25}$$

and, for every $\delta > 0$,

$$\lim_{A\to\infty} \limsup_{n\to\infty} \mathbb{P}\left\{ \sup_{v \in S^1} \|\mathcal{A}_{n,v,A}\|_{C^2([-R,R]^2)} > \delta \right\} = 0. \tag{6.26}$$

Proof. Write $a = a_{n,v}(r, z)$. The radial identity gives

$$\frac{1}{n} e^{-\kappa_n |a|^2} e^{a\cdot Z_i - |a|^2/2} = e^{\xi_i - \kappa_n |a|^2 - |Z_i - a|^2/2}.$$

Cover S^1 by $O(\rho_n)$ arcs of angular length $O(1/\rho_n)$, and on an arc centered at v_α write $v = \mathsf{R}_{y/\rho_n} v_\alpha$, with y in a fixed interval. On the resulting fixed boxes in (y, r, z), differentiation inserts a polynomial of bounded degree in $Z_i - \gamma_n a$, with coefficients bounded uniformly in n. Consequently, for every multi-index $|\beta| \le 3$, a summand of the corresponding derivative of $\mathcal{B}_{n,v}$ is bounded by $C_R e^{-\ell_n^{\mathrm{edge}}}$. Exponential tilting, as in the proof of Proposition 3.6, also gives

$$\sup_{\alpha,y,r,z} \sum_i \mathbb{E} \left| \partial^\beta \left[\frac{1}{n} e^{-\kappa_n |a|^2} e^{a\cdot Z_i - |a|^2/2} \mathbf{1}_{\{\xi_i < -\ell_n^{\mathrm{edge}}\}} \right] \right|^2 \le \frac{C_R}{\rho_n}.$$

Here ∂^β denotes derivatives in the scaled local variables (y, r, z), and the supremum ranges over their fixed coordinate boxes. The corresponding sum of pth moments is at most $C_R e^{-(p-2)\ell_n^{\mathrm{edge}}}/\rho_n$. For any fixed $p > 3$, Rosenthal's inequality therefore bounds the expected pth power of the $W^{3,p}$ norm on one box by $C_R \rho_n^{-p/2} + o(\rho_n^{-p/2})$. Summing over the $O(\rho_n)$ boxes still gives $o(1)$. The Sobolev–Morrey embedding $W^{3,p} \hookrightarrow C^2$ on each fixed box, followed by Markov's inequality, proves uniform C^2 concentration over all $v \in S^1$.

The mean is rotation invariant and exponential tilting gives

$$\mathbb{E}\mathcal{B}_{n,v}(r, z) = e^{-\kappa_n(\rho_n+r)^2} \mathbb{P}\{|G + (\rho_n + r)e_1|^2 \le \rho_n^2 - 2\ell_n^{\mathrm{edge}}\}, \qquad G \sim \mathcal{N}(0, I_2).$$

The boundary expansion in the proof of Proposition 3.6 is uniform after two derivatives in r; the mean has no z-dependence. Since $\ell_n^{\mathrm{edge}}/\rho_n \to 0$ and $\kappa_n \rho_n^2 \to \lambda$, the last display converges in C^2 to $e^{-\lambda}\Phi(-r)$. This proves (6.25).

For the strip, let $Y_i(v) = \rho_n \operatorname{ang}(U_i, v)$. Uniformly on the same compact set and for $|\beta| \le 2$, the radial identity and the elementary bound of a polynomial times a Gaussian give

$$|\partial^\beta_{r,z}\mathcal{A}_{n,v,A}(r,z)| \le C_R \sum_{i:\, -\ell_n^{\mathrm{edge}} \le \xi_i \le -A} e^{\xi_i}\{1 + |Y_i(v) - z|^4\}e^{-c|Y_i(v)-z|^2}.$$

Taking the supremum over $z \in [-R, R]$ only changes the constants and the Gaussian exponent. The slab-and-arc occupancy proof of Lemma 3.7 applies with the summable weight $(1 + q^4)e^{-cq^2}$ in place of $e^{-q^2/2}$. It bounds the last display by $C_R e^{-A} + o_{\mathbb{P}}(1)$, simultaneously for $|\beta| \le 2$, and proves (6.26). □

To convert each local maximizer into an exact contact count, we need two-derivative convergence of the fitted KKT certificate; convergence of likelihood values alone would not exclude nearby split atoms.

Proposition 6.9. *Suppose an NPMLE is localized to finitely many compact patches, its rescaled edge mass is bounded, and every associated radial excess is at least $\eta > 0$ from the relevant threshold. After the radial cutoffs are removed, an inactive patch contains no support point and an active patch contains exactly one, apart from an event of probability tending to zero. Here active means $x_{n,j} > s_n$ for $d \ge 3$ and $x_{n,j} > x_{\lambda_n}$ for $d = 2$.*

Proof. It is enough to argue along subsequences. The radial cutoffs and threshold buffer allow a further subsequence on which the number of candidate patches is constant and all marks $x_{n,j}$ converge, say to x_j. Write $a_{n,j}(u) = a^{(\ge 3)}_{n,j}(u)$ in $d \ge 3$, and $a_{n,j}(u) = a^{(2)}_{n,j}(r,z)$ for $u = (r,z)$ in $d = 2$. The candidate index set is fixed throughout this subsequence; only the lower observational cutoff A will later vary. For a localized measure $\nu = (\nu_1, \dots, \nu_J)$, let

$$M_{n,i}(\nu) = 1 + \frac{1}{n}\sum_j \int \{q_{n,\tau,a_{n,j}(u)}(i) - 1\}\,\nu_j(du).$$

Moving mass from the central atom into patch j has derivative

$$\Gamma_{n,j,\nu}(u) = \frac{1}{n}\sum_{i=1}^n \frac{q_{n,\tau,a_{n,j}(u)}(i) - 1}{M_{n,i}(\nu)}. \tag{6.27}$$

It is nonpositive throughout the patch and vanishes at every support point.

The limiting certificate in an active $d \ge 3$ patch is

$$\Gamma^*_j(h) = e^{-|h|^2/2} - 1;$$

it has a unique zero at zero and Hessian $-I_d$. In an active two-dimensional patch it is

$$\Gamma^*_j(r,z) = (1 - B_\lambda(r))\left\{\frac{R_{x_j}(r,z)}{\sup_{u,y} R_{x_j}(u,y)} - 1\right\},$$

whose unique zero is $(-a_\lambda, 0)$ and whose Hessian there is $-\{1 - e^{-\lambda}\Phi(a_\lambda)\}I_2$. A threshold buffer η also gives a fixed negative certificate gap in every inactive patch.

We next verify that (6.27) converges to these limits in C^2 on a smaller patch, uniformly over measures in a fixed compact mass ball. Fix the observational cutoff A. At the candidate observations, completing the square and differentiating three times gives uniform C^3 convergence of the associated response. Every response from a different exposed observation carries the factor $e^{-c(\ell_n^{\mathrm{sep}})^2}$, and the exposed cloud has tight cardinality.

For the lower observations in fixed $d \geq 3$, Lemma 6.4 gives, simultaneously in all candidate patches, C^2 convergence of the punctured lower field to zero. The lemma is centered at Z_j/γ_n, whereas the score-displacement patch is centered at $Z_j/\gamma_n - \tau$. Taking the Palm estimate on a slightly larger compact set and using $|\tau| = o_{\mathbb{P}}(1)$ transfers the bound to the score-displacement patch. The accompanying multiplier $e^{-\gamma_n a\cdot\tau}$ is $1 + o_{\mathbb{P}}(1)$ in C^2, because $\rho_n|\tau| = o_{\mathbb{P}}(1)$. Thus this is a deterministic translation of an already uniform C^2 estimate and requires no conditioning on a sample-centered exceedance set.

In $d = 2$, Lemma 6.8 gives C^2 convergence of the diffuse background to $e^{-\lambda}\Phi(-r)$ and removes the strip between the receding and fixed radial cutoffs in the C^2 topology. The estimate is uniform over the patch direction, so it remains valid for the random directions U_j. The KKT field differs from the field in Lemma 6.8 by the anchor multiplier

$$e^{-a\cdot\Delta_n^{\mathrm{anc}} - a\cdot\bar{Z}_n} = e^{-\gamma_n a\cdot\tau}.$$

Lemma 6.3 gives

$$\rho_n^2|\tau|^2 = O_{\mathbb{P}}\left\{\frac{(\log n)^{C_d+1}}{n\kappa_n}\right\} = o_{\mathbb{P}}(1),$$

and differentiation of the multiplier in the local coordinates only introduces further factors $O_{\mathbb{P}}(|\tau|)$. It is therefore $1 + o_{\mathbb{P}}(1)$ in C^2 on every fixed patch.

It remains to pass from the unfitted fields to the fitted certificate. Take $A \geq A_0(C, R, M)$ from Lemma 6.5. For every lower observation,

$$|M_{n,i}(\nu) - 1| \leq F_{n,i}(\nu) + M/n \leq C_{C,R}Me^{-A} + o(1) \leq \frac{1}{2},$$

uniformly over measures of mass at most M. Hence $|M_{n,i}^{-1} - 1| \leq 2|M_{n,i} - 1|$. Applying Lemma 6.5 after each of the at most two local differentiations gives

$$\sup_{\nu:\,\nu(\mathbb{R}^d)\leq M}\left\|\frac{1}{n}\sum_{i:x_{n,i}\leq -A}\{q_{n,\tau,a_{n,j(\cdot)}}(i) - 1\}\{M_{n,i}(\nu)^{-1} - 1\}\right\|_{C^2} = O_{\mathbb{P},C,R,J}(Me^{-A}) + o_{\mathbb{P},A}(1), \tag{6.28}$$

with a tight $O_{\mathbb{P}}(1)$ coefficient uniform for $A \geq A_0$. To see this directly, note that $M_{n,i}(\nu)$ is fixed when the prospective contact location u is varied. Thus all derivatives fall on the numerator, and (6.20), together with $\max_i |M_{n,i}^{-1} - 1| \leq CMe^{-A} + o(1)$, gives the displayed bound. The constant -1 at derivative order zero is controlled by the same maximum. At a candidate observation, the denominator converges to its local-program denominator. At a noncandidate exposed observation, every candidate response is $o_{\mathbb{P},A}(1)$, uniformly in C^2; the constant -1 in the numerator contributes only $O_{\mathbb{P},A}(1/n)$ after summing over the tight exposed cloud. Thus the complete fitted certificate converges sequentially, first as $n \to \infty$ and then as $A \to \infty$, to the displayed limiting certificates.

We now make the order of the cutoff and argmax limits explicit. Proposition 6.7 confines the total rescaled mass to a deterministic compact interval on events of arbitrarily high probability. Together with the fixed compact candidate-patch union, this gives a weakly compact space $\mathcal{M}_{M,R}$ of candidate measures. Let $\mathcal{J}_{\boldsymbol{x}}^\star$ be the sum of the limiting local programs for a vector $\boldsymbol{x} = (x_1, \ldots, x_J)$ of candidate marks, and let $\nu_{\boldsymbol{x}}^\star$ be its optimizer. The scalar optimizations above show that this optimizer is unique and depends continuously on $\boldsymbol{x}$ as long as the marks remain outside the threshold buffer.

Choose the patch radius R so that every such optimizer is in the interior. For a fixed small $\delta > 0$, let $\mathcal{U}_{\boldsymbol{x}}$ be its radius-δ weak neighborhood in $\mathcal{M}_{M,R}$. The mark vectors with $|x_j| \leq C$ and distance at least η from the threshold form a compact set. Joint continuity of the local programs and uniqueness of their optimizers therefore give the uniform separation

$$c_{\delta,\eta,C,M,R,J} := \inf_{\boldsymbol{x}} \left[\mathcal{J}_{\boldsymbol{x}}^{\star}(\nu_{\boldsymbol{x}}^{\star}) - \sup_{\nu \in \mathcal{M}_{M,R} \setminus \mathcal{U}_{\boldsymbol{x}}} \mathcal{J}_{\boldsymbol{x}}^{\star}(\nu) \right] > 0. \tag{6.29}$$

Decrease δ, if necessary, so that every active patch has mass bounded away from zero throughout $\mathcal{U}_{\boldsymbol{x}}$. All the objects in (6.29) depend only on the threshold-buffered candidate patches, not on the observational cutoff. Choose the deterministic cutoff $A \ge A_0(C, R, M)$ so large that, by (6.24), the finite-patch approximation error exceeds one quarter of the gap in (6.29) with arbitrarily small limiting probability. Then take n large. Uniform convergence of the local kernels shows on the complementary event that every finite-sample maximizer belongs to $\mathcal{U}_{\boldsymbol{x}}$: otherwise comparison with the pushforward of $\nu_{\boldsymbol{x}}^{\star}$ would lose at least half that gap. In particular, every active finite-sample patch carries positive mass and hence contains a KKT contact.

Increase this same deterministic A, if necessary, so that (6.28) exceeds one quarter of either the inactive-patch certificate gap or the absolute value of the largest active limiting Hessian eigenvalue with arbitrarily small limiting probability. These constants are uniform over the threshold-buffered compact mark set, while the candidate space does not change. An inactive patch then has no contact. In an active patch, C^0 convergence confines all contacts to a small convex neighborhood of the unique limiting zero, and C^2 convergence makes the finite-sample certificate strictly concave there. Since the certificate is nonpositive, strict concavity rules out two distinct zeros. Sending the two exceptional probabilities to zero proves the claim along every subsequence. □

6.2.3 Threshold pruning and the support law

Only observations within a fixed radial distance of the activation threshold can support atoms. The needed deleted-field estimate is a further consequence of the Bennett–Morrey argument in Section 4.4.

Lemma 6.10. *Assume $d \ge 3$ and $s_n = O(1)$. For fixed $a > 0$, define*

$$S_t^{(a)} = \frac{1}{n} e^{-\kappa_n |t|^2} \sum_{i=1}^{n} e^{t \cdot Z_i - |t|^2/2} \mathbf{1}_{\{|Z_i|^2/2 \le \log n + \lambda_n - a\}}.$$

If $\zeta \in (0,1)$ and $e^a \zeta > 1$, then

$$\mathbb{P}\left\{ \sup_{|t| \ge \rho_n/2} S_t^{(a)} \ge \zeta \right\} \to 0.$$

Proof. Put $R_a^2/2 = \log n + \lambda_n - a$ and $h = R_a - |t|$. On each fixed physical annulus $|h| \le H$, Lemma 4.5 gives

$$\mathbb{P}\{S_t^{(a)} \ge \zeta\} \le \left(\rho_n^{-1} e^{-\lambda_n}\right)^{e^{a+h^2/2}\zeta - o(1)}.$$

Lemma 4.4 reduces the annulus to a net of cardinality $\rho_n^{d-1+o(1)}$, with $o(1)$ oscillation. Since $\rho_n^{-1} e^{-\lambda_n} = \rho_n^{-(d-1)+o(1)}$ and $e^a \zeta > 1$, the union bound tends to zero.

For $h \ge H$, Lemma 4.6 gives

$$\mathbb{P}\{S_t^{(a)} \ge \zeta/2\} \le \exp\{-c_d e^{a+h^2/4} \log \rho_n\}.$$

A polynomial-size Euclidean net and the same derivative control make the union bound vanish after H is large. When $|t| \ge R_a + H$, every truncated summand decreases along its radial ray, so the outward region reduces to the controlled boundary. These ranges cover $|t| \ge \rho_n/2$. □

The two-dimensional analogue of the deleted-field estimate has an explicit profile gap. It is useful to record both the global residual bound and the stronger bound on the candidate patches, where the associated extreme observation has itself been deleted.

Lemma 6.11. *Assume $d = 2$ and $\lambda_n \to \lambda \in (0, \infty)$. Fix $\eta > 0$, put*

$$\mathcal{I}_n(\eta) = \{i : \xi_i > x_{\lambda_n} - \eta\}, \qquad \mathcal{R}_{n,\eta}(a) = \frac{1}{n} \sum_{i \notin \mathcal{I}_n(\eta)} q_{n,\tau,a}(i),$$

and let $\mathcal{P}_n(\eta, R)$ be the union of the radius-R local patches associated with $\mathcal{I}_n(\eta)$. Then

$$\sup_{|a| \geq \rho_n/2} \mathcal{R}_{n,\eta}(a) \leq 1 - \frac{1}{2}(1 - e^{-\eta})(1 - e^{-\lambda}) + o_{\mathbb{P}}(1), \tag{6.30}$$

$$\sup_{a \in \mathcal{P}_n(\eta,R)} \mathcal{R}_{n,\eta}(a) \leq e^{-\lambda} + o_{\mathbb{P},R}(1) \tag{6.31}$$

for each fixed R.

Proof. Fix $C, K, A < \infty$ and work first on $\max_i \xi_i \leq C$ and $\big||a| - \rho_n\big| \leq K$. Write $a = (\rho_n + r)v$, $|r| \leq K$. Lemma 3.8, its quantitative absorption bound, and the anchor estimate give, uniformly in r and v,

$$\mathcal{R}_{n,\eta}((\rho_n + r)v) = B_{\lambda_n}(r) + \sum_{-A < \xi_i \leq x_{\lambda_n} - \eta} C_{\lambda_n,\xi_i}(r, Y_i(v)) + O_{\mathbb{P},K}(e^{-A}) + o_{\mathbb{P},A,K}(1). \tag{6.32}$$

For fixed A, C, the exposed points have tight cardinality and, with probability tending to one, their pairwise scaled angular distances exceed a deterministic $\ell_n \to \infty$. If there are at most J exposed points, then for any v all but at most one have $|Y_i(v)| \geq \ell_n/2$; uniformly for $|r| \leq K$, their total contribution is at most $Je^{C+K^2/2}e^{-\ell_n^2/8}$. Tightness of J therefore makes this remainder $o_{\mathbb{P},A,C,K}(1)$, uniformly in v. Since every retained residual mark is at most $x_{\lambda_n} - \eta$,

$$\begin{aligned} B_{\lambda_n}(r) + C_{\lambda_n, x_{\lambda_n} - \eta}(r, z) &= e^{-\eta}\{B_{\lambda_n}(r) + C_{\lambda_n, x_{\lambda_n}}(r, z)\} + (1 - e^{-\eta})B_{\lambda_n}(r) \\ &\leq e^{-\eta} + (1 - e^{-\eta})e^{-\lambda_n} \\ &= 1 - (1 - e^{-\eta})(1 - e^{-\lambda_n}). \end{aligned}$$

Choose A so that the absorption error is less than one quarter of the limiting gap and then take $n \to \infty$. This proves the first bound on every fixed radial window, on $\{\max_i \xi_i \leq C\}$. As $|r| \to \infty$, the spike term is bounded by $e^{C - r^2/2 + o(1)}$, while $B_{\lambda_n}(r)$ tends to $e^{-\lambda}$ inward and to zero outward. The diffuse- and tail-field estimates in the proof of Proposition 3.2 make these bounds uniform as $K \to \infty$. Finally,

$$\limsup_{n \to \infty} \mathbb{P}\{\max_i \xi_i > C\} = 1 - \exp\{-e^{-C}\} \longrightarrow 0 \qquad (C \to \infty),$$

which removes the upper cutoff and proves (6.30).

If $a \in \mathcal{P}_n(\eta, R)$, its associated observation belongs to $\mathcal{I}_n(\eta)$ and is absent from the residual sum. Every exposed residual direction is then at scaled angular distance at least $\ell_n - R$, so the entire exposed sum in (6.32) is $o_{\mathbb{P},A,C,R}(1)$. The remaining profile is at most $\sup_r B_{\lambda_n}(r) = e^{-\lambda_n}$. Taking first $n \to \infty$, then $A, C \to \infty$, proves (6.31). ∎

Combining the edge and bulk gaps prunes subthreshold points before the mass-coercivity bound is used. Thus the number of candidate patches is tight independently of the lower radial cutoff.

Proposition 6.12. *Fix $\eta > 0$, and set*

$$\vartheta_n = \begin{cases} s_n, & d \geq 3, \\ x_{\lambda_n}, & d = 2. \end{cases}$$

In the corresponding critical window, every noncentral support point of an NPMLE lies, with probability tending to one after the radial cutoffs are removed, in a patch of some deterministic radius $R = R(\eta) < \infty$ associated with an observation whose radial excess satisfies $x_{n,i} > \vartheta_n - \eta$. The number of these patches is $O_{\mathbb{P}}(1)$. More precisely, put

$$\mathcal{I}_n(\eta) = \{i : x_{n,i} > \vartheta_n - \eta\},$$

and let $\mathcal{P}_n(\eta)$ denote the union of their radius-R patches. There are a deterministic $b < 1$ and random variables $C_{0,n} = O_{\mathbb{P}}(1)$ such that

$$\sup_{a \in \mathcal{P}_n(\eta)} \frac{1}{n} \sum_{i : x_{n,i} \leq \vartheta_n - \eta} q_{n,\tau,a}(i) \leq b, \qquad \sup_{a \in \mathcal{P}_n(\eta)} \max_{i : x_{n,i} > \vartheta_n - \eta} \frac{1}{n} q_{n,\tau,a}(i) \leq C_{0,n}.$$

Proof. Call the sum over $i \notin \mathcal{I}_n(\eta)$ the residual field. For $d \geq 3$, every residual observation satisfies

$$|Z_i - \bar{Z}_n|^2/2 \leq b_{n,d} + s_n - \eta = \log n + \lambda_n - \eta.$$

Uniformly in i,

$$\frac{|Z_i|^2}{2} \leq \frac{|Z_i - \bar{Z}_n|^2}{2} + \max_i |Z_i - \bar{Z}_n| \, |\bar{Z}_n| + \frac{|\bar{Z}_n|^2}{2} = \frac{|Z_i - \bar{Z}_n|^2}{2} + o_{\mathbb{P}}(1).$$

Thus, with probability tending to one, every residual observation lies below $\log n + \lambda_n - \eta/2$. Put

$$a_\eta = \eta/2, \qquad \zeta_\eta = \frac{1 + e^{-a_\eta}}{2}.$$

Since $e^{a_\eta} \zeta_\eta > 1$, Lemma 6.10 applied with $a = a_\eta$ gives a gap of $(1 - \zeta_\eta)/2$ below one throughout $|a| \geq \rho_n/2$, with probability tending to one. To pass from the raw field in Lemma 6.10 to the residual field, note that

$$\frac{1}{n} \sum_{\text{residual}} q_{n,\tau,a}(i) \leq e^{-a \cdot \Delta_n^{\text{anc}} - a \cdot \bar{Z}_n} S_a^{(a_\eta)}.$$

An NPMLE may be supported in the convex hull of the rescaled observations, so $|a| = O_{\mathbb{P}}(\rho_n)$; on this range (6.7) and $\rho_n |\bar{Z}_n| = o_{\mathbb{P}}(1)$ make the prefactor $1 + o_{\mathbb{P}}(1)$. Thus the residual field is at most $(1 + \zeta_\eta)/2 < 1$ throughout the $d \geq 3$ edge region.

In $d = 2$, Lemma 6.11 gives the explicit global edge gap

$$g_{2,\eta} = \frac{1}{2}(1 - e^{-\eta})(1 - e^{-\lambda}) > 0.$$

On the candidate patches themselves it gives the stronger residual bound $e^{-\lambda} + o_{\mathbb{P}}(1)$. Thus, in the first displayed inequality of the proposition, one may take any fixed $b \in (e^{-\lambda}, 1)$ for all sufficiently large n.

The bulk estimates in Proposition 4.10 for $d \geq 3$, and Lemma 3.5 for $d = 2$, retain a gap of order κ_n on the region between the fixed core and $\rho_n/2$. The candidate set has tight cardinality by radial Poisson convergence, and its maximum radial excess is tight. On an event of arbitrarily high probability, both its cardinality and its normalized peak heights are bounded by deterministic constants. The exact square completion (6.22) then bounds the total candidate contribution outside radius R of the patch centers by Ce^{-cR^2}. Choose R so this is smaller than one quarter of the relevant residual edge gap: $(1 - \zeta_\eta)/2$ for $d \geq 3$, and $g_{2,\eta}$ for $d = 2$. The full central field consequently has a fixed gap outside $\mathcal{P}_n(\eta)$ throughout the edge region. Equation (6.7) transfers both gaps to the central anchor. Finally, Lemma 6.1 preserves the edge gap because $W_n = o_{\mathbb{P}}(1)$ and the bulk gap because $W_n = o_{\mathbb{P}}(\kappa_n)$. KKT equality therefore excludes every support point outside the candidate patches.

On the fixed candidate patches, Lemma 6.10 for $d \geq 3$ and Lemma 6.11 for $d = 2$ give the first displayed inequality with deterministic $b < 1$. The tight upper radial cutoff, square completion, and angular separation give the second with $C_{0,n} = O_{\mathbb{P}}(1)$. □

We can now count the patches. The following buffered sandwich is the fixed-dimensional input to the common edge-count reduction.

Proposition 6.13. *Assume $\lambda_n \to \lambda \in (0, \infty)$ when $d = 2$, or $s_n^\star \to s \in \mathbb{R}$ when $d \geq 3$ is fixed. With $x_{n,i}$ defined in (6.8) and ϑ_n as in Proposition 6.12, for every fixed $\eta > 0$,*

$$\#\{i : x_{n,i} > \vartheta_n + \eta\} \leq \widehat{K}_n - 1 \leq \#\{i : x_{n,i} > \vartheta_n - \eta\}, \tag{6.33}$$

apart from an event whose limiting probability is $O(\eta)$. Under either critical-window assumption, the NPMLE is unique with probability tending to one.

Proof. First assume $\lambda_n \to \lambda \in (0, \infty)$ when $d = 2$, or $s_n^\star \to s \in \mathbb{R}$ when $d \geq 3$. Fix $\eta > 0$. Proposition 6.12 confines every noncentral support point to the $O_{\mathbb{P}}(1)$ patches associated with $x_{n,i} > \vartheta_n - \eta$, and supplies the two hypotheses of Proposition 6.7. Hence the rescaled edge mass is $O_{\mathbb{P}}(1)$. After choosing an upper radial cutoff and a deterministic bound on the number of patches, Proposition 6.6 and Proposition 6.9 show that every active patch contains exactly one fitted atom and every inactive candidate patch contains none. These are precisely the inequalities in (6.33). Removing the radial cutoffs costs $o(1)$; the only remaining exceptional event is a radial point in the threshold buffer, whose probability is $O(\eta)$.

On the event where Propositions 6.12 and 6.9 give the stated patch localization, the NPMLE is also unique. If G and H are two NPMLEs, then $(G + H)/2$ is also an NPMLE. Lemma 6.3, applied to all three measures, forces their central atoms to coincide: otherwise the central part of $(G + H)/2$ would have positive covariance. Propositions 6.12 and 6.9, applied to $(G + H)/2$, likewise force the edge contact in every active patch to coincide. Thus all NPMLEs are supported on the same finite contact set.

The likelihood vectors at these contacts are linearly independent. Indeed, on the rows indexed by the active observations, the edge-contact block has diagonal entries of order n, bounded away from zero after division by n, and off-diagonal entries $o(n)$. The finite-patch estimates also give an $O_{\mathbb{P}}(1)$ empirical average for each of the finitely many edge columns. Hence there is a noncandidate row on which every edge entry is $o(n)$; adjoining this row and the central column produces, after scaling the edge columns by n^{-1}, an asymptotically invertible triangular minor. Strict concavity of the log likelihood makes the fitted likelihood vector common to all maximizers, and linear independence then makes its representing weights unique. The probability of the exceptional event is $O(\eta) + o(1)$; letting $\eta \downarrow 0$ proves the uniqueness claim. □

6.3 Growing-dimensional support recovery

This subsection proves the support identity for every growing-dimensional regime in Theorem 1.1. In the Gaussian-bump coordinates of Subsection 2.1, the buffered reductions below identify every noncentral fitted atom with one radial exceedance while retaining the exact centering correction.

6.3.1 Exact edge coordinates

Recall the exact radial coordinate $\xi_{n,i}^\star$, scale $\delta_n^\star$, and identity (2.14) from Section 2.2. For fixed $C, J < \infty$ and $\eta > 0$, recall the buffered event $\mathcal{B}_n(C, J, \eta)$ from (6.97). The centered radial Poisson limit makes the complement of such buffered events negligible after $C, J \to \infty$ and then $\eta \downarrow 0$. It therefore remains to prove that every positive buffered excess creates exactly one noncentral support point.

When $d_nL_n = o(n^2)$, the exact coordinate agrees with the $2\log n$-based coordinate used in Sections 4 and 4.5. The next elementary comparison lets us invoke the radial estimates from those sections without carrying two normalizations through each proof.

Lemma 6.14. *Suppose $d_nL_n = o(n^2)$ and $s_n^\star = O(1)$. With s_n, $\xi_{n,i}$, and $\delta_n = a_n/\gamma_n$ as in (4.1) and (4.4),*

$$s_n - s_n^\star = o(1), \qquad \frac{\delta_n^\star}{\delta_n} = 1 + o(1).$$

Moreover, for every fixed $C < \infty$,

$$\max_{i:\,|\xi_{n,i}^\star|\le C} \left|\xi_{n,i}^\star - (\xi_{n,i} - s_n)\right| \longrightarrow_{\mathbb{P}} 0. \tag{6.34}$$

Proof. The second assertion is immediate from $\delta_n^\star = \sigma_n^2 a_n/\gamma_n$. The definitions give

$$s_n - s_n^\star = \frac{\gamma_n}{a_n}\left(L_n - \frac{R_n^\star}{\sigma_n^2}\right).$$

Since $L_n - R_n^\star = O(n^{-1})$ and $1-\sigma_n^2 = n^{-1}$, its absolute value is $O\{b_n/(na_n) + \gamma_n/(na_n)\}$. This tends to zero: it is $O(L_n/n)$ when $d_n = O(L_n)$, while the superlogarithmic Gamma expansions make it $O(\sqrt{d_nL_n}/n) = o(1)$ otherwise.

In fact, the two excess coordinates satisfy the exact identity

$$\xi_{n,i} - s_n = \sigma_n^2\xi_{n,i}^\star - \frac{L_n - R_n^\star}{\delta_n}. \tag{6.35}$$

Now $L_n - R_n^\star = O(n^{-1})$. Also $(n\delta_n)^{-1} = o(1)$: this is immediate when $d_n = O(L_n)$, and in the superlogarithmic range it is $O\{\sqrt{d_n/L_n}/n\} = o(L_n^{-1})$. Equation (6.35), together with $\sigma_n^2 = 1 - n^{-1}$, proves (6.34). □

For later use, if $c_1, \dots, c_n > 0$, define

$$\Gamma_c(w) = \frac{1}{n}\sum_{i=1}^n c_i \exp\{w \cdot q_i - |w|^2/2\}.$$

The Gibbs variational formula gives the weighted entropy dual

$$\log \Gamma_c(w) = \sup_{p\in\Delta_n}\left\{w \cdot \sum_i p_i q_i - \frac{|w|^2}{2} - K(p\|\mathbf{u}_n) + \sum_i p_i \log c_i\right\}. \tag{6.36}$$

The entropy argument used for one-atomicity must be made stable under the fitted denominators before it can count contacts. The following weighted version supplies precisely that stability. It is stated only in the random entropy regime, where it will be used.

Lemma 6.15. *Suppose (5.1) holds, $s_n = O(1)$, and $d_nL_n = o(n^2)$. Put*

$$q_i = \frac{Z_i - \bar{Z}_n}{\sqrt{\gamma_n}}, \qquad R_i = \frac{|q_i|^2}{2}, \qquad \delta_n = \frac{a_n}{\gamma_n}, \qquad \mu_n^{\mathrm{off}} = \max_{i\ne j}|q_i \cdot q_j|, \qquad \mathfrak{k}_n = \frac{\|(q_i \cdot q_j)_{i,j\le n}\|_{\mathrm{op}}}{2n}.$$

Fix $C, J < \infty$ and $\eta > 0$, and suppose

$$\max_i (R_i - L_n)/\delta_n \le C, \qquad \#\{i : R_i > L_n - \eta\delta_n\} \le J, \qquad \min_i |R_i - L_n| > \eta\delta_n.$$

Let $A = \{i : R_i > L_n\}$. Suppose a random vector $\beta = (\beta_1, \dots, \beta_n)$ satisfies, uniformly on this event,

$$\beta_i = -(R_i - L_n) + e_{n,i} \quad (i \in A), \qquad \max_{i\in A} |e_{n,i}| + \max_{j\notin A} |\beta_j| =: \omega_n, \tag{6.37}$$

where

$$\omega_n = O_{\mathbb{P}}\left\{\frac{e^{\mu_n^{\mathrm{off}}}}{n} + \frac{\delta_n L_n}{n}\right\} = n^{-1+o_{\mathbb{P}}(1)} = o_{\mathbb{P}}(\delta_n). \tag{6.38}$$

If

$$F_\beta(p) = \frac{1}{2}\left|\sum_i p_i q_i\right|^2 - K(p\|\mathbf{u}_n) + \sum_i p_i\beta_i,$$

then, with probability tending to one, every $p \in \Delta_n$ satisfying $F_\beta(p) \ge 0$ has one of the following two forms:

$$\|p - \mathbf{u}_n\|_1 \le n^{-1/2+o_{\mathbb{P}}(1)}, \qquad \textit{or} \qquad 1 - p_j \le n^{-1/2+o_{\mathbb{P}}(1)} \quad \textit{for some } j \in A. \tag{6.39}$$

In the first case $|\sum_i p_i q_i| = o_{\mathbb{P}}(1)$; in the second, $|\sum_i p_i q_i - q_j| = o_{\mathbb{P}}(1)$. Moreover, if $p = p(w)$ is the Gibbs vector of a weighted certificate with value at least one, then w lies within $n^{-1/2+o_{\mathbb{P}}(1)}$ of zero in the first case and of q_j in the second.

Proof. Lemma 5.6 and (5.1) imply that the deterministic bounds in Proposition 5.5 hold with probability tending to one. Replace μ_n^{off} and υ_n by deterministic upper bounds that hold with probability tending to one and preserve their asymptotic separation. For readability, continue to denote these deterministic bounds by μ_n^{off} and υ_n. We may then choose deterministic a_n° so that

$$1 + \mu_n^{\mathrm{off}} \ll a_n^\circ \ll \min\{\log(1/\upsilon_n), \log(n\delta_n), L_n\},$$

on an event of probability tending to one, and then set

$$T_n = e^{a_n^\circ}, \qquad L_n y_n = \{a_n^\circ \log(n\delta_n)\}^{1/2}.$$

These choices satisfy (5.12). Since $\max_i R_i = L_n + O(\delta_n)$, the extra diagonal term in (5.15) is

$$O(\delta_n)\sum_i b_i^2 = o\left(L_n \sum_i b_i^2\right)$$

and is absorbed by the heavy-entropy term in (5.16).

First suppose $\max_i p_i \le 1 - y_n$. Repeat the light-heavy decomposition in Proposition 5.5. Explicitly, put

$$f_i = np_i, \qquad \ell_i = f_i \wedge T_n, \qquad h_i = (f_i - T_n)_+, \qquad b_i = \frac{h_i}{n}, \qquad K_0 = \frac{1}{n}\sum_i \phi(\ell_i), \qquad z = \sum_i b_i,$$

where $\phi(x) = x\log x - x + 1$. The active diagonal excess is canceled by the active fitted weight: because $b_i^2 \le p_i$,

$$\sum_{i\in A}(R_i - L_n)b_i^2 + \sum_i p_i\beta_i \le \omega_n.$$

Keeping half of the slack in (5.14) and (5.16) therefore gives

$$F_\beta(p) \le -cK_0 - ca_n^\circ z + \omega_n. \tag{6.40}$$

Thus $F_\beta(p) \ge 0$ implies $K_0 + a_n^\circ z = O_{\mathbb{P}}(\omega_n)$. Moreover,

$$\|p - \mathbf{u}_n\|_1 \le C\left\{\left(\frac{T_n\omega_n}{a_n^\circ}\right)^{1/2} + \frac{\omega_n}{a_n^\circ}\right\} \le n^{-1/2+o_{\mathbb{P}}(1)}.$$

The last inequality uses $T_n = n^{o(1)}$ and (6.38); in particular, $L_n(T_n\omega_n/a_n^\circ)^{1/2} = o_{\mathbb{P}}(1)$. The Gram bound then gives $|\sum_i p_i q_i| = o_{\mathbb{P}}(1)$, proving the first alternative.

It remains to consider $p_i = 1 - y > 1 - y_n$. Write $\sum_j p_j q_j = (1-y)q_i + y\bar{q}$, as in the last part of the proof of Proposition 5.5. Put

$$\mathfrak{H}_n = L_n - \mu_n^{\text{off}} - O(\delta_n) - 2L_n y_n - \omega_n.$$

The preceding choices give $\mathfrak{H}_n \sim L_n$ and $e^{-\mathfrak{H}_n} = n^{-1+o(1)} = o(\delta_n)$. Convexity, the entropy bound $h_{\text{bin}}(y) \le y\log(e/y)$, and the fitted weights give

$$F_\beta(p) \le \alpha_i + y\left\{\log\frac{e}{y} - \mathfrak{H}_n\right\}, \tag{6.41}$$

where $\alpha_i = R_i - L_n + \beta_i$.

If $i \notin A$, the first term on the right of (6.41) is at most $-\eta\delta_n + \omega_n$. The second term on the right of (6.41) is at most $e^{-\mathfrak{H}_n} = o(\delta_n)$. Hence $F_\beta(p) < 0$.

If $i \in A$, then $|\alpha_i| \le \omega_n$. For $y > e^{-\mathfrak{H}_n/2}$,

$$\log(e/y) - \mathfrak{H}_n \le -\mathfrak{H}_n/3$$

for all large n. Nonnegative free energy consequently forces

$$y \le e^{-\mathfrak{H}_n/2} + \frac{3\omega_n}{\mathfrak{H}_n} \le n^{-1/2+o_{\mathbb{P}}(1)}.$$

Since $\max_j |q_j| = O_{\mathbb{P}}(\sqrt{L_n})$, this gives

$$\left|\sum_j p_j q_j - q_i\right| = y|\bar{q} - q_i| = o_{\mathbb{P}}(1),$$

which is the second alternative in (6.39).

For later use, define the weighted certificate and its Gibbs vector by

$$\Gamma_\beta(w) = \frac{1}{n}\sum_{i=1}^n \exp\left\{w\cdot q_i - \frac{|w|^2}{2} + \beta_i\right\}, \qquad p_i(w) = \frac{\exp\{w\cdot q_i + \beta_i\}}{\sum_j \exp\{w\cdot q_j + \beta_j\}}.$$

The exact identity

$$F_\beta\{p(w)\} - \log\Gamma_\beta(w) = \frac{1}{2}\left|w - \sum_i p_i(w) q_i\right|^2$$

shows that the two alternatives in (6.39) also localize w whenever $\Gamma_\beta(w) \geq 1$. On the diffuse branch, (6.40) gives $F_\beta\{p(w)\} = n^{-1+o_{\mathbb{P}}(1)}$, while the Gram estimate and (6.39) give $|\sum_i p_i(w)q_i| = n^{-1/2+o_{\mathbb{P}}(1)}$. On an active branch, (6.41) gives $F_\beta\{p(w)\} = n^{-1+o_{\mathbb{P}}(1)}$, and $|\sum_i p_i(w)q_i - q_j| = n^{-1/2+o_{\mathbb{P}}(1)}$. The displayed identity therefore proves the asserted spatial localization. Moreover,

$$\nabla^2 \log \Gamma_\beta(w) = \operatorname{Cov}_{p(w)}(q_i) - I.$$

On the diffuse branch the covariance norm is at most $2\mathfrak{k}_n + O\{L_n\|p - \mathbf{u}_n\|_1\} = o_{\mathbb{P}}(1)$; on an active branch it is $O\{L_n(1-p_i)\} = o_{\mathbb{P}}(1)$. Thus every relevant patch is strictly log-concave for all large n. □

The slowly growing range needs a local version of the truncated shell argument. Conditional first moments are enough. Exponential tilting shows that the supremum of the background on a unit edge patch has mean $\exp\{-\varepsilon_n L_n + O(\sqrt{d_n} + \varepsilon_n\sqrt{L_n})\}$, which beats the number of upper centers even when d_n grows arbitrarily slowly.

Lemma 6.16. *Suppose*

$$d_n \longrightarrow \infty, \qquad d_n \leq (\log L_n)^2, \qquad s_n = O(1),$$

and put $A_n = \sqrt{d_n}$. There is a random set $\mathcal{U}_n$, containing every index with $\xi_{n,i} - s_n > -\eta$, such that, with probability tending to one,

$$|\mathcal{U}_n| \leq e^{2A_n}, \qquad \min_{\substack{i,j\in\mathcal{U}_n\\ i\neq j}} |q_i - q_j|^2 = 4L_n\{1+o(1)\},$$

and, for every fixed $R < \infty$ and $0 \leq k \leq 3$,

$$\max_{j\in\mathcal{U}_n} \sup_{|w-q_j|\leq R} \left\| \nabla^k \sum_{i\notin\mathcal{U}_n} h_i e^{-|w-q_i|^2/2} \right\| = o_{\mathbb{P}}(1). \tag{6.42}$$

Proof. Work first with the independent uncentered vectors $Q_i = Z_i/\sqrt{\gamma_n}$, and let $\mathcal{U}_n^\circ$ contain the indices whose uncentered radial coordinate is at least $-A_n - 1$. Write $h_i^\circ = n^{-1}e^{|Q_i|^2/2}$ for the corresponding uncentered bump height. The Gamma hazard estimate and Markov's inequality give

$$|\mathcal{U}_n^\circ| \leq e^{2A_n}$$

with probability tending to one. Conditional on the radii, their directions are independent and uniform. Since $\log|\mathcal{U}_n^\circ| = O(A_n) = o(d_n)$, spherical concentration gives the displayed separation, first for the Q_i's and then for the centered q_i's.

Conditional on Q_j, $j \in \mathcal{U}_n^\circ$, all remaining vectors are independent of Q_j. Write $w = Q_j + h$, $|h| \leq R$. For every multi-index $|\alpha| \leq 3$, exponential tilting from $\mathcal{N}(0, \gamma_n^{-1}I)$ to $\mathcal{N}(Q_j/\gamma_n, \gamma_n^{-1}I)$ gives

$$\begin{aligned}\sum_{i\neq j} \mathbb{E}\left[\sup_{|h|\leq R} \left| \partial^\alpha \{h_i^\circ e^{-|Q_j+h-Q_i|^2/2}\} \right| \mathbf{1}_{\{i\notin\mathcal{U}_n^\circ\}} \,\middle|\, Q_j \right] \\ \leq C_{\alpha,R} \exp\left\{ -\frac{\varepsilon_n |Q_j|^2}{2} + C_{\alpha,R}\left(\sqrt{d_n} + \varepsilon_n|Q_j| + 1\right) \right\}.\end{aligned}$$

Indeed, after tilting, $Q_i - Q_j$ is Gaussian with mean $-\varepsilon_n Q_j$, covariance $\gamma_n^{-1}I$, and the supremum over h contributes at most a fixed polynomial in its norm times $e^{R|Q_i-Q_j|}$. Uniformly over $j \in \mathcal{U}_n^\circ$, $|Q_j|^2/2 = L_n + O(A_n)$, while

$$\varepsilon_n L_n = (1 - o(1))\frac{d_n}{2}\log\frac{2L_n}{d_n}.$$

This dominates $A_n + \sqrt{d_n} + \varepsilon_n\sqrt{L_n}$, since $d_n \to \infty$ and $d_n \le (\log L_n)^2$. Conditional Markov's inequality is applied to the random set through

$$\mathbb{P}\{\exists j \in \mathcal{U}_n^\circ : B_{n,j}\} \le \sum_{j=1}^{n} \mathbb{E}\left[\mathbf{1}_{\{j\in\mathcal{U}_n^\circ\}}\mathbb{P}(B_{n,j} \mid Q_j)\right],$$

where $B_{n,j}$ is the failure of the displayed derivative bound. Since $\mathbb{E}|\mathcal{U}_n^\circ| = e^{(1+o(1))A_n}$, the conditional exponential moment bound in the preceding display makes this sum $o(1)$. Passing from coordinate derivatives to the tensor norm costs at most $O(d_n^k)$ for order $k \le 3$, whose logarithm is negligible compared with $\varepsilon_n L_n$. This proves the uncentered version of (6.42).

Finally,

$$B_n(w) = e^{-w\cdot\bar{Z}_n/\sqrt{\gamma_n}}B_n^\circ(w), \qquad \frac{|\bar{Z}_n|}{\sqrt{\gamma_n}} = O_{\mathbb{P}}\{\sqrt{d_n/n}\}.$$

On the relevant balls $|w| = O(\sqrt{L_n})$, this multiplicative factor is $1 + o_{\mathbb{P}}(1)$, while each of its first four derivatives is $o_{\mathbb{P}}(1)$. The uniform centered–uncentered radial difference is $o_{\mathbb{P}}(1)$, so the one-unit cutoff buffer shows that $\mathcal{U}_n := \mathcal{U}_n^\circ$ contains every centered index above $-\eta$. This proves all three assertions. □

6.3.2 Uniform finite-contact continuation

The support-count argument needs a joint expansion over all active edge patches. The following lemma supplies it. The hypotheses are stated in terms of absolute derivative envelopes; this is important because the reciprocal fitted values used in the KKT certificate need not preserve cancellation among derivatives of different bumps.

Lemma 6.17. *Let $L = \log n$, and let $q_1, \dots, q_n \in \mathbb{R}^{d_n}$ satisfy*

$$\sum_{j=1}^{n} q_j = 0, \qquad \max_j |q_j| \le C_0\sqrt{L}.$$

Suppose

$$\frac{|q_i|^2}{2} = L + \delta v_i, \qquad 0 < \delta \le 2, \qquad \bar{\delta} = 1 \wedge \delta.$$

Let $\mathcal{U} \subseteq [n]$ contain every i with $v_i > -\eta$, put $A = \{i : v_i > 0\}$, and assume

$$|A| \le J, \qquad \eta \le v_i \le C \quad (i \in A), \qquad v_i \le -\eta \quad (i \in \mathcal{U} \setminus A).$$

For $0 \le k \le 3$, define

$$\mathcal{E}_{i,k}(r) = \sup_{|z|\le r} \frac{1}{n} \sum_{j\ne i} \{1 + |q_j - q_i - z|^k\} \exp\left\{(q_i + z)\cdot q_j - \frac{|q_i + z|^2}{2}\right\}. \tag{6.43}$$

Fix $0 < r_n \le R_n \le 1/4$, and put $e_{k,n} = \max_{i\in\mathcal{U}} \mathcal{E}_{i,k}(2R_n)$. Suppose that

$$\max_{i\in\mathcal{U}} \mathcal{E}_{i,0}(2R_n) = o_{\mathbb{P}}(\bar{\delta}), \qquad \max_{i\in\mathcal{U}} \mathcal{E}_{i,1}(2R_n) = o_{\mathbb{P}}(r_n), \tag{6.44}$$

and

$$\begin{aligned} &\max_{i\in\mathcal{U}}\{\mathcal{E}_{i,2}(2R_n) + \mathcal{E}_{i,3}(2R_n)\} = o_{\mathbb{P}}(1), \qquad n\bar{\delta} \longrightarrow \infty, \qquad r_n^2 = o(\bar{\delta}), \\ &\frac{\bar{\delta}L}{n} = o(r_n), \qquad r_n e_{1,n} + r_n^2 e_{2,n} + r_n^3 e_{3,n} \qquad = o_{\mathbb{P}}(\bar{\delta}). \end{aligned} \tag{6.45}$$

For $\tau \in \mathbb{R}^{d_n}$*,* $z_i \in \mathbb{R}^{d_n}$*, and* $m_i \geq 0$*, set*

$$\Pi_{\tau,z,m} = \left(1 - \frac{\sum_{i \in A} m_i}{n}\right)\delta_\tau + \frac{1}{n}\sum_{i \in A} m_i \delta_{q_i + z_i}. \tag{6.46}$$

Uniformly for

$$|\tau| \leq K\bar{\delta}\sqrt{L}/n, \qquad |z_i| \leq R_n, \qquad 0 \leq m_i \leq K\bar{\delta},$$

its relative log likelihood has the expansion

$$\ell_q(\Pi_{\tau,z,m}) = -\frac{n}{2}|\tau|^2 + \sum_{i \in A}\left[\log\{1 + m_i e^{\vartheta_i(\tau,z_i)}\} - m_i\right] + \mathcal{R}_n(\tau, z, m), \tag{6.47}$$

where

$$\vartheta_i(\tau, z) = \delta v_i - \tau \cdot q_i + \frac{|\tau|^2}{2} - \frac{|z|^2}{2}.$$

The remainder is $o_{\mathbb{P}}(\bar{\delta}^2)$*, uniformly on this set. The same expansion may be differentiated twice in the mass variables and three times in the spatial variables. In particular, uniformly,*

$$\begin{aligned} \|\partial_m \mathcal{R}_n\| &\leq O_{\mathbb{P}}(e_{0,n} + n^{-1}) = o_{\mathbb{P}}(\bar{\delta}), & \|\partial_m^2 \mathcal{R}_n\| &= o_{\mathbb{P}}(1), \\ \|\nabla_z \mathcal{R}_n\| &\leq O_{\mathbb{P}}(\bar{\delta} e_{1,n}) = o_{\mathbb{P}}(\bar{\delta} r_n), & \|\partial_m \nabla_z \mathcal{R}_n\| &= O_{\mathbb{P}}(e_{1,n}) = o_{\mathbb{P}}(r_n), \\ \|\nabla_z^2 \mathcal{R}_n\| + \|\nabla_z^3 \mathcal{R}_n\| &\leq O_{\mathbb{P}}\{\bar{\delta}(e_{2,n} + e_{3,n})\} = o_{\mathbb{P}}(\bar{\delta}), & \|\nabla_\tau \mathcal{R}_n\| &= o_{\mathbb{P}}(\bar{\delta}\sqrt{L}), \end{aligned} \tag{6.48}$$

with all mixed blocks interpreted in the maximum block norm.

There is a critical point of ℓ_q *in these patches, unique near the locations and masses displayed in* (6.49)–(6.50)*, for which*

$$\tau_n = -\frac{1}{n}\sum_{i \in A}(1 - e^{-\delta v_i})q_i + o_{\mathbb{P}}(\bar{\delta}\sqrt{L}/n), \qquad z_{n,i} = o_{\mathbb{P}}(r_n), \tag{6.49}$$

and

$$m_{n,i} = 1 - e^{-\delta v_i} + o_{\mathbb{P}}(\bar{\delta}). \tag{6.50}$$

In particular, all edge masses are positive. If $\widetilde{\Pi} = \Pi_{\tau_n, z_n, m_n}$*, then*

$$\Gamma_{\widetilde{\Pi}}(\tau_n) = \Gamma_{\widetilde{\Pi}}(q_i + z_{n,i}) = 1, \qquad \nabla\Gamma_{\widetilde{\Pi}}(\tau_n) = \nabla\Gamma_{\widetilde{\Pi}}(q_i + z_{n,i}) = 0.$$

Writing $c_j = M_j(\widetilde{\Pi})^{-1}$*, one has*

$$\max_{j \notin A}(\log c_j)_+ = O_{\mathbb{P}}(\bar{\delta}L/n), \qquad \log c_i = -\delta v_i + o_{\mathbb{P}}(\bar{\delta}) \quad (i \in A). \tag{6.51}$$

Moreover, each active patch has the C^2 *representation*

$$\Gamma_{\widetilde{\Pi}}(q_i + y) = A_{n,i}e^{-|y|^2/2} + E_{n,i}(y), \qquad A_{n,i} = 1 + o_{\mathbb{P}}(1), \tag{6.52}$$

where

$$\|E_{n,i}\|_{C^0(B(0,R_n))} = o_{\mathbb{P}}(\bar{\delta}), \qquad \|E_{n,i}\|_{C^2(B(0,R_n))} = o_{\mathbb{P}}(1).$$

Thus the certificate is strictly log-concave on every active patch and has exactly one level-one point there. On every inactive patch indexed by $\mathcal{U} \setminus A$*,*

$$\sup_{|y| \leq R_n} \Gamma_{\widetilde{\Pi}}(q_i + y) \leq 1 - c_\eta\bar{\delta}.$$

Proof. Write $k_w(j) = \exp\{w \cdot q_j - |w|^2/2\}$, $m_+ = \sum_{i\in A} m_i$, and $w_i = q_i + z_i$. Factoring out the moving central component gives the exact identity

$$M_j(\Pi_{\tau,z,m}) = k_\tau(j)(1+x_j), \qquad x_j = -\frac{m_+}{n} + \frac{1}{n}\sum_{i\in A} m_i K_{ij},$$

where

$$K_{ij} = \frac{k_{w_i}(j)}{k_\tau(j)} = \exp\left\{(w_i-\tau)\cdot(q_j-\tau) - \frac{|w_i-\tau|^2}{2}\right\}.$$

At the observation belonging to the ith patch,

$$\frac{K_{ii}}{n} = e^{\vartheta_i(\tau,z_i)}. \tag{6.53}$$

For $j \notin A$, positivity and (6.44) give, uniformly over the displayed parameter set,

$$\max_{j\notin A}|x_j| = o_{\mathbb{P}}(1), \qquad \sum_{j\notin A} x_j = -m_+ + o_{\mathbb{P}}(\bar\delta^2), \qquad \sum_{j\notin A} x_j^2 = o_{\mathbb{P}}(\bar\delta^2).$$

Indeed, after putting $a_i = w_i - \tau$, the positive response away from the own observation equals

$$\frac{m_i}{n}\sum_{j\neq i} K_{ij} = m_i e^{-a_i\cdot\tau}\frac{1}{n}\sum_{j\neq i} e^{a_i\cdot q_j - |a_i|^2/2} = o_{\mathbb{P}}(\bar\delta^2).$$

The negative part is m_+/n per observation. Omitting the at most J active observations therefore costs $O(\bar\delta/n)$, which is $o(\bar\delta^2)$ by $n\bar\delta \to \infty$. Positivity also gives

$$\frac{1}{n^2}\sum_{j\neq i} K_{ij}^2 \le \left(\frac{1}{n}\max_{j\neq i} K_{ij}\right)\left(\frac{1}{n}\sum_{j\neq i} K_{ij}\right) \le O_{\mathbb{P}}(e_{0,n}^2),$$

which controls the quadratic Taylor remainder. Applying the same product bound after inserting the polynomial factors from (6.43) yields

$$\begin{aligned} |\mathcal{R}_n| &\le O_{\mathbb{P}}\{\bar\delta e_{0,n} + e_{0,n}^2 + \bar\delta/n\}, && \\ \|\partial_m\mathcal{R}_n\| &\le O_{\mathbb{P}}(e_{0,n} + n^{-1}), && \|\partial_m^2\mathcal{R}_n\| = o_{\mathbb{P}}(1), \\ \|\nabla_z\mathcal{R}_n\| &\le O_{\mathbb{P}}(\bar\delta e_{1,n}), && \|\partial_m\nabla_z\mathcal{R}_n\| \le O_{\mathbb{P}}(e_{1,n}), \\ \|\nabla_z^2\mathcal{R}_n\| + \|\nabla_z^3\mathcal{R}_n\| &\le O_{\mathbb{P}}\{\bar\delta(e_{2,n} + e_{3,n})\}. && \end{aligned} \tag{6.54}$$

Differentiation with respect to τ produces the same relative displacements, together with terms bounded by $\sqrt{L}/n$ from the finitely many omitted observations. Hence

$$\|\nabla_\tau\mathcal{R}_n\| \le O_{\mathbb{P}}\{\bar\delta e_{1,n} + \bar\delta\sqrt{L}\, e_{0,n} + \bar\delta\sqrt{L}/n\} = o_{\mathbb{P}}(\bar\delta\sqrt{L}). \tag{6.55}$$

These estimates prove the remainder assertions in (6.48).

Since $\sum_j q_j = 0$, $\sum_j \log k_\tau(j) = -n|\tau|^2/2$. Taylor expansion of $\sum_{j\notin A}\log(1+x_j)$, together with (6.53), proves (6.47). Differentiating the same exact factorization proves (6.48). The finiteness of $|A|$ makes every bound uniform over simultaneous masses and locations.

It remains to justify the critical point without treating the fitted denominators as fixed. Define the contact map

$$\mathcal{F}(\tau,z,m) = \left(\nabla\Gamma_\Pi(\tau),\ \{\nabla\Gamma_\Pi(q_i+z_i)\}_{i\in A},\ \{\Gamma_\Pi(q_i+z_i) - \Gamma_\Pi(\tau)\}_{i\in A}\right).$$

Put $m_i^0 = 1 - e^{-\delta v_i}$, $\tau^0 = -n^{-1}\sum_i m_i^0 q_i$, and $\widetilde{m}_i^0 = 1 - e^{-\vartheta_i(\tau^0,0)}$. Then

$$|\tau^0| = O(\bar{\delta}\sqrt{L}/n), \qquad \widetilde{m}_i^0 - m_i^0 = o(\bar{\delta}),$$

and $m_i^0 \geq c_{\eta,C}\bar{\delta}$. Use the anisotropic norms

$$\|(f_0, \{f_i\}, \{g_i\})\|_{\text{out}} = \frac{n|f_0|}{\bar{\delta}\sqrt{L}} + \max_i \frac{|f_i|}{r_n} + \max_i \frac{|g_i|}{\bar{\delta}}$$

and

$$\|(\Delta\tau, \Delta z, \Delta m)\|_{\text{in}} = \frac{n|\Delta\tau|}{\bar{\delta}\sqrt{L}} + \max_i \frac{|\Delta z_i|}{r_n} + \max_i \frac{|\Delta m_i|}{\bar{\delta}}.$$

The likelihood derivatives and contact equations are related by

$$\partial_\tau \ell_q = n\left(1 - \frac{m_+}{n}\right)\nabla\Gamma_\Pi(\tau), \qquad \nabla_{z_i}\ell_q = m_i \nabla\Gamma_\Pi(q_i + z_i), \qquad \partial_{m_i}\ell_q = \Gamma_\Pi(q_i + z_i) - \Gamma_\Pi(\tau).$$

For the quantitative estimates, put

$$\begin{aligned}
\chi_n &= \frac{e_{0,n} + n^{-1}}{\bar{\delta}} + \frac{e_{1,n}}{r_n} + \frac{e_{0,n} + e_{1,n}}{\sqrt{L}} + \frac{L}{n}, \\
\zeta_n &= \frac{e_{0,n}}{\bar{\delta}} + \frac{e_{1,n}}{r_n} + e_{2,n} + e_{3,n} + \frac{r_n e_{1,n} + r_n^2 e_{2,n} + r_n^3 e_{3,n}}{\bar{\delta}} \\
&\quad + \frac{1}{n\bar{\delta}} + \frac{\bar{\delta}L}{n r_n}.
\end{aligned}$$

Every term is $o_{\mathbb{P}}(1)$ by the hypotheses. At $(\tau^0, 0, \widetilde{m}^0)$, the principal part of (6.47) is stationary in the spatial and mass coordinates. Its central derivative is $O(\bar{\delta}L^{3/2}/n)$. Hence (6.54) and (6.55) give the explicit residual estimate

$$\|\mathcal{F}(\tau^0, 0, \widetilde{m}^0)\|_{\text{out}} = O_{\mathbb{P}}(\chi_n) = o_{\mathbb{P}}(1). \tag{6.56}$$

Let $\alpha_i = e^{-\vartheta_i(\tau^0,0)}$, which is bounded above and below, and define

$$\mathcal{J}_n(\Delta\tau, \Delta z, \Delta m) = \left(-\Delta\tau - \frac{1}{n}\sum_i \alpha_i q_i \Delta m_i,\ \{-\Delta z_i\}_{i\in A},\ \{-\Delta m_i - \alpha_i q_i \cdot \Delta\tau\}_{i\in A}\right).$$

Direct differentiation of the exact factorization gives

$$\|D\mathcal{F}(\tau^0, 0, \widetilde{m}^0) - \mathcal{J}_n\|_{\text{in}\to\text{out}} = O_{\mathbb{P}}(\zeta_n) = o_{\mathbb{P}}(1). \tag{6.57}$$

The operator $\mathcal{J}_n$ and its inverse are uniformly bounded. Indeed, its only nontrivial Schur correction is bounded by

$$\left\|\frac{1}{n}\sum_{i\in A}\alpha_i^2 q_i q_i^\top\right\|_{\text{op}} = O_J(L/n) = o(1).$$

The own-response term requires one additional contribution in the Jacobian-variation estimate. On a scaled ball of radius ρ, one has $|z_i| \leq \rho r_n$, and differentiating $-|z_i|^2/2$ in the mass equation contributes at most $C\rho r_n^2/\bar{\delta}$ in the scaled operator norm. The remaining principal and remainder terms are controlled by ρ and ζ_n. Consequently, uniformly for $0 < \rho \leq 1$,

$$\sup_{\|(\tau,z,m)-(\tau^0,0,\widetilde{m}^0)\|_{\text{in}}\leq\rho} \|D\mathcal{F} - D\mathcal{F}(\tau^0, 0, \widetilde{m}^0)\|_{\text{in}\to\text{out}} \leq O_{\mathbb{P}}\left\{\zeta_n + \rho + \frac{\rho r_n^2}{\bar{\delta}}\right\}. \tag{6.58}$$

Since $r_n^2 = o(\bar{\delta})$, the right side is $o_{\mathbb{P}}(1) + O(\rho)$.

We finally verify rather than assume the existence of a vanishing Newton radius. Since $\chi_n \xrightarrow{\mathbb{P}} 0$, choose integers $N_k \uparrow \infty$ so that

$$\mathbb{P}\{\chi_n > k^{-2}\} \le k^{-1} \qquad (n \ge N_k),$$

and set $\varrho_n = k^{-1}$ when $N_k \le n < N_{k+1}$. Then $\varrho_n \downarrow 0$ and $\chi_n = o_{\mathbb{P}}(\varrho_n)$. Equations (6.56)– (6.58), the uniformly bounded inverse of $\mathcal{J}_n$, and Newton–Kantorovich give a unique zero in the scaled ball of radius ϱ_n. Its correction has input norm $o_{\mathbb{P}}(\varrho_n)$, and therefore

$$|\tau_n - \tau^0| = o_{\mathbb{P}}(\bar{\delta}\sqrt{L}/n), \qquad \max_i |z_{n,i}| = o_{\mathbb{P}}(r_n), \qquad \max_i |m_{n,i} - \widetilde{m}_i^0| = o_{\mathbb{P}}(\bar{\delta}).$$

Together with the definition of the base point, these estimates give (6.49) and (6.50). Notice in particular that the central block is $-I + o_{\mathbb{P}}(1)$. The Hessian $\nabla^2 B_n(0) \approx -\varepsilon_n I$ is the Hessian of the unfitted bump field; it is not the Jacobian obtained when the central atom and all fitted denominators move together.

Stationarity in τ and z_i gives the gradient contact equations, while stationarity in m_i makes all displayed certificate values equal. Since $\int \Gamma_{\widetilde{\Pi}}\, d\widetilde{\Pi} = 1$, their common value is one. For $j \notin A$,

$$M_j(\widetilde{\Pi}) \ge \left(1 - \frac{\sum_i m_{n,i}}{n}\right) k_{\tau_n}(j),$$

so $\max_{j \notin A}(\log c_j)_+ = O_{\mathbb{P}}(\bar{\delta}L/n)$. The own fitted value and (6.50) give $\log c_i = -\delta v_i + o_{\mathbb{P}}(\bar{\delta})$.

Finally, separate the ith observation in the certificate:

$$\Gamma_{\widetilde{\Pi}}(q_i + y) = c_i e^{\delta v_i} e^{-|y|^2/2} + \frac{1}{n} \sum_{j \ne i} c_j e^{(q_i+y)\cdot q_j - |q_i+y|^2/2}.$$

The positive fitted-weight bound and the absolute envelopes imply (6.52). The own Gaussian has Hessian $-I$ at its peak, so the certificate is strictly log-concave on a smaller patch containing its stationary contact. For $i \in \mathcal{U} \setminus A$, its adjusted own height is at most $\exp\{-\eta\delta + O_{\mathbb{P}}(\bar{\delta}L/n)\}$, and the punctured background is $o_{\mathbb{P}}(\bar{\delta})$. This proves the inactive-patch claim and completes the proof. □

6.3.3 Support identity when $d_n \log n = o(n^2)$

The finite-contact lemma is deterministic. We now verify its hypotheses throughout the range $d_n L_n = o(n^2)$ and express the resulting support identity in the exact coordinate of Theorem 1.1.

Proposition 6.18. *Suppose*

$$d_n \longrightarrow \infty, \qquad d_n L_n = o(n^2), \qquad s_n^\star = O(1).$$

For every fixed $C, J < \infty$ and $\eta > 0$,

$$\mathbb{P}\left\{\mathcal{B}_n(C, J, \eta), \quad \widehat{K}_n \ne 1 + \#\{i : \xi_{n,i}^\star > 0\}\right\} \longrightarrow 0. \tag{6.59}$$

On $\mathcal{B}_n(C, J, \eta)$, the NPMLE is also unique apart from an event whose probability tends to zero.

Proof. Lemma 6.14 allows us to use $v_{n,i} = \xi_{n,i} - s_n$ in place of $\xi_{n,i}^\star$. On a slightly smaller buffer the two active sets agree. Put $A_n = \{i : v_{n,i} > 0\}$; throughout the proof $|A_n| \le J$.

We first describe the finite perturbation common to all the dimension ranges. The centered directions are not conditionally independent, so we embed A_n in an uncentered buffered cloud. The uniform centered–uncentered edge-coordinate estimates in Lemmas 2.10 and 2.11 show that, with probability tending to one, every $i \in A_n$ belongs to

$$\mathcal{V}_n = \{i : \xi^\circ_{n,i} - s_n > -\eta/2\}.$$

This cloud has $O_{\mathbb{P}}(1)$ points. Conditional on its uncentered radii, the directions are independent and uniform. Since $d_n \to \infty$, spherical concentration gives the following separation for the uncentered score vectors. The common empirical-centering shift is $o_{\mathbb{P}}(\sqrt{L_n})$, so it does not change their leading pairwise inner products or distances. Consequently,

$$\max_{i \neq j \in A_n} \frac{|q_i \cdot q_j|}{L_n} = o(1), \qquad |q_i - q_j|^2 = 4L_n\{1 + o(1)\}. \tag{6.60}$$

The pairwise estimates in (6.60) also hold between an active point and any fixed number of other points in the upper radial cloud. Hence a neighborhood of q_i sees its own bump at order one and every other retained upper bump at $n^{-1+o(1)}$.

Let $\delta_n = a_n/\gamma_n$ and $\bar{\delta}_n = 1 \wedge \delta_n$. The regime expansions in Section 4.2 give $\delta_n \le 2$ for all large n. They also verify the scale assumptions in Lemma 6.17. If $d_n = O(L_n)$, then $\bar{\delta}_n$ is bounded away from zero. If $L_n = o(d_n)$, then

$$\delta_n = \sqrt{L_n/d_n}\{1 + o(1)\}, \qquad n\bar{\delta}_n = \frac{L_n}{\sqrt{d_n L_n/n^2}}\{1 + o(1)\} \longrightarrow \infty,$$

because $d_n L_n = o(n^2)$. Thus $n\bar{\delta}_n \to \infty$ in the slowly growing, direct-truncation, crossover, and entropy ranges alike. After replacing η by a smaller fixed constant, the coordinate comparison and the buffer imply that every nonactive index has $v_{n,i} \le -\eta$. We may therefore take $\mathcal{U} = A_n$ in Lemma 6.17. We verify the hypotheses of Lemma 6.17 uniformly over the finitely many active indices.

Fix a small constant $r_0 < 1/4$. In the slowly growing, direct truncation, and crossover ranges, let $R_n = r_0$. In the entropy range, let $R_n = L_n^{-1/2}$. Define the available-radius envelopes

$$\widetilde{e}_{k,n} = \max_{i \in A_n} \mathcal{E}_{i,k}(2R_n), \qquad 0 \le k \le 3,$$

and put

$$\Lambda_n^{\mathrm{loc}} = \widetilde{e}_{1,n} \vee \frac{\bar{\delta}_n L_n}{n}, \qquad U_n = R_n \wedge \sqrt{\bar{\delta}_n}.$$

The earlier field estimates imply, in each range, that

$$\frac{\Lambda_n^{\mathrm{loc}}}{U_n} = o_{\mathbb{P}}(1), \qquad \widetilde{e}_{0,n} = o_{\mathbb{P}}(\bar{\delta}_n), \qquad \widetilde{e}_{2,n} + \widetilde{e}_{3,n} = o_{\mathbb{P}}(1). \tag{6.61}$$

In the slowly growing range, the proof of Lemma 6.16 bounds the sum of the absolute derivatives of the lower cloud. The remaining upper cloud has at most $e^{2\sqrt{d_n}}$ points, and angular separation bounds its first three derivative envelopes by

$$e^{2\sqrt{d_n}} L_n^{3/2} e^{-2L_n + o(L_n)} = o(1).$$

Thus $\widetilde{e}_{0,n} = o_{\mathbb{P}}(1)$ and $\sum_{k=1}^3 \widetilde{e}_{k,n} = o_{\mathbb{P}}(1)$. Since $\bar{\delta}_n = 1 + o(1)$, U_n is bounded away from zero, and $L_n/n = o(1)$, these estimates give (6.61).

In the direct truncation range, positivity of the Gaussian bumps and the proof of Proposition 4.15 give

$$\widetilde{e}_{k,n} \le O_{\mathbb{P}}(\delta_n L_n^{-10+k/2}) + e^{-L_n + o(L_n)}, \qquad 0 \le k \le 3. \tag{6.62}$$

When $\bar\delta_n$ vanishes, the regime condition $d_n(\log d_n)^2 = o(L_n^3)$ and the superlogarithmic expansion imply $\bar\delta_n \ge L_n^{-1+o(1)}$. Hence $\widetilde e_{1,n}/U_n = o_{\mathbb P}(1)$ and $(\bar\delta_n L_n/n)/U_n = o(1)$. The bounds in (6.62) for $k = 0, 2, 3$ prove the other two assertions in (6.61).

In the crossover range, (4.44) gives

$$\widetilde e_{k,n} \le L_n^{k/2} e^{-L_n/2+O(\sqrt{L_n})} = o(\delta_n^m) \qquad (0 \le k \le 3) \tag{6.63}$$

for every fixed m. Since $\bar\delta_n \asymp (\log L_n)/L_n$ and $U_n \asymp \sqrt{\bar\delta_n}$, the superpolynomial bound and $\sqrt{L_n \log L_n}/n = o(1)$ give (6.61).

Finally, in the entropy range put

$$\mu_n^{\text{off}} = \max_{i\ne j} |q_i \cdot q_j|.$$

The pairwise Gram bound gives

$$\widetilde e_{k,n} \le \frac{e^{\mu_n^{\text{off}}+O_{\mathbb P}(1)} L_n^{k/2}}{n}, \qquad 0 \le k \le 3. \tag{6.64}$$

Writing $A_n^{\text{off}} = e^{\mu_n^{\text{off}}}/n$, the rates in Lemma 5.6 imply

$$A_n^{\text{off}} = o_{\mathbb P}(\delta_n), \qquad A_n^{\text{off}} L_n = o_{\mathbb P}(1), \qquad A_n^{\text{off}} L_n^{3/2} = o_{\mathbb P}(1).$$

Since $U_n^{-1} \le \sqrt{L_n} + \bar\delta_n^{-1/2}$, $A_n^{\text{off}}\sqrt{L_n/\bar\delta_n} = \{(A_n^{\text{off}}/\bar\delta_n)(A_n^{\text{off}}L_n)\}^{1/2} = o_{\mathbb P}(1)$. It follows that $\widetilde e_{1,n}/U_n = o_{\mathbb P}(1)$; also $(\bar\delta_n L_n/n)/U_n = o(1)$, since it is at most $\bar\delta_n L_n^{3/2}/n + \sqrt{\bar\delta_n} L_n/n$. The three rates for A_n^{off}, applied to (6.64), prove the remaining assertions in (6.61).

We now choose the Newton scale. From (6.61), choose integers $M_k \uparrow \infty$ so that

$$\mathbb P\left\{\Lambda_n^{\text{loc}}/U_n > k^{-2}\right\} \le k^{-1} \qquad (n \ge M_k),$$

and set $c_n = k^{-1}$ for $M_k \le n < M_{k+1}$. Then $c_n \downarrow 0$ and

$$\Lambda_n^{\text{loc}} = o_{\mathbb P}(c_n U_n).$$

Set $r_n = c_n U_n$. Then

$$r_n \le R_n, \qquad \widetilde e_{1,n} = o_{\mathbb P}(r_n), \qquad \frac{\bar\delta_n L_n}{n} = o(r_n), \qquad \frac{r_n^2}{\bar\delta_n} \le c_n^2 \longrightarrow 0.$$

Because the envelopes decrease with the patch radius, the three inequalities in (6.61) verify (6.44). Moreover,

$$r_n \widetilde e_{1,n} = o_{\mathbb P}(r_n^2) = o_{\mathbb P}(\bar\delta_n), \qquad r_n^2 \widetilde e_{2,n} + r_n^3 \widetilde e_{3,n} = o_{\mathbb P}(\bar\delta_n),$$

so every condition in (6.45) holds.

Lemma 6.17 therefore constructs a measure $\widetilde\Pi$ with one central contact τ_n and one contact $w_{n,i}$ in each active patch. Its edge weights satisfy

$$|\tau_n| = O_{\mathbb P}(\bar\delta_n\sqrt{L_n}/n), \qquad w_{n,i} = q_i + o_{\mathbb P}(r_n), \qquad n\pi_{n,i} = 1 - e^{-\delta_n v_{n,i}} + o_{\mathbb P}(\bar\delta_n). \tag{6.65}$$

The buffer keeps every displayed mass positive. In the entropy range, retaining the explicit right side of (6.64) in the contact-map proof gives

$$n\pi_{n,i} = 1 - e^{-\delta_n v_{n,i}} + O_{\mathbb P}\left\{\frac{e^{\mu_n^{\text{off}}}}{n} + \frac{\delta_n L_n}{n}\right\}. \tag{6.66}$$

It remains to check that the finite perturbation has not created a contact elsewhere. If c_j denotes the reciprocal fitted value at observation j, then

$$(\log c_j)_+ = O_{\mathbb{P}}\left\{\frac{\bar{\delta}_n L_n}{n}\right\} \quad (j \notin A_n), \qquad \log c_i = -\delta_n v_{n,i} + o_{\mathbb{P}}(\bar{\delta}_n) \le -c_\eta \bar{\delta}_n \quad (i \in A_n). \tag{6.67}$$

In the entropy range the contact equation gives the sharper cancellation

$$\max_{i \in A_n} \left|\delta_n v_{n,i} + \log c_i\right| + \max_{j \notin A_n} |\log c_j| = O_{\mathbb{P}}\left\{\frac{e^{\mu_n^{\text{off}}}}{n} + \frac{\delta_n L_n}{n}\right\} = n^{-1+o_{\mathbb{P}}(1)}. \tag{6.68}$$

We first control the inactive denominators without using the contact equation. If $m = \sum_{i \in A_n} n\pi_{n,i} = O_{\mathbb{P}}(\delta_n)$, then

$$M_j = \left(1 - \frac{m}{n}\right) e^{\tau_n \cdot q_j - |\tau_n|^2/2} + \frac{1}{n}\sum_{i \in A_n} n\pi_{n,i} e^{w_{n,i}\cdot q_j - |w_{n,i}|^2/2}.$$

Here $|\tau_n| = O_{\mathbb{P}}(\delta_n \sqrt{L_n}/n)$, so the central term differs from one by $O_{\mathbb{P}}(\delta_n L_n/n)$. For $j \notin A_n$, each edge response in the last sum is at most $n^{-1} e^{\mu_n^{\text{off}} + o_{\mathbb{P}}(1)}$. Consequently,

$$\max_{j \notin A_n} |M_j - 1| = O_{\mathbb{P}}\left\{\frac{\delta_n L_n}{n} + \frac{\delta_n e^{\mu_n^{\text{off}}}}{n^2}\right\},$$

which proves the inactive part of (6.68).

We may now evaluate the exact level-one contact equation in the ith edge patch. If w_i is its contact and E_i is the contribution of all observations other than i to the certificate there, then the own term gives the exact identity

$$\log M_i = \delta_n v_{n,i} - \frac{|w_i - q_i|^2}{2} - \log(1 - E_i). \tag{6.69}$$

The derivative envelope (6.64), the C^2 certificate expansion in Lemma 6.17, and the local stationarity equation give

$$|w_i - q_i|^2 = O_{\mathbb{P}}\left\{\frac{e^{2\mu_n^{\text{off}}} L_n}{n^2} + \frac{\delta_n^2 L_n^2}{n^2}\right\}, \qquad E_i = O_{\mathbb{P}}\left\{\frac{e^{\mu_n^{\text{off}}}}{n} + \frac{\delta_n L_n}{n}\right\}.$$

Indeed, a retained upper response is at most $e^{\mu_n^{\text{off}}}/n$, and the central displacement and lower fitted values contribute $O_{\mathbb{P}}(\delta_n L_n/n)$. Substitution in (6.69) proves the active part of (6.68). These estimates are uniform because $|A_n| \le J$, and the argument is noncircular: the inactive denominator bound was obtained before the certificate contribution E_i was estimated. The positive part in (6.67) is $o_{\mathbb{P}}$ of the smallest gap used in each one-atom exclusion proof. The following four-region argument verifies these comparisons while proving the weighted exclusion rather than applying the unweighted results formally. Write

$$\Gamma_{\widetilde{\Pi}}(w) = \Gamma_I(w) + \sum_{i \in A_n} G_i(w), \qquad G_i(w) = \frac{c_i}{n} e^{w\cdot q_i - |w|^2/2},$$

where $I = A_n^c$ and

$$\Gamma_I(w) = \frac{1}{n}\sum_{j \in I} c_j e^{w\cdot q_j - |w|^2/2}, \qquad D_I(w) = \frac{1}{n}\sum_{j \in I} e^{w\cdot q_j - |w|^2/2}.$$

If

$$\omega_n = \max_{j \in I} (\log c_j)_+,$$

then, pointwise and after every truncation used in the cited proofs,

$$\Gamma_I(w) \le e^{\omega_n} D_I(w). \tag{6.70}$$

Thus every Bernstein or net upper bound for the inactive field changes by at most $O(\omega_n)$.

There are four regions. First consider a fixed small ball around an active contact. The own term satisfies the exact relation

$$G_i(w_{n,i}) = 1 - E_i,$$

where E_i is the remaining certificate contribution, rather than being one by itself. The C^2 representation (6.52) gives $\nabla^2 \log \Gamma_{\widetilde{\Pi}} \preceq -cI$ throughout this ball. Since the contact is stationary and has value one, it is the unique level-one point there.

Next take a corresponding ball around an inactive upper point q_j. Its adjusted own peak is at most

$$\frac{c_j}{n} e^{|q_j|^2/2} \le \exp\{\omega_n + \delta_n v_{n,j}\} \le 1 - c_\eta \bar{\delta}_n.$$

The punctured C^3 estimates (4.35), (6.64), (6.42), and (6.63), in their respective ranges, make all remaining terms $o_{\mathbb{P}}(\bar{\delta}_n)$. Hence an inactive upper patch contains no contact.

In the central region, the weighted Hessian differs from the unweighted one by

$$O_{\mathbb{P}}(\omega_n L_n) + O_{\mathbb{P}}\left\{\frac{\bar{\delta}_n L_n^2}{n}\right\} = o_{\mathbb{P}}(\varepsilon_n).$$

To obtain this bound, write each fitted value as $M_j = 1 + x_j$. The negative part of x_j is uniformly $O_{\mathbb{P}}(\bar{\delta}_n L_n/n)$, while

$$\frac{1}{n}\sum_j (x_j)_+ \le C\frac{\bar{\delta}_n}{n} + O_{\mathbb{P}}\left\{\frac{\bar{\delta}_n L_n}{n}\right\};$$

the first term follows by averaging each edge response and using $B_n(w_{n,i}) = O_{\mathbb{P}}(1)$. Multiplication by $\max_j |q_j|^2 = O_{\mathbb{P}}(L_n)$ gives the displayed covariance perturbation. It is therefore negative definite in the microscopic ball around the translated central contact. On the surrounding annuli, (6.70) raises the field by at most $O_{\mathbb{P}}(\omega_n)$. We compare the active profiles with the annular gap explicitly. Uniformly for $i \in A_n$,

$$G_i(w) \le \exp\left\{-\frac{1}{2}(|q_i| - |w|)^2 + o_{\mathbb{P}}(1)\right\}, \qquad |q_i| = \sqrt{2L_n}\{1 + o_{\mathbb{P}}(1)\}.$$

Let $g_n(r) = 1 - e^{-\varepsilon_n r^2/2}$, and let $g_{0,n}$ be the smallest gap at the inner endpoint of the annular argument. If $\varepsilon_n r^2 \le 1$, then

$$g_n(r) \ge c\varepsilon_n r^2 \ge c g_{0,n}, \qquad G_i(w) \le n^{-1+o(1)} = o(g_{0,n}).$$

The last comparison is $n^{-1} = o(\varepsilon_n^3/L_n^3)$ in the slow and direct truncation ranges and is immediate in the crossover range. If $\varepsilon_n r^2 \ge 1$ and $r \le |q_i|/2$, then $g_n(r) \ge c$ and $G_i(w) \le n^{-1/4+o(1)}$. Hence

$$\sum_{i \in A_n} G_i(w) = o_{\mathbb{P}}\{g_n(|w|)\} \quad (r_{0,n} \le |w| \le \tfrac{1}{2}\sqrt{2L_n}). \tag{6.71}$$

Together with the annular estimates established separately in the slow, direct-truncation, crossover, and entropy regimes, this proves that the central contact is the only contact in the inner region.

Finally consider the complement of the central and upper patches. Choose their fixed radius $R = R(J)$ so large that $Je^{-R^2/2} < 1/20$; the patches are disjoint by angular separation. On the annulus between the implicit-function neighborhood and radius R, the punctured C^2 bound and the exact level-one contact show that the Gaussian profile generated by the nearest active observation has its unique maximum at the constructed contact. In the slow range this uniform statement is precisely Lemma 6.16; in the direct-truncation and crossover ranges it follows from the local C^2 estimates used to construct the active contacts. The inactive part is again controlled by (6.70). Conditional angular separation implies that at most one upper profile is nearest; outside its patch that profile loses a fixed Gaussian factor, while the sum of all other upper profiles is $n^{-1+o_{\mathbb{P}}(1)}$ in the direct truncation range, $e^{-L_n/2+o(L_n)}$ in the crossover range, and $o_{\mathbb{P}}(1)$ in the slow range by Lemma 6.16. Thus the outer-region decompositions from the unweighted proofs remain valid after the fitted denominators are inserted: equation (6.70) controls the inactive field, and the active tails retain their original bounds. These estimates exclude every remaining level-one point. More concretely, in the crossover proof (4.50) bounds the clipped inactive field by $0.9 + o_{\mathbb{P}}(1)$; choose R so that the active tails total less than 0.05. In the slow proof, choose R so that those tails are smaller than $\rho/4$, where ρ is the fixed slack in Lemma 4.12; the fitted-denominator error is also $o_{\mathbb{P}}(\rho)$. The direct-truncation outer estimate has an $o_{\mathbb{P}}(\delta_n)$ inactive background and is stronger still.

In the entropy range, Lemma 6.15, applied with $\beta_j = \log c_j$, directly localizes every level-one contact to the central patch or an active upper patch; (6.68) verifies all of the lemma's weight hypotheses. The strict log-concavity asserted there leaves only the constructed contacts.

The fitted-denominator error and the tails of the active profiles are smaller than the KKT gaps in each unweighted localization argument. More precisely,

$$\frac{L_n}{n} = o\left(\frac{\varepsilon_n^3}{L_n^3}\right) \quad \text{in the slow range,} \qquad \frac{\delta_n L_n}{n} + \delta_n \mathfrak{j}_n = o(g_{0,n}) \quad \text{in the direct range,}$$

where $g_{0,n}$ is from Proposition 4.16; in the crossover range, $\delta_n L_n/n = o(L_n^{-3})$ and $e^{-L_n/2+o(L_n)} = o(\delta_n^m)$ for every fixed m. Thus the weighted estimates retain the gaps and negative curvature used in all four regions.

This verifies the global KKT inequality for the constructed measure. Strict concavity of the log likelihood makes its fitted vector unique. The contact likelihood vectors are linearly independent: their active-coordinate block has diagonal entries $n^{1+o(1)}$ and off-diagonal entries $n^{o(1)}$, while the central vector is bounded. Since $n^{-1}\sum_j k_{w_{n,i}}(j) = B_n(w_{n,i}) = O_{\mathbb{P}}(1)$ for each of the finitely many edge columns, their combined column sum is $O_{\mathbb{P}}(n)$. Choose a deterministic $K_n \to \infty$ so slowly that $K_n = o(n)$. Then

$$\#\left\{j : \max_{i\in A_n} k_{w_{n,i}}(j) > K_n\right\} \le \frac{1}{K_n}\sum_{i\in A_n}\sum_{j=1}^{n} k_{w_{n,i}}(j) = o_{\mathbb{P}}(n).$$

Since $|A_n| \le J$, with probability tending to one there is an inactive observation at which every edge entry is at most $K_n = o(n)$. Adjoin that observation as one extra row; its central entry is $1 + o_{\mathbb{P}}(1)$. After scaling the edge columns by n^{-1}, their active-coordinate block is asymptotically diagonal with diagonal bounded away from zero, while all edge entries in the added row are $o(1)$. The resulting square minor is therefore nonsingular. Every maximizing mixing measure must therefore use all and only the $1 + |A_n|$ contacts.

Finally, pass to a subsequence on which $\mathfrak{r}_n = d_n(\log d_n)^2/L_n^3$ tends to zero, stays bounded above and away from zero, or tends to infinity. These cases invoke, respectively, the slow/direct estimates (split at $d_n = (\log L_n)^2$), the crossover estimates, and the entropy estimates. They exhaust $d_n L_n = o(n^2)$ and prove (6.59). □

6.3.4 Ordered contacts in the simplex regime

When $d_n L_n/n^2$ is bounded below, the centered sample is close to a regular simplex and the isolated-bump model no longer applies. A radial crossing then creates a contact near an ordered Potts state. Write

$$\chi_n = \frac{n-2}{n-1}, \qquad p_i^{(i)} = \frac{n-1}{n}, \qquad p_j^{(i)} = \frac{1}{n(n-1)} \quad (j \neq i).$$

The identities $\sum_j q_j = 0$ and (2.6) give, exactly,

$$\sum_j p_j^{(i)} q_j = \chi_n q_i, \qquad \frac{1}{2}\left|\sum_j p_j^{(i)} q_j\right|^2 - K(p^{(i)} \| \mathbf{u}_n) = \chi_n^2 \delta_n^\star \xi_{n,i}^\star. \tag{6.72}$$

Thus the sign of the ordered-state free energy is the sign of the exact radial excess, with no centering approximation.

Proposition 6.19. *Let $n \geq 3$, let $q_1, \ldots, q_n$ be a centered triangular array, and write*

$$R_i = \frac{|q_i|^2}{2}, \qquad u_i = \frac{R_i - R_n^\star}{\delta_n}, \qquad G_n = (q_i \cdot q_j)_{i,j \leq n}.$$

Suppose $\delta_n \downarrow 0$, $\delta_n = O(L_n/n)$, and, for fixed $C, J < \infty$ and $\eta > 0$,

$$\max_i u_i \leq C, \qquad \#\{i : u_i > -\eta\} \leq J, \qquad \min_i |u_i| > \eta.$$

Assume also that

$$\max_{i \neq j} |q_i \cdot q_j| = o(L_n^{-1}), \qquad \|G_n\|_{\mathrm{op}} = O(L_n), \qquad \frac{C_n}{\delta_n L_n^2} \longrightarrow 0, \qquad \frac{C_n^2}{n\delta_n} \longrightarrow 0, \tag{6.73}$$

where

$$C_n = \max_{i \neq j} \left| q_i \cdot q_j + \frac{|q_i|^2}{n-1} \right|.$$

Then, for all sufficiently large n, every NPMLE has

$$|\operatorname{supp}(\Pi)| = 1 + \#\{i : u_i > 0\}. \tag{6.74}$$

There is one central contact and, for each i with $u_i > 0$, one contact $w_i = \chi_n q_i + o(1)$; its weight satisfies

$$n\pi_i = \frac{n}{n-1}\delta_n u_i + o(\delta_n). \tag{6.75}$$

All remainders are uniform when C, J, η in the buffer are fixed. Moreover, the NPMLE is unique.

Proof. Put $A = \{i : u_i > 0\}$, so $|A| \leq J$, and set $w_i^0 = \chi_n q_i$ for $i \in A$. The buffer and the definition of R_i imply

$$\max_i |q_i| = O(\sqrt{L_n}), \qquad C_n = o(L_n^{-1}), \qquad \delta_n L_n = o(1), \qquad \frac{C_n^2}{n} = o(\delta_n). \tag{6.76}$$

Indeed, the first assertion follows from $R_i \leq R_n^\star + C\delta_n = O(L_n)$, and the second follows from the first condition in (6.73) and $|q_i|^2/(n-1) = O(L_n/n)$. The last two assertions follow directly from the assumed bounds on δ_n and C_n.

If $A = \varnothing$, then $\max_i R_i \le R_n^\star - \eta\delta_n$. Proposition 5.4, with $g_n = \eta\delta_n$, applies by (6.73). Its proof is strict away from $p = \mathbf{u}_n$, so the one-atom certificate equals one only at the origin. The KKT conditions then force every NPMLE to be δ_0, and the conclusion follows. We henceforth assume that $A \neq \varnothing$.

We begin by locating the maxima of the one-atom certificate. If

$$e_{ij} = q_i \cdot q_j + \frac{|q_i|^2}{n-1},$$

then $|e_{ij}| \le C_n$, and direct substitution gives

$$\begin{aligned} k_{w_i^0}(i) &= (n-1)\exp\{(2\chi_n - \chi_n^2)\delta_n u_i\}, \\ k_{w_i^0}(j) &= \frac{1}{n-1}\exp\{-(\chi_n^2 + 2\chi_n/(n-1))\delta_n u_i + \chi_n e_{ij}\}, \qquad j \neq i. \end{aligned} \tag{6.77}$$

At $u_i = 0$ and $e_{ij} = 0$, these responses are $n-1$ and $(n-1)^{-1}$, respectively, and their average is one. Moreover, $\sum_{j\neq i} e_{ij} = 0$, by centering. The linear row error therefore cancels when the off-diagonal responses are summed. Expanding (6.77), using $C_n = o(1)$, gives

$$\log B_n(w_i^0) = \chi_n^2\delta_n u_i + O_{C,J}\left(\delta_n^2 + \frac{C_n^2}{n}\right) = \chi_n^2\delta_n u_i + o(\delta_n). \tag{6.78}$$

The coefficient can also be read directly from the exact dual identity (6.72).

The Gibbs weights at w_i^0 differ from $p^{(i)}$ by $O_{C,J}\{(C_n + \delta_n)/n\}$ in total variation. Differentiating the log-sum-exp representation therefore gives, uniformly for $i \in A$,

$$\begin{aligned} |\nabla \log B_n(w_i^0)| &\le K_{C,J}\frac{\sqrt{L_n}}{n}(C_n + \delta_n), \\ \nabla^2 \log B_n(w_i^0) &= -I + O(L_n/n) + o(1) = -I + o(1). \end{aligned} \tag{6.79}$$

The same estimate holds uniformly on $B(w_i^0, L_n^{-2})$: moving by L_n^{-2} changes every exponent by $O(L_n^{-3/2})$, while the total Gibbs mass outside coordinate i remains $O(n^{-1})$. The Newton–Kantorovich theorem now gives a unique critical point $\bar{w}_i$ in that ball, with

$$|\bar{w}_i - w_i^0| \le K_{C,J}\frac{\sqrt{L_n}}{n}(C_n + \delta_n). \tag{6.80}$$

It is a strict local maximum, and Taylor's theorem gives

$$h_i := \log B_n(\bar{w}_i) = \chi_n^2\delta_n u_i + \varepsilon_{n,i}, \qquad \max_{i\in A}|\varepsilon_{n,i}| = o(\delta_n). \tag{6.81}$$

More explicitly, the remainder is bounded by a constant times

$$\delta_n^2 + \frac{C_n^2}{n} + \frac{L_n(C_n + \delta_n)^2}{n^2},$$

which is $o(\delta_n)$ by (6.76) and (6.73). In particular, $h_i \ge c_\eta\delta_n$ for every $i \in A$ and all sufficiently large n.

For $i \in A$, let $v_i = (k_{\bar{w}_i}(1), \dots, k_{\bar{w}_i}(n))$. Equations (6.77)–(6.79) give

$$A_{ij}^{(0)} := \frac{1}{n^2}\sum_{r=1}^{n}\{v_i(r) - 1\}\{v_j(r) - 1\} = \beta_n^{\mathrm{P}}\mathbf{1}_{\{i=j\}} + e_{n,ij}, \qquad \max_{i,j\in A}|e_{n,ij}| = o(1), \tag{6.82}$$

where

$$\beta_n^{\mathrm{P}} = \frac{(n-2)^2}{n(n-1)}.$$

For instance, at the unperturbed ordered state the diagonal expression is exactly

$$\frac{1}{n^2}\left\{(n-2)^2 + (n-1)\left(\frac{1}{n-1} - 1\right)^2\right\} = \beta_n^{\mathrm{P}},$$

whereas the cross expressions are $O(n^{-1})$. The perturbations in (6.77) and (6.80) contribute $o(1)$, uniformly over the at most J active indices.

We next construct exact contacts and masses. For $m = (m_i)_{i\in A}$ and $x = (\tau, (w_i)_{i\in A})$, define

$$\Pi_{x,m} = \left(1 - \frac{1}{n}\sum_{i\in A} m_i\right)\delta_\tau + \frac{1}{n}\sum_{i\in A} m_i \delta_{w_i},$$

and write $\Gamma_{x,m}$ for its certificate. Consider the stationarity map

$$\mathcal{S}(x,m) = \left(\nabla \log \Gamma_{x,m}(\tau), \left(\nabla \log \Gamma_{x,m}(w_i)\right)_{i\in A}\right).$$

At $m = 0$ and $x = x^0 := (0, (\bar{w}_i)_{i\in A})$, this map vanishes. Equip the product of location spaces with the maximum of the Euclidean block norms. For each fixed $M < \infty$, the response table gives, uniformly on

$$|m|_\infty \le M\delta_n, \qquad |\tau| \le L_n^{-2}, \qquad |w_i - \bar{w}_i| \le L_n^{-2},$$

the bounds

$$\|D_x\mathcal{S}(x^0,0)^{-1}\| \le 2, \quad \|D_x\mathcal{S}(x,m) - D_x\mathcal{S}(x^0,0)\| = o(1), \quad \|D_m\mathcal{S}(x,m)\| \le K_{C,J,M}\frac{\sqrt{L_n}}{n}. \tag{6.83}$$

To verify the first two bounds, the diagonal location blocks are $-I + o(1)$ by (6.79), while the central block is

$$\frac{1}{n}\sum_{r=1}^n q_r q_r^\top - I = -I + O(L_n/n).$$

At zero mass, the derivative of an active equation with respect to the central location is also negligible: if the central atom is at τ, then

$$\log \Gamma_{\delta_\tau}(w) = \log B_n(w-\tau) - (w-\tau)\cdot\tau,$$

so at $(\tau, w) = (0, \bar{w}_i)$,

$$D_\tau \nabla_w \log \Gamma_{\delta_\tau}(w) = -\nabla^2 \log B_n(\bar{w}_i) - I = o(1).$$

All other cross-location blocks vanish at zero mass and are $O_{C,J}(\delta_n L_n) = o(1)$ on the displayed neighborhood. Finally, one differentiation with respect to m_j produces a normalized product of a central or ordered response and its first moment. The response table bounds this by $K\sqrt{L_n}/n$, proving the last estimate in (6.83).

The quantitative implicit-function theorem now supplies unique smooth locations satisfying the stationarity equations. Write

$$x(m) = \left(\tau(m), (w_i(m))_{i\in A}\right).$$

They obey

$$|\tau(m)| + \max_{i\in A}|w_i(m) - \bar{w}_i| \le K_{C,J,M}\frac{\sqrt{L_n}}{n}|m|_1. \tag{6.84}$$

Define the logarithmic contact-height differences

$$\mathcal{H}_i(m) = \log \Gamma_{x(m),m}\{w_i(m)\} - \log \Gamma_{x(m),m}\{\tau(m)\}.$$

Stationarity permits the envelope theorem. At zero mass, $\mathcal{H}_i(0) = h_i$, and direct differentiation gives

$$-\partial_{m_j}\mathcal{H}_i(0) = A^{(0)}_{ij} + o(1) = \beta^{\mathrm{P}}_n \mathbf{1}_{\{i=j\}} + o(1). \tag{6.85}$$

The additional $o(1)$, beyond (6.82), includes the movement of the central contact; by (6.83) its contribution is $O(L_n/n)$.

The response averages in (6.82) also control the second derivative. In particular, for any $i, j, k \in A$,

$$\frac{1}{n^3}\sum_{r=1}^{n}\{1 + v_i(r)\}\{1 + v_j(r)\}\{1 + v_k(r)\} = O_{C,J}(1).$$

The fitted denominators stay between $1/2$ and 2 on $|m|_\infty \le M\delta_n$, so

$$\sup_{|m|_\infty \le M\delta_n} \|D^2_m \mathcal{H}(m)\| = O_{C,J,M}(1). \tag{6.86}$$

Consequently,

$$\mathcal{H}(m) = h - A_n m + R_n(m), \qquad A_n = \beta^{\mathrm{P}}_n I_{|A|} + o(1), \qquad |R_n(m)|_\infty \le K_{C,J}|m|^2_\infty. \tag{6.87}$$

Choose M larger than $2C$. Since $u_i \in [\eta, C]$ on A, the map $m \mapsto A_n^{-1}\{h + R_n(m)\}$ is a contraction on a ball of radius $o(\delta_n)$ about $A_n^{-1}h$. It therefore has a unique fixed point in $|m|_\infty \le M\delta_n$, and

$$m_i = \frac{\chi^2_n}{\beta^{\mathrm{P}}_n}\delta_n u_i + o(\delta_n) = \frac{n}{n-1}\delta_n u_i + o(\delta_n). \tag{6.88}$$

Every component is positive for all sufficiently large n. The resulting measure $\widetilde{\Pi} = \Pi_{x(m),m}$ has one stationary central contact and one stationary contact in each active patch, all at the same certificate level. Since $\int \Gamma_{\widetilde{\Pi}}\, d\widetilde{\Pi} = 1$, that level equals one.

We finally prove the global KKT inequality. Put $c_r = M_r(\widetilde{\Pi})^{-1}$. The response expansion and the mass formula imply

$$|\log c_r| = O_{C,J}(\delta_n L_n/n) \quad (r \notin A), \qquad \log c_i = -\chi_n \delta_n u_i + o(\delta_n) \quad (i \in A). \tag{6.89}$$

Indeed, (6.84) gives $|\tau(m)| = O_{C,J}(\delta_n\sqrt{L_n}/n)$ and $\max_i |w_i(m) - \bar{w}_i| = O_{C,J}(\delta_n\sqrt{L_n}/n)$. For $r \notin A$, the central response is

$$e^{\tau(m)\cdot q_r - |\tau(m)|^2/2} = 1 + O_{C,J}(\delta_n L_n/n),$$

the total mass removed from the central atom is $O_{C,J}(\delta_n/n)$, and all active responses together contribute $O_{C,J}(\delta_n/n^2)$. Thus $M_r = 1 + O_{C,J}(\delta_n L_n/n)$. At an active observation i, (6.77) and (6.88) instead give

$$M_i = 1 + \frac{n-2}{n}\,\frac{n}{n-1}\delta_n u_i + o(\delta_n) = 1 + \chi_n \delta_n u_i + o(\delta_n).$$

Taking negative logarithms proves (6.89).

For clarity, put

$$F_{\log c}(p) = \frac{1}{2}\left|\sum_r p_r q_r\right|^2 - K(p\|\mathbf{u}_n) + \sum_r p_r \log c_r.$$

The weighted entropy dual (6.36) shows that $\Gamma_{\widetilde{\Pi}}(w) \geq 1$ only if its Gibbs vector $p = p(w)$ satisfies $F_{\log c}(p) \geq 0$. If $\max_r p_r \leq 1 - L_n^{-2}$, the stability inequality (5.3) may be applied after retaining the active diagonal terms. Indeed, writing $r = p - \mathbf{u}_n$, their contribution is at most

$$\sum_{i \in A} \delta_n u_i r_i^2 \leq \chi_n \sum_{i \in A} \delta_n u_i p_i + O_{C,J}(\delta_n/n^2),$$

where the scalar inequality $(p_i - 1/n)^2 \leq \chi_n p_i + n^{-2}$ holds for $0 \leq p_i \leq 1$. This is canceled by (6.89). The remaining active-weight error is $o(\delta_n)$, and the positive weight on inactive coordinates is $O(\delta_n L_n/n) = o(\delta_n)$. Consequently, (5.3) and (6.73) give, for a deterministic sequence $\varpi_n \downarrow 0$,

$$F_{\log c}(p) \leq -\frac{c}{L_n}|p - \mathbf{u}_n|_1^2 + \varpi_n \delta_n, \qquad \varpi_n \longrightarrow 0.$$

Thus nonnegative weighted free energy forces

$$|p - \mathbf{u}_n|_1 = o\{\sqrt{\delta_n L_n}\} = o(L_n/\sqrt{n}).$$

Because $\sum_r q_r = 0$, the corresponding $\mu(p) = \sum_r p_r q_r$ satisfies

$$|\mu(p)| \leq \max_r |q_r|\,|p - \mathbf{u}_n|_1 = o(L_n\sqrt{\delta_n}) = o(L_n^{3/2}/\sqrt{n}) = o(1).$$

For the Gibbs vector of an actual certificate point, the exact square identity is

$$\log \Gamma_{\widetilde{\Pi}}(w) = F_{\log c}\{p(w)\} - \frac{1}{2}|w - \mu\{p(w)\}|^2. \tag{6.90}$$

Therefore $\Gamma_{\widetilde{\Pi}}(w) \geq 1$ implies $w = o(1)$ in the diffuse case.

Suppose instead that $p_i = 1 - y > 1 - L_n^{-2}$, and decompose p as in the proof of Proposition 5.4: write $p_j = y r_j$ for $j \neq i$ and

$$v_r = \sum_{j \neq i}\left(r_j - \frac{1}{n-1}\right) q_j, \qquad \mathbf{u}_{n-1} = \left(\frac{1}{n-1}, \dots, \frac{1}{n-1}\right).$$

Applying the local-simplex decomposition from that proposition with $R_i = R_n^\star + \delta_n u_i$ gives

$$\begin{aligned}\frac{1}{2}\left|\sum_r p_r q_r\right|^2 &- K(p\|\mathbf{u}_n) \\ &\leq \delta_n u_i c_y^2 - \{k_n(y) - R_n^\star c_y^2\} + y|c_y|C_n, \qquad c_y = 1 - \frac{ny}{n-1}.\end{aligned} \tag{6.91}$$

Before the row error is bounded, the decomposition from the proof of Proposition 5.4 also retains

$$-yK(r\|\mathbf{u}_{n-1}) + \frac{y^2}{2}|v_r|^2 \leq -\{1 - o(1)\}yK(r\|\mathbf{u}_{n-1}), \tag{6.92}$$

uniformly for $y \leq L_n^{-2}$. Indeed,

$$|v_r|^2 \leq \frac{\|G_n\|_{\mathrm{op}}}{\log(n-1)} K(r\|\mathbf{u}_{n-1}) = O\{K(r\|\mathbf{u}_{n-1})\}.$$

We use a sharpened form of the last term. Pinsker's inequality and the definition of C_n give

$$|q_i \cdot v_r| \leq C_n \|r - \mathbf{u}_{n-1}\|_1 \leq C_n \{2K(r\|\mathbf{u}_{n-1})\}^{1/2}.$$

The spectral estimate in Proposition 5.4 retains a fixed fraction of the conditional entropy $yK(r\|\mathbf{u}_{n-1})$. Young's inequality therefore absorbs the cross term $yc_y q_i \cdot v_r$, at the cost of $O(yC_n^2)$. Thus the right side of (6.91) may be replaced by

$$\delta_n u_i c_y^2 - \{k_n(y) - R_n^\star c_y^2\} - c\, yK(r\|\mathbf{u}_{n-1}) + O(yC_n^2). \tag{6.93}$$

The scalar function in braces is nonnegative and, on $0 \leq y \leq L_n^{-2}$, vanishes only at $y = 1/n$; its second derivative there is $n + O(L_n)$. More generally, its explicit formula gives, on this interval,

$$k_n(y) - R_n^\star c_y^2 \geq \frac{c}{n}\{z \log z - z + 1\}, \qquad z = ny. \tag{6.94}$$

This follows by differentiating the two sides on either side of $z = 1$; both vanish with zero derivative at $z = 1$, while the second derivative of the left side after the change of variables is $n^{-1}z^{-1}\{1 + o(1)\}$. Combining this estimate with (6.89) gives the following dichotomy. If $i \notin A$, then $u_i \leq -\eta$, and the weighted free energy is at most $-c\eta\delta_n$ for all large n. If $i \in A$, the active fitted weight cancels the radial term: uniformly on the present interval,

$$\delta_n u_i c_y^2 + (1 - y)\log c_i \leq O_{C,J}(\delta_n/n) + o(\delta_n), \tag{6.95}$$

because $\sup_{0 \leq y \leq L_n^{-2}}\{c_y^2 - \chi_n(1-y)\} = 1/(n-1)$. Let $\phi(z) = z\log z - z + 1$. Outside $z = ny \in [1/2, 2]$, one has $\phi(z) \geq c(1+z)$, so the term $yC_n^2 = zC_n^2/n$ is absorbed by (6.94). Inside that interval it is $O(C_n^2/n) = o(\delta_n)$. Nonnegative weighted free energy therefore implies

$$\phi(ny) + nyK(r\|\mathbf{u}_{n-1}) = o(n\delta_n) = o(L_n).$$

It follows that

$$y = o\left(\frac{L_n}{n \log L_n}\right), \qquad yK(r\|\mathbf{u}_{n-1}) = o(\delta_n).$$

Hence

$$y|v_r| \leq C\{y^2 K(r\|\mathbf{u}_{n-1})\}^{1/2} = o\{y\delta_n\}^{1/2} = o(1).$$

Since $|c_y - \chi_n| = |1 - ny|/(n-1)$, we obtain

$$\mu(p) = c_y q_i + y v_r = \chi_n q_i + o(1).$$

Substituting (6.95) and the bounds $yK(r\|\mathbf{u}_{n-1}) = o(\delta_n)$ and $y|v_r| = o(1)$ into the definition of $F_{\log c}$ gives $F_{\log c}(p) = o(\delta_n)$. Equation (6.90) then shows that every certificate point with value at least one lies within $o(1)$ of $\chi_n q_i$.

It remains to prove uniqueness inside the localized patches. Fix a small constant $r_0 > 0$. In $B(\tau, r_0)$, the Gibbs probabilities are at most $n^{-1+o(1)}$, and hence

$$\|\operatorname{Cov}_{p(w)}(q)\|_{\mathrm{op}} \leq n^{-1+o(1)}\|G_n\|_{\mathrm{op}} = o(1).$$

In $B(w_i, r_0)$, the total Gibbs mass outside coordinate i is $n^{-1+o(1)}$, so the covariance operator is again $o(1)$. Therefore

$$\nabla^2 \log \Gamma_{\widetilde{\Pi}}(w) = \operatorname{Cov}_{p(w)}(q) - I \preceq -\frac{1}{2}I$$

on every localized patch. Each patch contains exactly one stationary point, namely the contact already constructed there, and that point is its unique maximum. The diffuse and near-vertex localization arguments in this proof now give $\Gamma_{\widetilde{\Pi}} \le 1$, with equality exactly at the central and active contacts.

The constructed measure is an NPMLE. Its fitted vector is unique by strict concavity in the fitted likelihood vector. The likelihood vectors at its contacts are linearly independent. Indeed, on rows indexed by A, the active-contact block has diagonal entries $(n-1)\{1+o(1)\}$ and off-diagonal entries $(n-1)^{-1}\{1+o(1)\}$, while the central column is $1+o(1)$. Choose any inactive row, which exists because $|A| \le J < n$ for large n. On that row the central entry is $1+o(1)$, and every active entry is $(n-1)^{-1}\{1+o(1)\}$. After scaling the active columns by n^{-1}, this $(1+|A|)$-square minor converges to an invertible triangular matrix. The KKT conditions for the unique fitted likelihood vector restrict the support of every maximizing measure to these contacts; linear independence then makes its representing weights unique. Thus every maximizing measure uses the same positive weights on precisely these contacts. Equations (6.74) and (6.75) follow from (6.88). □

For the random Gaussian cloud, all assumptions of Proposition 6.19 hold whenever $d_n L_n/n^2$ is bounded below.

Corollary 6.20. *Suppose*

$$d_n L_n/n^2 \ge c > 0, \qquad s_n^\star = O(1).$$

For every fixed $C, J < \infty$ and $\eta > 0$,

$$\mathbb{P}\left\{\mathcal{B}_n(C,J,\eta), \quad \widehat{K}_n \ne 1 + \#\{i : \xi_{n,i}^\star > 0\}\right\} \longrightarrow 0. \tag{6.96}$$

On $\mathcal{B}_n(C,J,\eta)$, the NPMLE is also unique apart from an event whose probability tends to zero.

Proof. Here $L_n = o(d_n)$, and Lemma 5.6 gives

$$\delta_n^\star = \sqrt{L_n/d_n}\{1+o(1)\}, \qquad \mu_n^{\mathrm{off}} := \max_{i\ne j} |q_i \cdot q_j| = o_{\mathbb{P}}(L_n^{-1}),$$
$$\|G_n\|_{\mathrm{op}} = O_{\mathbb{P}}(L_n), \qquad C_n = O_{\mathbb{P}}(L_n^{3/2}/\sqrt{d_n}).$$

Consequently,

$$\frac{C_n}{\delta_n^\star L_n^2} = O_{\mathbb{P}}(L_n^{-1}), \qquad \frac{C_n^2}{n\delta_n^\star} = O_{\mathbb{P}}(\delta_n^\star L_n^2/n) = o_{\mathbb{P}}(1),$$

and $\delta_n^\star = O(L_n/n) \to 0$. Proposition 6.19 applies on an event whose probability tends to one. □

6.4 Proof of the main theorem

We finish by combining the dimension-specific localization estimates with the radial Poisson limits. This short argument proves all assertions in Theorem 1.1.

For $d_n \ge 3$, let

$$\mathcal{B}_n(C,J,\eta) = \left\{\max_i \xi_{n,i}^\star \le C, \quad \#\{i : \xi_{n,i}^\star > -\eta\} \le J, \quad \min_i |\xi_{n,i}^\star| > \eta\right\}. \tag{6.97}$$

Proof of Theorem 1.1. First let $d = 2$, and put $\xi_{n,i} = |Z_i|^2/2 - \log n$. For every fixed $\eta > 0$, Proposition 6.13 gives

$$\#\{i : \xi_{n,i} > x_\lambda + \eta\} \le \widehat{K}_n - 1 \le \#\{i : \xi_{n,i} > x_\lambda - \eta\},$$

apart from an event whose limiting probability is $O(\eta)$. The extremal process of the $\xi_{n,i}$'s converges to a Poisson process with intensity $e^{-x}\,dx$. Letting $n \to \infty$ and then $\eta \downarrow 0$ gives

$$\widehat{K}_n - 1 \Rightarrow \mathrm{Pois}(e^{-x_\lambda}).$$

Taking the probability of zero gives $\mathbb{P}\{\widehat{K}_n = 1\} \to Q_2(\lambda)$. Proposition 6.13 also shows that the NPMLE is unique with probability tending to one.

Now suppose $d_n \geq 3$ and $s_n^\star \to s$. We first connect support points to radial exceedances. For every fixed $C, J < \infty$ and $\eta > 0$,

$$\#\{i : \xi_{n,i}^\star > \eta\} \leq \widehat{K}_n - 1 \leq \#\{i : \xi_{n,i}^\star > -\eta\} \tag{6.98}$$

on $\mathcal{B}_n(C, J, \eta)$, apart from an event whose probability tends to zero. To verify this uniformly, argue along subsequences. If d_n is bounded, pass to a further subsequence on which $d_n = d \geq 3$ is fixed. Lemma 2.9 converts the exact excess and threshold into their fixed-dimensional versions, and Proposition 6.13 gives the claim. If $d_n \to \infty$, pass to a further subsequence on which either $d_n \log n = o(n^2)$ or $d_n \log n / n^2$ is bounded below. The exact support identity follows respectively from Proposition 6.18 or Corollary 6.20. These alternatives cover every subsequence and prove (6.98). They also show that, on $\mathcal{B}_n(C, J, \eta)$, the NPMLE is unique apart from an event whose probability tends to zero.

The centered radial exceedance count satisfies

$$\#\{i : \xi_{n,i}^\star > 0\} \;\Rightarrow\; \mathrm{Pois}(e^{-s}). \tag{6.99}$$

For bounded d_n, this follows from Lemmas 4.1 and 2.9. If $d_n \to \infty$ and $d_n \log n = o(n^2)$, Proposition 2.8 gives the uncentered point-process limit. Lemmas 2.10 and 2.11 compare the centered and uncentered radial energies uniformly over the sample. They therefore transfer counts at thresholds separated from zero by a fixed amount; letting the separation decrease proves (6.99). In the complementary range, the count assertion is part of Proposition 2.12.

For fixed $\eta > 0$, the expected number of excesses in $[-\eta, \eta]$ tends to

$$e^{-s}(e^{\eta} - e^{-\eta}) = O_s(\eta).$$

The radial point-process limits used in (6.99) also make the maximum excess and the number of excesses above $-\eta$ tight. We may therefore choose C, J so that $\mathcal{B}_n(C, J, \eta)$ fails, apart from an excess in $[-\eta, \eta]$, with arbitrarily small probability. The sandwich (6.98), followed by $n \to \infty$, $C, J \to \infty$, and $\eta \downarrow 0$, gives

$$\widehat{K}_n - 1 \Rightarrow \mathrm{Pois}(e^{-s}).$$

The uniqueness conclusions in Proposition 6.13, Proposition 6.18, and Corollary 6.20, together with the preceding buffer estimate, show that the probability of nonuniqueness tends to zero. Again, the probability of zero gives the one-atom limit.

It remains to prove the off-critical tightness assertions in Theorem 1.1. In dimension two, pass along subsequences on which λ_n converges in $(0, \infty]$. A finite limit is covered above. If $\lambda_n \to \infty$, monotonicity compares the one-atom event with every fixed λ_0; its limiting probability is $Q_2(\lambda_0) \uparrow 1$. For $d_n \geq 3$, any subsequence on which $s_n^\star$ is bounded below has a further subsequence on which $s_n^\star$ converges in $\mathbb{R}$ or tends to $+\infty$. The first case is covered by the Poisson limit. In the second, monotonicity and comparison with a fixed s_0, followed by $s_0 \to \infty$, show that $\widehat{K}_n = 1$ with probability tending to one. □

Finally, monotonicity converts the one-atom limits into the samplewise critical-defect laws stated in the introduction. In dimension two, the map

$$c_{2,n}(\varepsilon) = \frac{\rho_n^2 \varepsilon}{2(1-\varepsilon)}$$

is continuous and strictly increasing. Hence, for every $\lambda > 0$,

$$\mathbb{P}\{c_{2,n}(\varepsilon_*(Z_{1:n})) \le \lambda\} = \mathbb{P}\{E_n^{(1)}(\varepsilon_n(\lambda))\} \longrightarrow Q_2(\lambda),$$

where $\varepsilon_n(\lambda) = c_{2,n}^{-1}(\lambda)$. For $d_n \ge 3$, use instead the critical coordinate

$$c_{\ge 3,n}(\varepsilon) = \frac{R_n^\star/[\sigma_n^2(1-\varepsilon)] - b_n}{a_n}$$

is also continuous and strictly increasing. Evaluating the one-atom limit at $c_{\ge 3,n}(\varepsilon) = s$ gives $\mathbb{P}\{S_{*,n} \le s\} \to \exp\{-e^{-s}\}$.

6.5 The regularization path

This subsection proves Theorem 1.2 and then refines it with the mass and spatial-mark limits in Proposition 6.23. The fixed-parameter support identities can be made uniform across a compact part of the critical window. To obtain a path limit, we must also resolve what happens at each activation threshold: as the variance defect passes the threshold associated with one radial observation, the corresponding edge mass reaches zero transversally. We first record the radial process that supplies these thresholds, then prove the required uniform continuation.

For dimension two and dimensions at least three, respectively, set

$$\mathcal{X}_n^{(2)} = \sum_{i=1}^n \delta_{\zeta_{n,i}^{(2)}}, \quad \zeta_{n,i}^{(2)} = \frac{|Z_i|^2}{2} - \log n, \qquad \mathcal{X}_n^{(\ge 3)} = \sum_{i=1}^n \delta_{\zeta_{n,i}^{(\ge 3)}}, \quad \zeta_{n,i}^{(\ge 3)} = \frac{|Z_i - \bar{Z}_n|^2/(2\sigma_n^2) - b_n}{a_n}. \tag{6.100}$$

Thus the exact excess in (2.13), evaluated along (1.13), is $\xi_{n,i}^\star(s) = \zeta_{n,i}^{(\ge 3)} - s$.

The one-level exceedance limits used to prove Theorem 1.1 hold jointly at every finite collection of levels. Equivalently, the whole radial cloud has the following limit.

Proposition 6.21. *In dimension two, and for every sequence $d_n \ge 3$,*

$$\mathcal{X}_n^{(2)} \Longrightarrow \mathcal{X}, \qquad \mathcal{X}_n^{(\ge 3)} \Longrightarrow \mathcal{X} \tag{6.101}$$

vaguely on $\mathbb{R}$, where $\mathcal{X}$ has intensity $e^{-x}\,dx$. For each fixed $d \ge 2$, the point-process convergence remains valid when each radial excess is marked by the direction of its observation; the limiting directions are independent and uniform on S^{d-1}.

Proof. For $d = 2$, the variables $|Z_i|^2/2$ are standard exponential, and hence $n\mathbb{P}\{\zeta_{n,i}^{(2)} > x\} = e^{-x}$ for every fixed x. Counts in disjoint bounded intervals are multinomial rare-event counts, which proves the first convergence.

For uncentered Gamma energies, Lemma 2.7 gives, uniformly for bounded x,

$$n\overline{F}_n(b_n + a_n x) \longrightarrow e^{-x}.$$

The rare-event multinomial calculation from the $d = 2$ case proves convergence of the uncentered point process. In the range where empirical centering is negligible on the a_n-scale, Lemmas 2.10 and 2.11 transfer the process. In the remaining range, the proof of Proposition 2.12 gives joint upper-tail factorization uniformly when finitely many thresholds lie in $b_n + a_n O(1)$. Although (2.22) is displayed there with a common threshold, its determinant-and-tilting proof also allows bounded coordinate-dependent offsets: use the corresponding diagonal matrix of tilt parameters in place of the scalar tilt. The determinant comparison and tilted local limit are unchanged. Inclusion–exclusion then gives the factorization for disjoint intervals, so the factorial-moment criterion proves the second convergence in (6.101). For fixed dimension, radius and direction are independent; the marking theorem completes the proof. ◻

We next follow the optimizer while one point crosses its activation level. In dimension two, let $a_\lambda > 0$ be the unique solution of $\Phi(a_\lambda) + \phi(a_\lambda)/a_\lambda = e^\lambda$, and put

$$b_\lambda = 1 - e^{-\lambda}\Phi(a_\lambda), \qquad q_{\lambda,x} = \frac{[1 - e^{-(x - x_\lambda)}]_+}{b_\lambda}, \tag{6.102}$$

and recall that Λ is the inverse of $\lambda \mapsto x_\lambda$. For $d_n \geq 3$, define

$$\gamma_n(s) = \frac{\sigma_n^2(b_n + a_n s)}{R_n^\star}, \qquad \delta_n(s) = \frac{\sigma_n^2 a_n}{\gamma_n(s)}. \tag{6.103}$$

The next proposition is the uniform edge reduction. Its separation hypothesis is only a proof device; Proposition 6.21 will remove it.

Proposition 6.22. *Fix compact intervals $J \subset \operatorname{int}(J^+) \Subset (0,\infty)$ and $I \subset \operatorname{int}(I^+) \subset \mathbb{R}$, and constants $C, M < \infty$ and $\eta > 0$. In dimension two, work on the event*

$$\max_i \zeta_{n,i}^{(2)} \leq C, \qquad \#\{i : \zeta_{n,i}^{(2)} \geq x_{\inf J^+}\} \leq M, \tag{6.104}$$

on which the values $\Lambda(\zeta_{n,i}^{(2)}) \in J^+$ are pairwise at least 4η apart and are at least 4η from the endpoints of J^+. Apart from an event whose probability tends to zero, each such point has one activation value $\widehat{\lambda}_{n,i}$, and

$$\max_i |\widehat{\lambda}_{n,i} - \Lambda(\zeta_{n,i}^{(2)})| = o_{\mathbb{P}}(1). \tag{6.105}$$

Its patch contains one fitted atom when $\lambda < \widehat{\lambda}_{n,i}$ and none when $\lambda \geq \widehat{\lambda}_{n,i}$, and, uniformly for $\lambda \in J$,

$$n\pi_{n,i}^{(2)}(\lambda) = q_{\lambda,\zeta_{n,i}^{(2)}} + o_{\mathbb{P}}(1). \tag{6.106}$$

Points above $x_{\sup J^+}$ are active throughout J, points below $x_{\inf J^+}$ are inactive throughout J, and the mass estimate applies to every point satisfying the second cutoff in (6.104).

For every sequence $d_n \geq 3$, work instead on the event

$$\max_i \zeta_{n,i}^{(\geq 3)} \leq C, \qquad \#\{i : \zeta_{n,i}^{(\geq 3)} \geq \inf I^+ - 1\} \leq M, \tag{6.107}$$

on which the points in I^+ are pairwise at least 4η apart and at least 4η from the endpoints of I^+. Apart from an event whose probability tends to zero, each point in I^+ has one activation value $\widehat{s}_{n,i}$, and

$$\max_i |\widehat{s}_{n,i} - \zeta_{n,i}^{(\geq 3)}| = o_{\mathbb{P}}(1). \tag{6.108}$$

Its patch contains one fitted atom when $s < \widehat{s}_{n,i}$ and none when $s \geq \widehat{s}_{n,i}$. Fixing $s_0 \in I$ and writing $\bar{\delta}_n = 1 \wedge \delta_n(s_0)$, its mass satisfies, uniformly for $s \in I$,

$$n\pi_{n,i}^{(\geq 3)}(s) = \left[1 - e^{-\delta_n(s)(\zeta_{n,i}^{(\geq 3)} - s)}\right]_+ + o_{\mathbb{P}}(\bar{\delta}_n), \tag{6.109}$$

with the exponential expression replaced by its linearization when $d_n L_n / n^2$ is bounded below, the regime in which the centered cloud is asymptotically simplex-like. Points above $\sup I^+$ are active throughout I, points below $\inf I^+$ are inactive throughout I, and the mass estimate applies to every point in the second cutoff in (6.107). In both dimensions there are no other noncentral contacts, and the NPMLE is unique throughout the original compact interval.

Proof. We first explain why the fixed-parameter estimates are uniform. In dimension two, the local diffuse background and the response of a radial point of excess x are

$$B_\lambda(r) = e^{-\lambda}\Phi(-r), \qquad C_{\lambda,x}(r,z) = e^{-\lambda}e^{x-r^2/2-z^2/2}.$$

Their spatial derivatives through order two and their first λ-derivatives are bounded uniformly when λ ranges over a compact subset of $(0,\infty)$. At finite n,

$$\partial_\lambda e^{-\kappa_n|t|^2} = -\frac{|t|^2}{\rho_n^2}e^{-\kappa_n|t|^2}, \qquad \kappa_n = \lambda/\rho_n^2,$$

so the Gaussian envelopes used for the undifferentiated field also dominate its λ-derivative on $|t| = \rho_n + O(1)$. Applying the compact decomposition in Proposition 3.6, the lower-cloud bounds, and the C^2 estimates of Section 6.2 on a deterministic λ-grid, then using these derivative bounds between grid points, makes every fixed-parameter estimate uniform in $C^2(r,z)\cap C^1(\lambda)$. The radial compactness and outside-window gaps in Proposition 3.2 and Lemmas 3.3–3.4 are uniform for the same reason: on J^+, every damping constant stays in a fixed compact subset of $(0,\infty)$.

It remains to inspect the limiting patch program (6.14). Its unique spatial optimizer is $(-a_\lambda, 0)$. After the other patch variables have been optimized, the derivative at zero rescaled mass is

$$S_{\lambda,x}(0) = b_\lambda\{e^{x-x_\lambda} - 1\}. \tag{6.110}$$

The functions $a_\lambda, x_\lambda, b_\lambda$ are smooth on compact intervals, with $x'_\lambda > 0$ and $b_\lambda > 0$. At the unique zero $\lambda = \Lambda(x)$,

$$\partial_\lambda S_{\lambda,x}(0) = -b_\lambda x'_\lambda,$$

which is bounded away from zero. The spatial Hessian there is $-b_\lambda I_2$, and the mass curvature is strictly negative. The uniform finite-patch expansion therefore gives one empirical zero satisfying (6.105). Strict concavity places positive mass below that zero and zero mass at and above it. Solving the limiting interior score equation gives (6.106). Continuity of the optimizer at zero makes the estimate uniform through the crossing.

For $d_n \geq 3$, write

$$h_n(s) = \frac{\gamma_n'(s)}{\gamma_n(s)} = \frac{a_n}{b_n + a_n s}.$$

If $q_j(s) = y_j/\sqrt{\gamma_n(s)}$, differentiation of an edge kernel gives

$$\begin{aligned}&\left|\partial_s \exp\left\{(q_i(s)+z)\cdot q_j(s) - \frac{|q_i(s)+z|^2}{2}\right\}\right| \\ &\qquad \leq C h_n(s)\{\log n + |q_j(s) - q_i(s) - z|^2\}\exp\left\{(q_i(s)+z)\cdot q_j(s) - \frac{|q_i(s)+z|^2}{2}\right\}.\end{aligned} \tag{6.111}$$

Moreover $h_n(s)\log n \asymp \delta_n(s)$, uniformly on I. Thus the order-zero and order-two absolute envelopes already proved in the spatial, entropy, and support-recovery arguments also control one derivative in s. As in dimension two, a slowly refining deterministic grid makes all buffered support identities uniform on I.

Near one crossing, optimize the central contact and every other patch, leaving only the rescaled mass $m_i = n\pi_i \geq 0$ in the crossing patch free. The algebraic expansion in the proof of Lemma 6.17 is obtained before the threshold buffer is used: its envelope assumptions do not depend on the sign of the crossing excess.

Holding this one mass free therefore gives the expansion in Lemma 6.17 uniformly while $|\zeta_{n,i}^{(\geq 3)} - s| \leq 2\eta$, with every other patch still buffered. In particular, the boundary score and its derivative are

$$\partial_{m_i} \mathcal{L}_{n,i}(s, 0) = e^{\delta_n(s)(\zeta_{n,i}^{(\geq 3)} - s)} - 1 + o_{\mathbb{P}}(\bar{\delta}_n), \tag{6.112}$$

$$\partial_s \partial_{m_i} \mathcal{L}_{n,i}(s, 0) = -\delta_n(s) + o_{\mathbb{P}}(\bar{\delta}_n), \tag{6.113}$$

while its second mass derivative is $-1 + o_{\mathbb{P}}(1)$. Indeed, $\delta_n' = -h_n \delta_n$, so differentiating the leading exponent adds only $O\{h_n \delta_n |\zeta_{n,i}^{(\geq 3)} - s|\}$ to $-\delta_n$; inequality (6.111) keeps every remainder $o_{\mathbb{P}}(\bar{\delta}_n)$ after differentiation. When $d_n L_n / n^2$ is bounded below, the ordered-contact argument gives (6.112)–(6.113) with the exponential linearized. More explicitly, the ordered-contact response remainder in (6.81) is bounded by a constant times

$$\delta_n^2 + \frac{C_n^2}{n} + \frac{\log n \, (C_n + \delta_n)^2}{n^2}.$$

Here $\delta_n' = -h_n \delta_n$, and the Gram-error array defining C_n is rescaled by $\gamma_n(s)^{-1}$, so its upper Dini derivative is at most $h_n C_n$. The rates in (6.76) therefore show directly that one s-derivative leaves this remainder $o_{\mathbb{P}}(\delta_n)$.

Equation (6.113) is uniformly negative. Hence each boundary score has exactly one zero, which proves (6.108); strict mass curvature and the spatial certificate give one contact below the zero and none at or above it. Solving the interior score equation gives (6.109). Finally, the uniform residual KKT gaps proved in the dimension-specific support identities exclude contacts outside these patches. The strict concavity of the local mass programs and the uniqueness of each spatial contact then prove simultaneous uniqueness, including at an activation value, where the crossing patch has zero mass. □

Proof of Theorem 1.2. Enlarge the parameter interval slightly. Proposition 6.21 makes the radial maximum and the number of points in the enlarged critical window tight. The limiting process is simple, so after imposing fixed radial and cardinality cutoffs, its activation levels are separated from one another and from the enlarged interval's endpoints with probability tending to one as the separation constant decreases to zero. Apply Proposition 6.22 first with fixed cutoffs and separation, then let the cutoffs tend to infinity and the separation tend to zero. A diagonal choice gives a single event of probability tending to one on which all of its conclusions hold.

In dimension two, the support-count path is the tail-count path of the perturbed activation measure $\sum_i \delta_{\widehat{\lambda}_{n,i}}$. Equation (6.105) allows a piecewise-linear time change that moves each $\widehat{\lambda}_{n,i}$ to $\Lambda(\zeta_{n,i}^{(2)})$ and converges uniformly to the identity. Points above the enlarged interval contribute the same constant to both paths. After adding them, the resulting path is exactly $\mathcal{X}_n^{(2)}((x_\lambda, \infty))$. Proposition 6.21 and the continuous-mapping theorem for tail counts of a simple point process prove (1.14) in the J_1 topology.

In dimensions at least three, equation (6.108) supplies an analogous time change. After that time change and the addition of the points active throughout I, the path is $\mathcal{X}_n^{(\geq 3)}((s, \infty))$. This proves (1.15). Simultaneous uniqueness is part of Proposition 6.22. □

Proposition 6.22 also identifies the mass and location carried by every jump. We state the mass limit explicitly because it is stronger than the support-count path and shows that the fitted mixing distribution itself changes continuously even when its support size jumps.

Proposition 6.23. *Associate each noncentral fitted atom with its radial observation, and extend its mass by zero after the corresponding patch becomes inactive. In dimension two, for compact $J \Subset (0, \infty)$,*

$$\max_{i:\, \zeta_{n,i}^{(2)} \geq x_{\inf J} - 1} \sup_{\lambda \in J} \left| n \pi_{n,i}^{(2)}(\lambda) - q_{\lambda, \zeta_{n,i}^{(2)}} \right| \xrightarrow{\mathbb{P}} 0, \tag{6.114}$$

after first imposing and then removing a fixed upper radial cutoff. Moreover,

$$\left\{ n \sum_i \pi_{n,i}^{(2)}(\lambda) : \lambda \in J \right\} \Longrightarrow \left\{ \sum_{x \in \mathcal{X}} q_{\lambda,x} : \lambda \in J \right\} \qquad \text{in } C(J). \tag{6.115}$$

For $d_n \geq 3$, fix $s_0 \in I$, pass to a subsequence on which $\delta_n(s_0) \to \delta \in [0, \infty]$, and define

$$\psi_\delta(y) = \begin{cases} (1 - e^{-\delta y})/(1 \wedge \delta), & 0 < \delta < \infty, \\ y, & \delta = 0, \\ 1, & \delta = \infty, \end{cases} \qquad y > 0, \quad \psi_\delta(y) = 0 \quad (y \leq 0).$$

Then

$$\max_{i:\, \zeta_{n,i}^{(\geq 3)} \geq \inf I - 1} \sup_{s \in I} \left| \frac{n\pi_{n,i}^{(\geq 3)}(s)}{\bar{\delta}_n} - \psi_\delta((\zeta_{n,i}^{(\geq 3)} - s)_+) \right| \xrightarrow{\mathbb{P}} 0, \tag{6.116}$$

and the sum of these profiles converges in $C(I)$ to $s \mapsto \sum_{x \in \mathcal{X}} \psi_\delta((x - s)_+)$. For every fixed $d \geq 2$, the convergence is joint with the radial directions; each active atom has the direction of its associated extreme observation asymptotically. In dimension two its radial score location is $\rho_n - a_\lambda + o_{\mathbb{P}}(1)$, uniformly away from activation.

Proof. Equations (6.106) and (6.109) give the individual limits on events with fixed radial and cardinality cutoffs. Lemma 2.7 gives $a_n/b_n \to 0$, and hence $\delta_n(s)/\delta_n(s_0) \to 1$ uniformly on I; this turns (6.109), after division by $\bar{\delta}_n$, into the profile ψ_δ. Proposition 6.21 removes these cutoffs. The profiles are jointly continuous in the path parameter and radial mark, including at activation where they vanish. Only finitely many limiting Poisson points contribute on a compact parameter interval. Summation is therefore a continuous map into $C(J)$ or $C(I)$, which proves the total-mass limits. The marked part of Proposition 6.21, together with the unique spatial optimizer in each local likelihood program, gives the final assertions. ◻

References

[Ano26a] Support divergence of mixture NPMLEs under finite mixtures, 2026. In preparation.

[Ano26b] Variance misspecification and the one-dimensional Gaussian mixture NPMLE, 2026. In preparation.

[CW26] Jiafeng Chen and Yihong Wu. Sharp regret-Hellinger bounds for Gaussian empirical Bayes via polynomial approximation. *arXiv preprint arXiv:2605.02070*, 2026.

[DWYZ23] Natalie Doss, Yihong Wu, Pengkun Yang, and Harrison H Zhou. Optimal estimation of high-dimensional Gaussian location mixtures. *Ann. Stat.*, 51(1):62–95, 2023.

[DZ16] Lee H Dicker and Sihai D Zhao. High-dimensional classification via nonparametric empirical Bayes and maximum likelihood inference. *Biometrika*, 103(1):21–34, 2016.

[GGMT26] Subhroshekhar Ghosh, Adityanand Guntuboyina, Satyaki Mukherjee, and Hoang-Son Tran. Gaussian mixtures and non-parametric likelihoods through the lens of statistical mechanics. *arXiv preprint arXiv:2603.23196*, 2026.

[GKV18] Jiaying Gu, Roger Koenker, and Stanislav Volgushev. Testing for homogeneity in mixture models. *Econometric Theory*, 34(4):850–895, 2018.

[GVDV01] Subhashis Ghosal and Aad W Van Der Vaart. Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Stat.*, 29(5):1233–1263, 2001.

[HY01] Peter Hall and Matthew York. On the calibration of Silverman's test for multimodality. *Statist. Sinica*, 11(2):515–536, 2001.

[JZ09] Wenhua Jiang and Cun-Hui Zhang. General maximum likelihood empirical Bayes estimation of normal means. *Ann. Stat.*, 37(4):1647–1684, 2009.

[KM14] Roger Koenker and Ivan Mizera. Convex optimization, shape constraints, compound decisions, and empirical Bayes rules. *Journal of the American Statistical Association*, 109(506):674–685, 2014.

[KS26] Taehyun Kim and Bodhisattva Sen. Empirical Bayes estimation and inference via smooth nonparametric maximum likelihood. *arXiv preprint arXiv:2603.27843*, 2026.

[KW56] Jack Kiefer and Jacob Wolfowitz. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Stat.*, pages 887–906, 1956.

[Lin83] Bruce G Lindsay. The Geometry of Mixture Likelihoods: A General Theory. *Ann. Stat.*, 11(1):86–94, 1983.

[PS25] Yury Polyanskiy and Mark Sellke. Nonparametric MLE for Gaussian location mixtures: Certified computation and generic behavior. *arXiv preprint arXiv:2503.20193*, 2025.

[PW20] Yury Polyanskiy and Yihong Wu. Self-Regularizing Property of Nonparametric Maximum Likelihood Estimator in Mixture Models. *arXiv preprint arXiv:2008.08244*, 2020.

[Rob50] Herbert Robbins. A generalization of the method of maximum likelihood: Estimating a mixing distribution (Abstract). *Annals of Mathematical Statistics*, 21(2):314–315, 1950.

[SG20] Sujayam Saha and Adityanand Guntuboyina. On the nonparametric maximum likelihood estimator for Gaussian location mixture densities with application to Gaussian denoising. *Ann. Stat.*, 48(2):738–762, 2020.

[SGS25] Jake A Soloff, Adityanand Guntuboyina, and Bodhisattva Sen. Multivariate, heteroscedastic empirical Bayes via nonparametric maximum likelihood. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(1):1–32, 2025.

[Sil81] B. W. Silverman. Using kernel density estimates to investigate multimodality. *J. Roy. Statist. Soc. Ser. B*, 43(1):97–99, 1981.

[vdVW96] Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer, 1996.

[Zha09] Cun-Hui Zhang. Generalized maximum likelihood estimation of normal mixture densities. *Statistica Sinica*, pages 1297–1318, 2009.

[ZV26] Yan Zhang and Stanislav Volgushev. Adaptivity of the NPMLE to finitely discrete mixing distributions in Gaussian/Poisson mixtures. *arXiv preprint arXiv:2604.12087*, 2026.