Benign population landscapes do not transfer to empirical risk minimization in nonsmooth optimization

Penghao Wang* Mateo Díaz† Liwei Jiang*

Abstract

Benign landscapes provide a geometric explanation for the success of iterative methods on nonconvex problems arising in data science. The landscape is benign if every local minimizer is global and every other critical point is a strict saddle. For methods that avoid strict saddles, convergence to a critical point therefore implies convergence to a global minimizer. For smooth stochastic optimization, a common strategy for establishing such landscapes proceeds in two steps: (*i*) establish benign geometry for the population objective—which is often easier to analyze—and (*ii*) transfer this geometry to the empirical objective using concentration of gradients and Hessians. Recent work has extended saddle-point avoidance guarantees to weakly convex nonsmooth optimization through the notion of active strict saddles. We show, however, that benign population geometry need not transfer to empirical objectives in this setting. Specifically, for robust phase retrieval with Gaussian measurements—a well-studied and highly structured problem—the population objective is benign. Nevertheless, the empirical objective has sharp spurious local minima with probability bounded away from zero as the sample size tends to infinity. This asymptotic probability lower bound is independent of the dimension.

1 Introduction

Simple iterative methods, such as gradient descent, have proven to be highly effective at solving nonconvex optimization problems stemming from data science, machine learning, and signal processing. A working geometric explanation for this success is the benign landscape of the objective: all local minima are global, and every other critical point is a strict saddle—meaning that its Hessian has a strictly negative eigenvalue. Under this geometry, many suitably tuned first-order methods escape strict saddles and consequently recover global minimizers whenever they converge [41, 53]. Further, stochastically perturbed variants of gradient descent enjoy strong nonasymptotic guarantees [36]. This geometric explanation is supported by a number of high-impact problems that exhibit benign landscapes under suitable assumptions [8, 32, 33, 47, 49, 50, 61, 65, 69].

Most of these examples arise from statistical models and admit both population and empirical objectives. Formally, for samples $z_1, \dots, z_n$ from some distribution, these objectives are defined as

$$F(x) := \mathbb{E}_z f(x, z), \qquad \text{and} \qquad F_n(x) := \frac{1}{n}\sum_{i=1}^{n} f(x, z_i),$$

where f is problem specific. Although the objective being minimized is often F_n, the population objective F tends to be simpler to analyze. This motivates a common two-stage approach to establish benign geometry: first, show that it holds for the population objective F and then prove that this

*Edwardson School of Industrial Engineering, Purdue University.

†Department of Statistics & Data Science Institute, University of Chicago. MD was partially supported by NSF awards CCF 2442615 and DMS 2502377, and a Sloan Research Fellowship.

geometry persists for the empirical objective F_n whenever n is sufficiently large. The second step is carried out by exploiting the concentration of gradients and Hessians of F_n around those of F.

The seminal work of Davis and Drusvyatskiy [17] extended saddle-avoidance guarantees to nonsmooth weakly convex optimization—a broad class encompassing many problems arising in data science [15]. Unlike in the smooth setting, a classical Hessian is not available for these functions. To overcome this obstacle, they introduce the notion of an active strict saddle. In a nutshell, these are critical points lying on a manifold along which the objective is smooth—and has a direction of strictly negative curvature—and away from which the function grows sharply; see Figure 1 for an illustration and Section 2 for a formal definition. Just as in the smooth setting, several first-order methods are guaranteed to avoid these active strict saddles [4, 14, 17, 19, 55]. These results provide a reason to seek nonsmooth objectives with benign landscapes and raise a natural question:

Does benign population geometry transfer to the empirical landscape in nonsmooth optimization?

We answer this question negatively: for general weakly convex problems, this is not the case. To do so, we consider phase retrieval, an inverse problem with a long history in physics and signal processing [29, 60]. Although phase retrieval is commonly formulated over the complex numbers, we study a stylized real version in which the goal is to recover a vector $x^\star \in \mathbb{R}^d$, up to sign, from n nonlinear measurements of the form $\langle a_i, x^\star\rangle^2$, where the vectors a_i are i.i.d. standard Gaussians. To solve this problem, we consider the robust formulation

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \left| \langle a_i, x\rangle^2 - \langle a_i, x^\star\rangle^2 \right|. \tag{1.1}$$

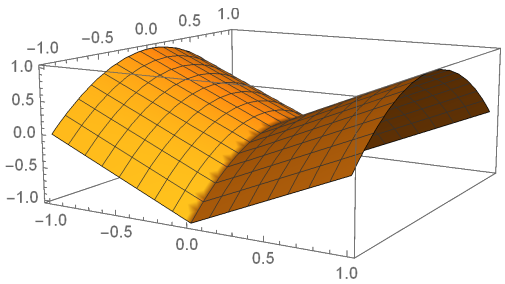


Figure 1: The function $(x, y) \mapsto |x| - y^2$ has an active strict saddle at zero, with the y-axis as the active manifold.

This objective is highly structured: it is weakly convex, a single-index model, and exhibits composite structure, i.e., it is a composition of a convex function with a smooth map [26]. Further, this formulation and related low-rank matrix recovery formulations have been studied extensively and serve as standard testbeds for theory and algorithms in weakly convex optimization [12, 13, 15, 20, 21, 23, 57].

Population objective. The population counterpart of (1.1) and its first-order landscape were characterized in [21]. For a unit-norm $x^\star$, the critical points of $F(x) = \mathbb{E}|\langle a, x\rangle^2 - \langle a, x^\star\rangle^2|$ are

$$\{\pm x^\star\} \cup \{0\} \cup \mathcal{S} \qquad \text{with} \qquad \mathcal{S} \coloneqq \{\tau u \mid \|u\| = 1 \text{ and } u \perp x^\star\},$$

where $\tau \approx 0.4416$ is the unique solution of a certain equation. That is, the critical points correspond to the true underlying solutions $\pm x^\star$, zero, and a $(d-2)$-dimensional sphere orthogonal to the solution. We show here that F is twice continuously differentiable on $\mathbb{R}^d \setminus \{\pm x^\star\}$. Further, zero and the vectors on the sphere $\mathcal{S}$ are all standard saddles. Close to the ground-truth solutions, the population objective is sharp, i.e., the objective can be bounded below by a constant multiple of the distance to the minimizer set. In turn, the landscape is benign: the only local minimizers are the ground-truth solutions, and all other critical points are (standard) strict saddles—which are consequently active strict saddles with an active manifold corresponding to the full domain.

Empirical landscape. The geometry of F_n differs substantially from that of its population counterpart. We establish three complementary results.

1. (**Active strict saddles**) Suppose $n \geq 2d$. Then, almost surely,

   every critical point of F_n is either a strict local minimizer or an active strict saddle.

   This dichotomy classifies the critical points, but leaves open whether every local minimizer is global. Our next result shows that this need not be the case.

2. (**Expected number of spurious minimizers**) Let $M_{d,n}$ denote the number of sharp spurious local minimizers. Here, sharpness means that the objective grows at least linearly with distance from each such minimizer in a sufficiently small neighborhood. For every fixed dimension $d \geq 4$, we show that

   $$\mathbb{E}M_{d,n} = \Omega\left(n^{d/2-1}\right) \quad \text{as } n \to \infty.$$

   Thus, the expected number of spurious minimizers grows at least polynomially with the sample size. However, this alone does not rule out their absence with high probability: the growing expectation could be driven by rare events with very many spurious minimizers.

3. (**Probability of having a spurious minimizer**) We therefore establish a stronger guarantee:

   $$\liminf_{n\to\infty} \mathbb{P}(F_n \text{ has a spurious sharp local minimum}) > 0.136, \quad \text{for any fixed } d \geq 4.$$

   This lower bound is independent of dimension. Moreover, the spurious minimizers constructed in our proof approach the population saddle set $\mathcal{S}$ as $n \to \infty$. Thus, sharp empirical local minima can accumulate at population strict saddles, even as the sample size grows.

Intuition. Our results may initially seem counterintuitive in light of the population-to-empirical landscape transfer familiar from smooth optimization. The key difference is that nonsmooth empirical perturbations can introduce linear growth that dominates the negative curvature of the population objective. To illustrate this point, consider the one-dimensional empirical objective

$$H_n(t) = -t^2 + \frac{1}{n}\sum_{i=1}^{n} \alpha_i |t|,$$

where $\alpha_1, \ldots, \alpha_n$ are i.i.d. standard Gaussian random variables. The population objective is $\mathbb{E}H_n(t) = -t^2$, so the origin is a strict maximizer and hence a strict saddle. Yet, for every finite n, the empirical coefficient $n^{-1}\sum_{i=1}^{n} \alpha_i$ is positive with probability $1/2$. On this event, the linear term dominates the negative quadratic term sufficiently close to the origin, creating a sharp local minimum. Although the perturbation vanishes as n grows, the probability of creating this local minimum remains unchanged. This illustrates the mechanism behind the sharp local minima that arise near the population saddle set $\mathcal{S}$.

Outline. The remainder of this section covers related work. Section 2 introduces the necessary background for the rest of the paper. Section 3 presents our results on the population and empirical landscapes, and Section 4 outlines the proof strategy. Section 5 studies algorithmic recovery from perturbations of the population saddle sphere. Finally, Section 6 summarizes our conclusions and discusses limitations and future directions. All technical proofs are deferred to the appendix.

Related work

Population-to-empirical transfer. For smooth losses, [31, 51] established general results for the uniform concentration of gradients and Hessians under mild assumptions. For weakly convex optimization, [16] showed the graphical convergence of subdifferentials for weakly convex stochastic

problems. Similarly, [57] studied uniform convergence of the subdifferentials for a more structured family of weakly convex functions and used it to analyze robust phase retrieval [21]. To our knowledge, our result is the first to identify a second-order obstruction in this setting. Beyond weakly convex optimization, [59] established uniform laws of large numbers for random set-valued mappings and applied them to the consistency of stationary points in nonsmooth stochastic programming. On the other hand, [63, 64] provided negative and positive results for uniform convergence of Clarke subdifferentials of Lipschitz objectives.

Nonsmooth landscapes. For symmetric rank-one matrix factorization with entrywise ℓ_1 loss, the recent works [34, 39] established that every second-order stationary point, defined through the second subderivative, is globally optimal. Moreover, [34] showed that nonglobal stationary points may lack active manifolds and that arbitrarily small linear perturbations can lead to spurious local minima. More recently, [66] developed a deterministic second-order branching criterion that rules out the existence of identifiable manifolds at critical points of finite-max composite problems. They showed that this criterion holds with high probability in overparameterized robust low-rank recovery and almost surely at common interpolators of random minimax regression, suggesting that the absence of identifiable manifolds can be typical in modern nonsmooth models. This line of work studies the failure of manifold identification at a fixed critical point, whereas our obstruction is the creation of sharp empirical local minima near a population saddle. The asymmetric analogue of this matrix factorization formulation was studied in [38], which characterized precisely when it has no spurious local minima. For orthogonal dictionary learning, [3] used uniform concentration of subdifferentials to establish recovery guarantees for a randomly initialized Riemannian subgradient method applied to a sphere-constrained ℓ_1 objective. For the nonsmooth amplitude least-squares formulation of phase retrieval, which differs from (1.1), McRae [48] establishes benign empirical landscape guarantees with constant rank overparameterization. The paper leaves open whether the original formulation without overparameterization has spurious local minima. The authors of [68] established the benign empirical landscape of a smoothed version of (1.1) with a sufficiently small fixed positive smoothing parameter. Intuitively, this landscape seems to smooth out the sharp kinks we identify in the purely nonsmooth objective. The second-order analysis and possible existence of spurious local minima for robust phase retrieval (1.1) were explicitly identified as relevant unresolved problems in [34, 68].

Nonsmooth algorithmic guarantees. Sharpness, together with Lipschitzness and weak convexity, yields fast local convergence of subgradient and prox-linear methods [12, 20, 26, 54]. These convergence results require initialization near the solution set and do not deal with stationary points outside that region. Beyond local convergence, [37] developed a smoothing algorithm to find approximate second-order stationary points of a smooth population risk from a sufficiently accurate, potentially nonsmooth uniform approximation. To our knowledge, all general-purpose guarantees for convergence to second-order critical points of nonsmooth objectives rely on active strict saddles [4, 14, 17, 19, 55]. Our results suggest that we need to reevaluate this algorithmic explanation if we wish to explain the empirical success of first-order methods on statistical recovery problems.

2 Preliminaries

In this section, we introduce the notation and background required for the rest of the paper.

Basic notation. We use $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ for the Euclidean inner product and norm, and $\|\cdot\|_{\mathrm{op}}$ for the operator norm. Set addition and scalar multiplication are understood elementwise: $A + B = \{a + b :$

$a \in A, b \in B\}$ and $Sv = \{sv : s \in S\}$ for $S \subset \mathbb{R}$. Given a differentiable map $g\colon \mathbb{R}^d \to \mathbb{R}^m$, we use ∇g to denote its Jacobian. When $m = 1$, this corresponds to the gradient map, and, further, when g is a twice differentiable function, the symbol $\nabla^2 g$ denotes its Hessian map.

Manifolds. Our treatment of manifolds here is rather brief since we will only focus on Euclidean manifolds, i.e., affine subspaces. We refer the interested reader to the textbooks by Boumal [9] and Lee [42]. A C^p manifold $\mathcal{M} \subset \mathbb{R}^d$ is locally a level set $G^{-1}(0)$ of a C^p map whose Jacobian has full row rank. Its tangent space at x is $T_{\mathcal{M}}(x) = \ker \nabla G(x)$. When $\mathcal{M}$ is an affine subspace, $T_{\mathcal{M}}(x)$ is equal to the same affine subspace shifted to the origin. We endow $\mathcal{M}$ with a Riemannian inner product induced by the Euclidean inner product. Given a C^2 function $g\colon \mathcal{M} \to \mathbb{R}$, we use the symbol $\nabla^2_{\mathcal{M}} g(x)$ to denote the Riemannian Hessian at x. When $\mathcal{M}$ is affine, $\nabla^2_{\mathcal{M}} g(x)$ is represented by $V^\top \nabla^2 \widetilde{g}(x) V$, where the columns of V form an orthonormal basis of $T_{\mathcal{M}}(x)$ and $\widetilde{g}$ is any C^2 extension of g to a neighborhood of x.

Nonsmooth analysis. A proper closed function $f\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is ρ-weakly convex if $x \mapsto f(x) + \frac{\rho}{2}\|x\|^2$ is convex, where $\rho \geq 0$. For $x \in \operatorname{dom} f$, its subdifferential $\partial f(x)$ consists of the vectors $v \in \mathbb{R}^d$ satisfying

$$f(y) \geq f(x) + \langle v, y - x\rangle - \frac{\rho}{2}\|y - x\|^2 \qquad \text{for all } y \in \mathbb{R}^d.$$

We call x a critical point if $0 \in \partial f(x)$. This class includes compositions $h \circ G$ with h convex and L-Lipschitz and G having a β-Lipschitz Jacobian: such functions are $L\beta$-weakly convex [20]. For the empirical objective F_n in (1.1), we show in Theorem 3.2 that F_n is $2\left\|\frac{1}{n}\sum_{i=1}^n a_i a_i^\top\right\|_{\mathrm{op}}$-weakly convex for every realization of the measurements.

Nonsmooth optimization admits several different notions of second order information [56]. We adopt the notion of an active strict saddle introduced by Davis and Drusvyatskiy [17], which underlies saddle-avoidance guarantees for several first-order methods [4, 19]. To introduce this notion, we require an auxiliary definition.

Definition 2.1 (Active manifold)**.** *Consider a closed, weakly convex function f, and let $\bar{x}$ be a critical point. A set $\mathcal{M}$ containing $\bar{x}$ is an* active C^p manifold *at $\bar{x}$ if there is some neighborhood U of $\bar{x}$ for which the following holds.*

1. *(**Smoothness**) The set $\mathcal{M} \cap U$ is a C^p manifold and the restriction $f|_{\mathcal{M}\cap U}$ is C^p-smooth.*
2. *(**Sharpness**) Subgradients are bounded away from zero off the manifold,*

$$\inf\{\|v\| : v \in \partial f(x),\ x \in U \setminus \mathcal{M}\} > 0.$$

The notion of "active sets" has a long history in nonsmooth optimization [1, 10, 11, 24, 27, 28, 30, 58]. Beyond the definition we follow here, this notion has been modeled in a variety of ways, including identifiable surfaces [67], partial smoothness [44], $\mathcal{VU}$-decompositions [43], and cone reducibility [6]. Furthermore, these notions have led to a range of algorithmic developments [18, 22, 40, 45, 46, 52]. Armed with Definition 2.1 we can now introduce active strict saddles.

Definition 2.2 (Active strict saddles)**.** *A critical point $\bar{x}$ of a weakly convex function f is an* active strict saddle *if it admits an active C^2 manifold $\mathcal{M}$ such that $\nabla^2_{\mathcal{M}}(f|_{\mathcal{M}})(\bar{x})$ has a negative eigenvalue.*[1] *A function f exhibits the* active strict saddle property *if every critical point is either a local minimizer or an active strict saddle.*

[1] This is equivalent to the parabolic-subderivative definition in [17]; see [4, Definition 4 and footnote 2].

Intuitively, at an active strict saddle, the function has a direction of negative curvature along the active manifold and grows sharply away from it. The existence of an active manifold appears crucial for saddle avoidance by simple iterative methods: without one, subgradient methods may converge to critical points that are not local minimizers [17]. Finally, a function $f\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ with a minimizer set $\mathcal{X}_\star$ is sharp if there exists a constant $\mu > 0$ such that

$$f(x) - \inf f \geq \mu \operatorname{dist}(x, \mathcal{X}_\star) \quad \text{for all } x \in \mathbb{R}^d.$$

A point $\bar{x} \in \operatorname{dom} f$ is a *sharp local minimizer* if there exist constants $\mu, r > 0$ such that

$$f(x) - f(\bar{x}) \geq \mu \|x - \bar{x}\| \qquad \text{whenever } \|x - \bar{x}\| < r.$$

Whenever f is ρ-weakly convex, μ-sharp and L-Lipschitz, then several first order methods converge fast towards a minimum provided they are initialized near enough [20, 25].

3 The second-order nonsmooth landscape of phase retrieval

In this section, we formally describe our results on the second-order landscape of robust phase retrieval. The first subsection covers the population objective, while the second subsection covers the empirical objective. Throughout, we make the blanket assumption that the signal we wish to recover has norm one, i.e., $\|x^\star\| = 1$. The assumption is made only for convenience, as it simplifies expressions that would otherwise depend on its norm.

3.1 Population landscape

We start by showing that F is C^2-smooth everywhere but at the solutions and that each critical point is either a global minimum or a saddle. The proof of this proposition appears in Appendix A.

Proposition 3.1. *Fix a unit-norm vector $x^\star \in \mathbb{R}^d$ and consider the population objective $F(x) = \mathbb{E}|\langle a, x\rangle^2 - \langle a, x^\star\rangle^2|$ with $a \sim N(0, I_d)$. Then, $F(\cdot)$ is a closed, weakly convex function, and its critical points correspond to*

$$\{\pm x^\star\} \cup \{0\} \cup \mathcal{S} \qquad \text{with} \qquad \mathcal{S} \coloneqq \big\{\tau u \mid \|u\| = 1 \text{ and } u \perp x^\star\big\}, \tag{3.1}$$

where τ is the unique solution of the equation $\arctan\tau + \tau/(1+\tau^2) = \pi/4$. The function F is C^2 on $\mathbb{R}^d \setminus \{\pm x^\star\}$, and its Hessian evaluates to

$$\nabla^2 F(0) = -2I_d, \quad \text{and} \quad \nabla^2 F(\tau u) = \frac{16}{\pi} \cdot \frac{\tau}{(1+\tau^2)^2} \cdot (uu^\top - x^\star x^{\star\top}) \quad \text{for all } \tau u \in \mathcal{S}.$$

Further, F is $2/\pi$-sharp with global minimizers $\pm x^\star$.

The characterization of critical points in (3.1) was first established in [21]. As direct consequences, the origin is a strict local maximizer, while each point $\tau u \in \mathcal{S}$ is a saddle with positive curvature in the radial direction u, negative curvature along $x^\star$, and zero curvature in directions tangent to $\mathcal{S}$. Since F is C^2 near these nonglobal critical points, each is also an active strict saddle, with active manifold equal to the whole space $\mathbb{R}^d$. Together with the sharpness of the global minimizers $\pm x^\star$, this establishes the benign geometry of the population landscape.

3.2 Empirical landscape

In this subsection, we state our main results for the empirical objective F_n, defined in (1.1). To describe these guarantees, we first define the active index set at a vector $x \in \mathbb{R}^d$ as

$$I(x) = \big\{i \in \{1, \ldots, n\} : \langle a_i, x\rangle^2 = \langle a_i, x^\star\rangle^2\big\}. \tag{3.2}$$

As it turns out, when $n \geq 2d$, almost surely every critical point $x \notin \{\pm x^\star\}$ has an active manifold encoded via $I(x)$. The proof of the following theorem appears in Appendix B.

Theorem 3.2 (Active strict saddle property)**.** *Fix a unit-norm vector $x^\star \in \mathbb{R}^d$. The empirical objective F_n in (1.1) is $2\left\|\frac{1}{n}\sum_{i=1}^n a_i a_i^\top\right\|_{\mathrm{op}}$-weakly convex for any realization of the measurements $\{a_i\}$. If $n \geq 2d$, then almost surely every critical point $x_\sharp$ of F_n is either a strict local minimizer or an active strict saddle. In the latter case, the active manifold is*

$$\mathcal{M}_{x_\sharp} = \left\{x \in \mathbb{R}^d \mid \langle a_i, x\rangle = \langle a_i, x_\sharp\rangle \text{ for all } i \in I(x_\sharp)\right\}.$$

Let us make a few remarks. This guarantee, in tandem with the saddle-avoidance results from [4, 19], shows that certain stochastically perturbed variants of the subgradient method can converge only to local minimizers of the empirical objective F_n. Since any subset of d Gaussian vectors $\{a_i\}$ is almost surely linearly independent, at every critical point $x_\sharp \notin \{\pm x^\star\}$ the index set satisfies $|I(x_\sharp)| \leq d$, and moreover, when equality holds, the manifold reduces to a single point and $x_\sharp$ is a sharp local minimizer. This simple observation will be key in our later developments.

Theorem 3.2, together with the benign population landscape described in Proposition 3.1, might give the impression that the empirical objective exhibits benign geometry as well. The next result shows that this is not the case. Recall that $M_{d,n}$ denotes the random variable that counts the number of spurious sharp local minima of F_n. We defer the proof of this guarantee to Appendix C.

Theorem 3.3 (**Existence of spurious sharp local minima**)**.** *Fix a unit-norm vector $x^\star \in \mathbb{R}^d$, consider the empirical objective F_n in (1.1), and suppose that $d \geq 4$. Then, there exists a positive constant c_d depending solely on d such that the following hold true.*

1. *(**Bounded moments**) The first moment is asymptotically bounded by*

$$\liminf_{n\to\infty} \frac{\mathbb{E}M_{d,n}}{n^{d/2-1}} \geq c_d.$$

2. *(**Constant probability**) Furthermore,*

$$\liminf_{n\to\infty} \mathbb{P}(F_n \text{ has a sharp spurious local minimum}) \geq 0.136.$$

Remarkably, the limiting probability lower bound is uniform over $d \geq 4$. Our proof constructs these sharp local minima within a shrinking tube around the population saddle sphere $\mathcal{S}$, so their distance to this saddle set tends to zero as n increases. The next section outlines the proof strategy. Section 5 studies algorithmic recovery from perturbations of $\mathcal{S}$.

4 Proof strategy

In this section, we sketch our proof strategy for establishing Theorem 3.3. A full proof appears in Appendix C. Before starting, let us provide a roadmap. The key idea is to define a random variable $N_{d,n}$ that counts a selected family of sharp spurious local minima inside a tube, denoted T_n, around the sphere of saddles $\mathcal{S}$. This tube shrinks as n goes to infinity. Since $M_{d,n}$ counts all sharp spurious local minima, it follows that

$$M_{d,n} \geq N_{d,n}, \qquad \text{and so} \qquad \mathbb{P}(M_{d,n} > 0) \geq \mathbb{P}(N_{d,n} > 0). \tag{4.1}$$

Thus, it suffices to lower bound the probability that our selected family contains at least one minimum. The proof progresses in three steps: (i) define the family counted by $N_{d,n}$, (ii) compute the asymptotic first moment of $N_{d,n}$, (iii) compute the second moment of $N_{d,n}$ and use the Paley–Zygmund inequality to obtain the desired probability bound. Throughout, we assume that $d \geq 4$ is fixed, and let n tend to infinity.

A family of spurious sharp minimizers. The following lemma, proved in Appendix B.1, characterizes critical points and sharp local minima. Recall the active index set $I(x)$ in (3.2).

Lemma 4.1 (Characterization of critical points and sharp local minima)**.** *Suppose $n \geq 2d$. Then, almost surely, every $x \in \mathbb{R}^d \setminus \{\pm x^\star\}$ satisfies $|I(x)| \leq d$. Such a point is critical if, and only if, there exist scalars $\lambda_i \in (-1, 1)$ with $i \in I(x)$, satisfying*

$$0 = \sum_{i \notin I(x)} \operatorname{sign}(\langle a_i, x\rangle^2 - \langle a_i, x^\star\rangle^2)\, \langle a_i, x\rangle\, a_i + \sum_{i \in I(x)} \lambda_i \langle a_i, x\rangle\, a_i. \tag{4.2}$$

Moreover, a critical point x is a sharp local minimum if, and only if, it is critical and $|I(x)| = d$.

The lemma implies a simple way to count sharp spurious local minima: enumerate the candidates determined by their active indices and check which are critical. We will use this strategy to bound $N_{d,n}$, but first we will describe the family of sharp local minima that $N_{d,n}$ counts. For $I = \{i_1 < \cdots < i_d\} \subset \{1, \ldots, n\}$ and $s \in \{\pm 1\}^d$, the active equations $a_{i_j}^\top x = s_j a_{i_j}^\top x^\star$, almost surely, determine a unique candidate

$$x_{I,s} = A_I^{-1} \operatorname{diag}(s) A_I x^\star. \tag{4.3}$$

Here A_I is the matrix with rows $a_{i_j}^\top$. To define $N_{d,n}$, we fix the pattern $s^\circ = (1, \ldots, 1, -1, -1)$, with $d-2$ plus signs, and write $x_I = x_{I,s^\circ}$. Counting sharp minima with only this pattern suffices for a lower bound as in (4.1). Yet, we will count an even smaller subset of these x_I.

We express the criticality condition (4.2) in a form that will help us estimate the probability that a candidate x_I is a sharp local minimum. Define $S_I(x) = \sum_{i \notin I} \operatorname{sign}((a_i^\top x)^2 - (a_i^\top x^\star)^2)(a_i^\top x) a_i$ and let $Q_I(x)$ have columns $(a_i^\top x) a_i$, $i \in I$, in increasing index order. By Lemma 4.1, x_I is a sharp local minimum precisely when

$$-S_I(x_I) \in Q_I(x_I)(-1, 1)^d, \tag{4.4}$$

where the set on the right consists of the vectors $Q_I(x_I)\lambda$ with $\lambda \in (-1, 1)^d$.

To make the probability of (4.4) easier to estimate, we impose two further restrictions. First, we require the selected measurement vectors $\{a_i\}_{i \in I}$ to have bounded norms and the determinants $|\det A_I|$ and $|\det Q_I(x_I)|$ to be bounded below by positive constants. These bounds prevent the equations defining x_I from becoming nearly dependent and ensure that the set on the right-hand side of (4.4) has volume bounded away from zero. The constants depend only on d and are chosen so that these bounds hold with positive probability. We do not include them here since their precise values are not particularly insightful. Second, we count only candidates in the tube

$$T_n = \left\{ n^{-1/2} z_1 x^\star + (\tau + n^{-1/2} z_2) u \mid u \perp x^\star,\ \|u\| = 1,\ \text{and } z \in [-1, 1]^2 \right\}. \tag{4.5}$$

The set T_n is a tube around the population saddle sphere $\mathcal{S}$ that shrinks to $\mathcal{S}$ as $n \to \infty$. The width $n^{-1/2}$ matches the fluctuations exhibited by critical points of the empirical objective [57].

The first moment. For any fixed set I, the probability that x_I is counted in the family described above is exactly the same, thus

$$\mathbb{E} N_{d,n} = \sum_I \mathbb{P}(x_I \text{ is counted}) = \binom{n}{d} \mathbb{P}(x_I \text{ is counted}). \tag{4.6}$$

We therefore fix one set I and focus on computing this probability. Our plan to do so is to integrate over the possible locations of the candidate x_I. For a proposed location x, write $X_I(x) = A_I x - \operatorname{diag}(s^\circ) A_I x^\star$, so that $X_I(x) = 0$ precisely when $x = x_I$. Let $H_I(x)$ be a boolean function that is equal to one precisely when the selected measurements a_i satisfy our bounds on their norms and associated determinants, and the condition (4.4) holds with x_I replaced by x, and zero otherwise.

The Kac–Rice formula [2] yields the following identity

$$\mathbb{P}(x_I \text{ is counted}) = \int_{T_n} p_{X_I(x)}(0)\, \mathbb{E}\big[|\det A_I| H_I(x) \;\big|\; X_I(x) = 0\big]\, \mathrm{d}x, \tag{4.7}$$

where $p_{X_I(x)}(0)$ is the probability density of $X_I(x)$ evaluated at zero.

To describe the asymptotic value of (4.7), we establish the following two estimates, with the first holding uniformly over $x \in T_n$,

$$\mathbb{E}\big[|\det A_I| H_I(x) \;\big|\; X_I(x) = 0\big] \asymp_d n^{-d/2}, \qquad \text{and} \qquad \int_{T_n} p_{X_I(x)}(0) dx \asymp_d n^{-1}. \tag{4.8}$$

Here $\asymp_d$ denotes comparison up to positive constants depending only on d. Before explaining the rationale behind these two results, let us derive an immediate consequence. Recall that $\binom{n}{d} \asymp_d n^d$; combining this fact with (4.6), (4.7), and (4.8) yields $\mathbb{E}N_{d,n} \asymp_d n^d \cdot n^{-d/2} \cdot n^{-1} = n^{d/2-1}$. This establishes the asymptotic lower bound in Theorem 3.3. A detailed computation accounting for the hidden constants yields two positive constants γ_d and $\mathcal{I}_d$ such that

$$n^{-d/2+1}\mathbb{E}N_{d,n} \to \gamma_d \mathcal{I}_d \qquad \text{as } n \to \infty. \tag{4.9}$$

Here $\gamma_d > 0$ accounts for the selected measurements, while $\mathcal{I}_d > 0$ accounts for the Gaussian density across the tube. Their precise expressions and the full calculation appear in Appendix C.2.

Let us quickly summarize the arguments behind (4.8). For a fixed location $x \in T_n$, we condition further on the selected measurement vectors $\{a_i\}_{i\in I}$. This fixes the set $Q_I(x)(-1,1)^d$, while the remaining measurements retain their independent standard Gaussian distributions. When our norm and determinant bounds hold, the determinant factor $|\det A_I|$ is bounded above and below by positive constants independent of n, and the set $Q_I(x)(-1,1)^d$ is bounded with volume bounded away from zero. The condition $H_I(x) = 1$ then reduces to asking whether $-S_I(x)$ lies in this fixed set. Each summand in $S_I(x)$ has mean $\frac{1}{2}\nabla F(x)$. Since ∇F vanishes on $\mathcal{S}$ and is Lipschitz continuous nearby, this mean has norm $O(n^{-1/2})$ on T_n, whose points lie within distance $O(n^{-1/2})$ of $\mathcal{S}$. Summing the $n-d$ terms, the mean of $S_I(x)$ therefore has norm of order at most $\sqrt{n}$, uniformly over $x \in T_n$. We can show that its standard deviation in every unit direction is also of order $\sqrt{n}$, uniformly over $x \in T_n$. A local limit theorem implies a Gaussian approximation to the probability of entering this region, yielding a probability of order $n^{-d/2}$ uniformly over the selected measurements satisfying our bounds and the locations in T_n. The first estimate in (4.8) then follows.

For the second estimate in (4.8), we consider the volume of T_n. The tube restricts the component along $x^\star$ to within $n^{-1/2}$ of zero and the length of the perpendicular component to within $n^{-1/2}$ of τ. These two restrictions give its volume order n^{-1}. We can prove that the density $p_{X_I(x)}(0)$ is bounded above and below by positive constants on T_n, so its integral has order n^{-1}.

A probability bound uniform in dimension. We use the Paley–Zygmund inequality to turn moment estimates into a lower bound on the probability that $N_{d,n} > 0$, namely

$$\liminf_{n\to\infty} \mathbb{P}(N_{d,n} > 0) \geq \lim_{n\to\infty} \frac{(\mathbb{E}N_{d,n})^2}{\mathbb{E}N_{d,n}^2}. \tag{4.10}$$

Thus, to derive our lower bound we also asymptotically characterize $\mathbb{E}N_{d,n}^2$. Computing the second moment follows a similar Kac–Rice approach as for the first moment, but the dependence between candidate pairs (x_I, x_J) makes the argument significantly more technical. We defer the proof to Appendix C.3 and simply state here the resulting estimate

$$n^{-d+2}\mathbb{E}N_{d,n}^2 \to \gamma_d^2 \mathcal{J}_d \qquad \text{as } n \to \infty,$$

where $\mathcal{J}_d > 0$ is finite and depends on d. Combining this estimate with (4.9) and (4.10) implies

$$\liminf_{n\to\infty} \mathbb{P}(N_{d,n} > 0) \geq \frac{\mathcal{I}_d^2}{\mathcal{J}_d},$$

the γ_d factor cancels exactly. By arguing directly with the definitions of $\mathcal{I}_d$ and $\mathcal{J}_d$ we show that this ratio can be further lower bounded by an explicit quantity $c_\star$ that is independent of the dimension. Combining analytic estimates with certified numerical computations, as described in Appendix C.5, we show that the value of $c_\star$ is lower bounded by 0.1361.

5 Numerical experiments

In this section, we numerically test our theory. The objective of our experiments is two-fold: (*i*) to show that for moderate d and n, we observe empirically that the probability of having a spurious sharp minimum is large and (*ii*) to show that despite our negative results, simple subgradient methods do not often get trapped at these minima. We present two numerical experiments supporting these claims. For both experiments we set $x^\star = e_1$, the first canonical vector. The code to reproduce these experiments can be found at https://github.com/PenghaoWang1225/Phase-Retrieval.

Probability of having a spurious sharp minimum. We estimate the probability that F_n has a sharp spurious minimum for $d = 4$ and $n \in \{12, 16, 24, 32, 48, 64\}$. For each n, we draw 100 independent problem instances with n i.i.d. Gaussian measurement vectors. For each instance, we enumerate all pairs (I, s) with $|I| = d$ and nonconstant $s \in \{\pm 1\}^d$, and test the candidates $x_{I,s}$ defined in (4.3). We count an instance as positive if at least one candidate satisfies the criticality condition (4.2) up to the numerical tolerances specified in Appendix G. Each estimate is the fraction of positive instances, and the shaded band shows 95% confidence intervals.

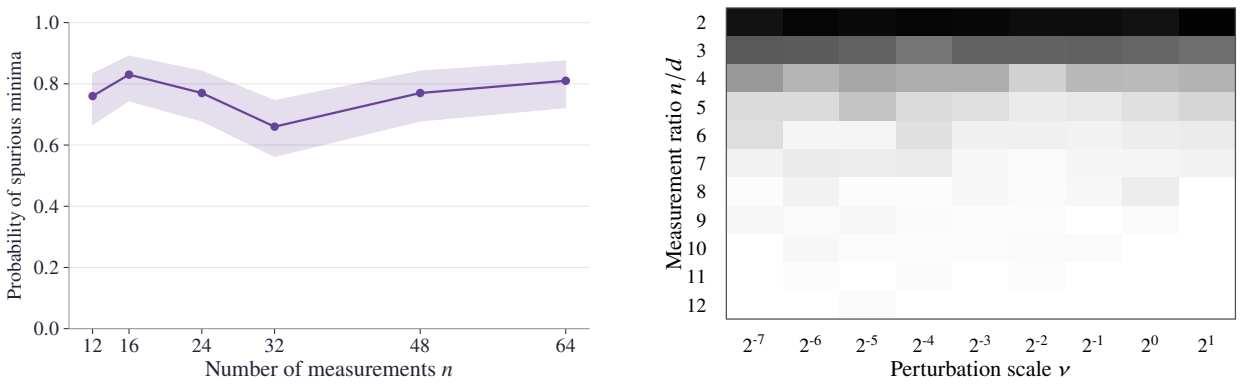
Figure 2(a) shows that the estimates change little with n and remain above 0.5 across all tested settings. These results are qualitatively consistent with Theorem 3.3, but suggest that its asymptotic lower bound of 0.136 is conservative. This may reflect the fact that our proof counts only a strict subset of the sharp spurious minima.

Recovery near the saddle sphere. For this experiment, we estimate how often a perturbed subgradient method recovers $x^\star$ up to sign. We fix $d = 100$ and vary the measurement ratio $n/d \in \{2, 3, \ldots, 12\}$ and perturbation radius $\nu \in \{2^{-7}, \ldots, 2^1\}$. For each measurement ratio, we conduct 100 independent trials. In each trial, we draw n i.i.d. Gaussian measurement vectors. For each radius ν, we draw fresh independent $u \sim \mathrm{Unif}(\mathcal{S})$ and $v \sim \mathrm{Unif}(\mathbb{S}^{d-1})$, then initialize $x_0 = u + \nu v$. We run the stochastically perturbed subgradient method analyzed in [19] for at most 5×10^5 updates per initialization, with updates

$$x_{k+1} = x_k - \frac{0.05}{(k+1)^{0.6}} \big(g(x_k) + 0.2\xi_k\big),$$

where $g(x_k) \in \partial F_n(x_k)$ and the vectors $\xi_k \sim \mathrm{Unif}(\mathbb{S}^{d-1})$ are independent. We count a run as recovered if an updated iterate satisfies both $\mathrm{dist}(x_k, \{\pm e_1\}) \leq 10^{-3}$ and $F_n(x_k) \leq 10^{-4}$.

Figure 2(b) suggests that the perturbation radius $\nu = \|x_0 - u\|$ has little effect on recovery frequency over the tested range. The measurement ratio n/d has a stronger effect, with a sharp increase in recovery near $n/d = 4$. Thus, sharp spurious minima near the saddle $\mathcal{S}$ do not preclude high recovery rates for this method for a sufficiently large number of samples.



(a) (b)

Figure 2: (a) Estimated probability of a sharp spurious minimum. (b) Recovery frequency of the perturbed subgradient method. White denotes a recovery fraction of one and black denotes zero.

6 Conclusion and limitations

Our main result shows that even for highly structured weakly convex stochastic objectives, benign population geometry need not transfer to empirical risk minimization. As a consequence, active strict saddle theory alone cannot explain the practical success of iterative methods applied to robust phase retrieval. We conjecture that a similar phenomenon occurs in other nonsmooth formulations for low-rank matrix recovery problems. This opens two directions for future research: (*i*) to understand which other geometric properties enable these iterative methods to find global solutions for these nonsmooth formulations, and (*ii*) to characterize general geometric conditions of weakly convex stochastic problems under which population-to-empirical transfer holds. Our analysis has several limitations: it covers only noiseless real Gaussian measurements and takes $n \to \infty$ at each fixed dimension. Although the limiting probability bound is independent of d, our proof does not quantify how large n must be relative to d or characterize regimes in which both grow. Extending the analysis to noisy or non-Gaussian measurements would clarify the scope of this obstruction.

7 Use of AI in proof discovery

We established Proposition 3.1 and Theorem 3.2 in 2021. After failing to prove a population-to-empirical landscape transfer, we questioned whether such a result could hold and conjectured instead that the empirical objective admits spurious sharp local minima. We did not make this work public at the time because the conjecture remained unresolved.

In August 2026, we began interacting with ChatGPT 5.6 Sol and asked it to test the conjecture numerically. The resulting experiments convinced us that the conjecture was true, and we began working with the model on a proof. At that stage, we did not expect to obtain a positive probability lower bound independent of d as $n \to \infty$ for every fixed $d \geq 4$.

In the subsequent proof discussions, ChatGPT proposed counting spurious sharp local minima in a tube of width $O(n^{-1/2})$ around the population saddle sphere. The preliminary argument only yielded a probability lower bound that tended to zero as $n \to \infty$. We then supplied this argument to `qmd-prover`, a proof-development harness developed by Xiao Ma.[2] We used separate instances of ChatGPT 5.6 Sol with the ultra reasoning setting as prover and verifier. These interactions produced

[2]The repository and its creator are not affiliated with the authors.

draft first- and second-moment arguments based on a weighted Kac–Rice formula and local limit theorems, leading to the dimension-independent probability bound in Theorem 3.3.

In later interactions with GPT-5.6 and GPT-6, we checked the hypotheses of the probabilistic tools, filled in details for special cases, and reorganized the dependencies among the lemmas. The authors parsed, completed, and simplified the resulting arguments and rewrote all proofs in the final manuscript with human readability in mind. Beyond proof discovery, we used ChatGPT 6.0 Astra to proofread the manuscript, polish the writing, and implement the numerical experiments. We reviewed all AI-assisted work and take full responsibility for the content of this paper.

References

[1] F. Al-Khayyal and J. Kyparisis. Finite convergence of algorithms for nonlinear programs and variational inequalities. *Journal of Optimization Theory and Applications*, 70(2):319–332, 1991.

[2] Jean-Marc Azaïs and Mario Wschebor. *Level Sets and Extrema of Random Processes and Fields*. John Wiley & Sons, 2009.

[3] Yu Bai, Qijia Jiang, and Ju Sun. Subgradient descent learns orthogonal dictionaries. In *International Conference on Learning Representations*, 2019.

[4] Pascal Bianchi, Walid Hachem, and Sholom Schechtman. Stochastic subgradient descent escapes active strict saddles on weakly convex functions. *Mathematics of Operations Research*, 49(3):1761–1790, 2024.

[5] Sergey G. Bobkov. Refinements of Berry–Esseen inequalities in terms of Lyapunov coefficients. *Journal of Fourier Analysis and Applications*, 29(6):72, 2023.

[6] J. Frédéric Bonnans and Alexander Shapiro. *Perturbation Analysis of Optimization Problems*. Springer, New York, 2000.

[7] Alexander Alekseevich Borovkov and Anatolii Al'fredovich Mogul'skii. Integro-local theorems for sums of independent random vectors in the series scheme. *Mathematical Notes*, 79(3–4):468–482, 2006.

[8] Nicolas Boumal. Nonconvex phase synchronization. *SIAM Journal on Optimization*, 26(4):2355–2377, 2016.

[9] Nicolas Boumal. *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, 2023.

[10] James V. Burke and Jorge J. Moré. On the identification of active constraints. *SIAM Journal on Numerical Analysis*, 25(5):1197–1211, 1988.

[11] Jim Burke. On the identification of active constraints II: The nonconvex case. *SIAM Journal on Numerical Analysis*, 27(4):1081–1102, 1990.

[12] Vasileios Charisopoulos, Yudong Chen, Damek Davis, Mateo Díaz, Lijun Ding, and Dmitriy Drusvyatskiy. Low-rank matrix recovery with composite optimization: Good conditioning and rapid convergence. *Foundations of Computational Mathematics*, 21(6):1505–1593, 2021.

[13] Vasileios Charisopoulos, Damek Davis, Mateo Díaz, and Dmitriy Drusvyatskiy. Composite optimization for robust rank one bilinear sensing. *Information and Inference: A Journal of the IMA*, 10(2):333–396, 2021.

[14] Damek Davis, Mateo Díaz, and Dmitriy Drusvyatskiy. Escaping strict saddle points of the Moreau envelope in nonsmooth optimization. *SIAM Journal on Optimization*, 32(3):1958–1983, 2022.

[15] Damek Davis and Dmitriy Drusvyatskiy. Stochastic model-based minimization of weakly convex functions. *SIAM Journal on Optimization*, 29(1):207–239, 2019.

[16] Damek Davis and Dmitriy Drusvyatskiy. Graphical convergence of subgradients in nonconvex optimization and learning. *Mathematics of Operations Research*, 47(1):209–231, 2022.

[17] Damek Davis and Dmitriy Drusvyatskiy. Proximal methods avoid active strict saddles of weakly convex functions. *Foundations of Computational Mathematics*, 22(2):561–606, 2022.

[18] Damek Davis, Dmitriy Drusvyatskiy, and Liwei Jiang. Asymptotic normality and optimality in nonsmooth stochastic approximation. *The Annals of Statistics*, 52(4):1485–1508, 2024.

[19] Damek Davis, Dmitriy Drusvyatskiy, and Liwei Jiang. Active manifolds, stratifications, and convergence to local minima in nonsmooth optimization. *Foundations of Computational Mathematics*, 26(2):779–861, 2026.

[20] Damek Davis, Dmitriy Drusvyatskiy, Kellie J. MacPhee, and Courtney Paquette. Subgradient methods for sharp weakly convex functions. *Journal of Optimization Theory and Applications*, 179(3):962–982, 2018.

[21] Damek Davis, Dmitriy Drusvyatskiy, and Courtney Paquette. The nonsmooth landscape of phase retrieval. *IMA Journal of Numerical Analysis*, 40(4):2652–2695, 2020.

[22] Damek Davis and Liwei Jiang. A local nearly linearly convergent first-order method for nonsmooth functions with quadratic growth. *Foundations of Computational Mathematics*, 25(3):943–1024, 2025.

[23] Mateo Díaz. The nonsmooth landscape of blind deconvolution. In *11th Annual Workshop on Optimization for Machine Learning*, 2019.

[24] Dmitriy Drusvyatskiy and Adrian S Lewis. Optimality, identifiability, and sensitivity. *Mathematical Programming*, 147(1–2):467–498, 2014.

[25] Dmitriy Drusvyatskiy and Courtney Paquette. Efficiency of minimizing compositions of convex functions and smooth maps. *Mathematical Programming*, 178:503–558, 2019.

[26] John C. Duchi and Feng Ruan. Solving (most) of a set of quadratic equalities: composite optimization for robust phase retrieval. *Information and Inference: A Journal of the IMA*, 8(3):471–529, 2019.

[27] J. C. Dunn. On the convergence of projected gradient processes to singular critical points. *Journal of Optimization Theory and Applications*, 55(2):203–216, 1987.

[28] Michael C. Ferris. Finite termination of the proximal point algorithm. *Mathematical Programming*, 50:359–366, 1991.

[29] James R. Fienup. Phase retrieval algorithms: a comparison. *Applied Optics*, 21(15):2758–2769, 1982.

[30] Sjur D. Flåm. On finite convergence and constraint identification of subgradient projection methods. *Mathematical Programming*, 57:427–437, 1992.

[31] Dylan J. Foster, Ayush Sekhari, and Karthik Sridharan. Uniform convergence of gradients for non-convex learning and optimization. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

[32] Rong Ge, Furong Huang, Chi Jin, and Yang Yuan. Escaping from saddle points—online stochastic gradient for tensor decomposition. In *Proceedings of the 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pages 797–842. PMLR, 2015.

[33] Rong Ge, Chi Jin, and Yi Zheng. No spurious local minima in nonconvex low rank problems: A unified geometric analysis. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1233–1242. PMLR, 2017.

[34] Jiewen Guan and Anthony Man-Cho So. ℓ_1-norm rank-one symmetric matrix factorization has no spurious second-order stationary points. *arXiv preprint arXiv:2410.05025*, 2024.

[35] Mourad E. H. Ismail and Jim Pitman. Algebraic evaluations of some Euler integrals, duplication formulae for Appell's hypergeometric function F_1, and Brownian variations. *Canadian Journal of Mathematics*, 52(5):961–981, 2000.

[36] Chi Jin, Rong Ge, Praneeth Netrapalli, Sham M. Kakade, and Michael I. Jordan. How to escape saddle points efficiently. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1724–1732. PMLR, 2017.

[37] Chi Jin, Lydia T. Liu, Rong Ge, and Michael I. Jordan. On the local minima of the empirical risk. In *Advances in Neural Information Processing Systems*, volume 31, pages 4896–4905, 2018.

[38] Cédric Josz and Lexiao Lai. Nonsmooth rank-one matrix factorization landscape. *Optimization Letters*, 16(6):1611–1631, 2022.

[39] Cédric Josz and Lexiao Lai. Nonsmooth rank-one symmetric matrix factorization landscape. *Optimization Letters*, 19(8):1691–1698, 2025.

[40] Siyu Kong and Adrian S. Lewis. Lipschitz minimization and the Goldstein modulus. *Mathematical Programming*, 219:315–344, 2026.

[41] Jason D. Lee, Ioannis Panageas, Georgios Piliouras, Max Simchowitz, Michael I. Jordan, and Benjamin Recht. First-order methods almost always avoid strict saddle points. *Mathematical Programming*, 176(1–2):311–337, 2019.

[42] John M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, New York, NY, second edition, 2012.

[43] Claude Lemaréchal, François Oustry, and Claudia Sagastizábal. The $\mathcal{U}$-lagrangian of a convex function. *Transactions of the American Mathematical Society*, 352(2):711–729, 2000.

[44] Adrian S. Lewis. Active sets, nonsmoothness, and sensitivity. *SIAM Journal on Optimization*, 13(3):702–725, 2002.

[45] Adrian S. Lewis and Calvin Wylie. Active-set Newton methods and partial smoothness. *Mathematics of Operations Research*, 46(2):712–725, 2021.

[46] Jingwei Liang, Jalal Fadili, and Gabriel Peyré. Activity identification and local linear convergence of Forward–Backward-type methods. *SIAM Journal on Optimization*, 27(1):408–437, 2017.

[47] Andrew D. McRae. Phase retrieval and matrix sensing via benign and overparametrized nonconvex optimization. *IEEE Transactions on Information Theory*, 72(6):4203–4220, 2026.

[48] Andrew D. McRae. Phase retrieval via overparametrized nonconvex optimization: nonsmooth amplitude loss landscapes. *IEEE Transactions on Information Theory*, 72(9):7049–7068, 2026.

[49] Andrew D. McRae, Pedro Abdalla, Afonso S. Bandeira, and Nicolas Boumal. Nonconvex landscapes for $\mathbf{Z}_2$ synchronization and graph clustering are benign near exact recovery thresholds. *Information and Inference: A Journal of the IMA*, 14(2):iaaf012, 2025.

[50] Andrew D. McRae and Nicolas Boumal. Benign landscapes of low-dimensional relaxations for orthogonal synchronization on general graphs. *SIAM Journal on Optimization*, 34(2):1427–1454, 2024.

[51] Song Mei, Yu Bai, and Andrea Montanari. The landscape of empirical risk for nonconvex losses. *The Annals of Statistics*, 46(6A):2747–2774, 2018.

[52] Robert Mifflin and Claudia Sagastizábal. A $\mathcal{VU}$-algorithm for convex minimization. *Mathematical Programming*, 104(2–3):583–608, 2005.

[53] Andreea-Alexandra Muşat and Nicolas Boumal. Gradient descent avoids strict saddles with a simple line-search method too. *arXiv preprint arXiv:2507.13804*, 2025.

[54] Boris T. Polyak. Minimization of unsmooth functionals. *USSR Computational Mathematics and Mathematical Physics*, 9(3):14–29, 1969.

[55] Junwen Qiu, Bohao Ma, Andre Milzarek, and Junyu Zhang. Stochastic saddle avoidance beyond unit excitation and smoothness: A pathwise Lyapunov–Perron framework. *arXiv preprint arXiv:2608.03001*, 2026.

[56] R. Tyrrell Rockafellar and Roger J. B. Wets. *Variational analysis*. Springer, 1998.

[57] Feng Ruan. On the uniform convergence of subdifferentials in stochastic optimization and learning. *Mathematics of Operations Research*, 51(3):2323–2353, 2026.

[58] Alexander Shapiro. On a class of nonsmooth composite functions. *Mathematics of Operations Research*, 28(4):677–692, 2003.

[59] Alexander Shapiro and Huifu Xu. Uniform laws of large numbers for set-valued mappings and subdifferentials of random functions. *Journal of Mathematical Analysis and Applications*, 325(2):1390–1399, 2007.

[60] Yoav Shechtman, Yonina C. Eldar, Oren Cohen, Henry N. Chapman, Jianwei Miao, and Mordechai Segev. Phase retrieval with application to optical imaging: A contemporary overview. *IEEE Signal Processing Magazine*, 32(3):87–109, 2015.

[61] Ju Sun, Qing Qu, and John Wright. A geometric analysis of phase retrieval. *Foundations of Computational Mathematics*, 18(5):1131–1198, 2018.

[62] Terence Tao. *Topics in Random Matrix Theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

[63] Lai Tian and Johannes O. Royset. Failure of uniform laws of large numbers for subdifferentials and beyond. arXiv preprint arXiv:2511.16568, 2025. Revised 16 March 2026.

[64] Lai Tian and Johannes O. Royset. Lipschitzian SLLNs for random functions. arXiv preprint arXiv:2607.20411, 2026.

[65] Kaizheng Wang, Yuling Yan, and Mateo Díaz. Efficient clustering for stretched mixtures: Landscape and optimality. In *Advances in Neural Information Processing Systems*, volume 33, pages 21309–21320, 2020.

[66] Yifan Wang, Jianhao Ma, and Salar Fattahi. On the absence of identifiable manifolds in finite-max composite optimization. *arXiv preprint arXiv:2608.16684*, 2026.

[67] Stephen J. Wright. Identifiable surfaces in constrained optimization. *SIAM Journal on Control and Optimization*, 31(4):1063–1079, 1993.

[68] Zhong Zheng and Lingzhou Xue. Smoothed robust phase retrieval. *arXiv preprint arXiv:2409.01570*, 2024.

[69] Zhihui Zhu, Daniel Soudry, Yonina C. Eldar, and Michael B. Wakin. The global optimization geometry of shallow linear neural networks. *Journal of Mathematical Imaging and Vision*, 62(3):279–292, 2020.

This appendix is rather long, so we start by providing the reader with a roadmap. Appendices A, B, and C present the proofs of our three main results, Proposition 3.1, Theorem 3.2, and Theorem 3.3, respectively. The proof of the last of these is quite involved and requires several auxiliary results. Appendix D develops the probabilistic tools used to count spurious sharp local minima, while Appendix E establishes certain estimates appearing in our characterization of the second moment $\mathbb{E}N_{d,n}^2$. Appendix F collects the remaining auxiliary lemmas, and Appendix G provides the details of the numerical criticality checks in our first experiment.

Appendix contents

A Proof of Proposition 3.1 **17**

B Proof of Theorem 3.2 **18**
B.1 Part 1: Characterization of critical points and sharp local minima 18
B.2 Part 2: Existence of active manifolds 21
B.3 Part 3: Active strict saddle property 22

C Proof of Theorem 3.3 **23**
C.1 Common notation 23
C.2 First moment characterization 25
C.3 Second moment characterization 27
C.4 Constant-probability lower bound 30
C.5 Numerical certification of the probability bound 33

D Probabilistic tools for counting minima **36**
D.1 Proof of Lemma C.1 . . . 36
D.2 Joint density bounds . . . 38
D.3 Kac–Rice formulas . . . 42
D.4 Uniform local limit theorems . . . 44
D.5 Estimates in the shrinking tube . . . 49

E Three ingredients for second moment **51**
E.1 Separated region . . . 51
E.2 Tangential region . . . 53
E.3 Normal-dominated region . . . 63

F Auxiliary lemmas **67**

G Numerical experiment details **71**

A Proof of Proposition 3.1

Proof. The set of critical points of F is characterized in [21, Theorem 5.3]. The function F is continuous and hence closed. Using the identity $|u-v| = \max\{u-v, v-u\}$, we derive

$$F(x) + \|x\|^2 = \mathbb{E}\max\{2(a^\top x)^2 - (a^\top x^\star)^2, (a^\top x^\star)^2\},$$

which is convex since it is the expectation of a maximum of convex quadratics, so F is closed and weakly convex. Next, we prove the statements about smoothness, the Hessian, and sharpness.

We write

$$F(x) = \mathbb{E}\left[\left|(a^\top x)^2 - (a^\top x^\star)^2\right|\right] = \mathbb{E}\left[\left|a^\top(x-x^\star)\right|\left|a^\top(x+x^\star)\right|\right] = \mathbb{E}[|YZ|],$$

where $Y = a^\top(x-x^\star)$ and $Z = a^\top(x+x^\star)$. Both Y and Z are real Gaussian random variables with

$$\mathbb{E}[Y] = \mathbb{E}[Z] = 0, \quad \mathbb{E}[Y^2] = \|x-x^\star\|^2, \quad \mathbb{E}[Z^2] = \|x+x^\star\|^2, \quad \text{and} \quad \mathbb{E}[YZ] = \|x\|^2 - 1.$$

Take $x \notin \operatorname{span}\{x^\star\}$ and write $x = tx^\star + ru$, where u is a unit vector orthogonal to $x^\star$ and $r > 0$. Applying [35, Lemma 1] to the normalized Y and Z, we obtain

$$F(x) = \frac{2}{\pi}\left(2r + (t^2+r^2-1)\arctan\frac{t^2+r^2-1}{2r}\right). \tag{A.1}$$

We denote the function above by $\psi(t,r)$. Since ψ is C^2 when $r > 0$, the chain rule implies that F is C^2 on $\mathbb{R}^d \setminus \operatorname{span}\{x^\star\}$. Moreover, for any $x = \tau u$ on the saddle sphere $\mathcal{S}$, a direct computation shows

$$\frac{\partial\psi}{\partial r}(0,\tau) = \frac{\partial^2\psi}{\partial t\,\partial r}(0,\tau) = 0 \qquad \text{and} \qquad \frac{\partial^2\psi}{\partial r^2}(0,\tau) = -\frac{\partial^2\psi}{\partial t^2}(0,\tau) = \frac{16\tau}{\pi(1+\tau^2)^2}.$$

The chain rule yields

$$\begin{aligned}\nabla^2 F(\tau u) &= \frac{\partial^2\psi}{\partial t^2}(0,\tau)x^\star(x^\star)^\top + \frac{\partial^2\psi}{\partial t\,\partial r}(0,\tau)(x^\star u^\top + u(x^\star)^\top) \\ &\quad + \frac{\partial^2\psi}{\partial r^2}(0,\tau)uu^\top + \frac{1}{\tau}\frac{\partial\psi}{\partial r}(0,\tau)(I_d - x^\star(x^\star)^\top - uu^\top) \\ &= \frac{16\tau}{\pi(1+\tau^2)^2}(uu^\top - x^\star(x^\star)^\top).\end{aligned}$$

Now, let us study smoothness near a point $t_0 x^\star$ of the axis $\mathrm{span}\{x^\star\}$ with $t_0 \neq \pm 1$. For x near $t_0 x^\star$, set $t = (x^\star)^\top x$, $r = \|x - t x^\star\|$, and $q = \|x\|^2 - 1$, which stays away from zero. Using the identity $\arctan s = \mathrm{sign}(s)\,\pi/2 - \arctan(1/s)$ for $s \neq 0$, and a Taylor expansion of $y - \arctan y$ at 0, we derive from (A.1) that

$$F(x) = |q| + \frac{2}{\pi}\left[2r - |q| \arctan\frac{2r}{|q|}\right] = |q| + r^3 h(t, r^2),$$

where h is smooth near $(t_0, 0)$; in particular, the remainder term $r^3 h(t, r^2)$ is $O(r^3)$ as $r \to 0$. Both expressions remain valid on the axis, where $r = 0$ and $F(x) = |q|\,\mathbb{E}(a^\top x^\star)^2 = |q|$. Since t, r^2, and $|q|$ are smooth in x, and $r^3 = \|x - t x^\star\|^3$ is C^2, the function F is C^2 near $t_0 x^\star$, and hence on $\mathbb{R}^d \setminus \{\pm x^\star\}$. At the origin, the same representation implies

$$F(x) = 1 - \|x\|^2 + O(\|x\|^3) \qquad \text{as } x \to 0,$$

so $\nabla^2 F(0) = -2 I_d$.

Finally, $F \geq 0$ and $F(\pm x^\star) = 0$, so $\pm x^\star$ are global minimizers. To establish sharpness, fix $x \neq \pm x^\star$ and let $p = \|x - x^\star\|\,\|x + x^\star\|$. The normalized variables $Y/\|x - x^\star\|$ and $Z/\|x + x^\star\|$ are standard normal with correlation $\rho = (\|x\|^2 - 1)/p \in [-1, 1]$. The Gaussian product formula [35, Lemma 1, equation (45)] therefore implies

$$F(x) = \frac{2p}{\pi}\left(\sqrt{1 - \rho^2} + \rho \arcsin\rho\right) \geq \frac{2p}{\pi}.$$

The inequality holds because the expression in parentheses has minimum 1 at $\rho = 0$: its derivative on $(-1, 1)$ is $\arcsin\rho$, which is negative for $\rho < 0$ and positive for $\rho > 0$. It remains to compare p to the distance to the solution set. The triangle inequality implies $2\max\{\|x - x^\star\|, \|x + x^\star\|\} \geq \|x - x^\star\| + \|x + x^\star\| \geq \|2x^\star\| = 2$. Thus,

$$\begin{aligned} p &= \min\{\|x - x^\star\|, \|x + x^\star\|\}\max\{\|x - x^\star\|, \|x + x^\star\|\} \\ &\geq \min\{\|x - x^\star\|, \|x + x^\star\|\} \\ &= \mathrm{dist}(x, \{\pm x^\star\}). \end{aligned}$$

Combining the two bounds with $\min F = 0$, we obtain $F(x) - \min F \geq \frac{2}{\pi}\mathrm{dist}(x, \{\pm x^\star\})$. The same inequality holds, with equality, at $x = \pm x^\star$. Hence, F is $2/\pi$-sharp with global minimizers $\pm x^\star$. □

B Proof of Theorem 3.2

The proof has three parts. First, we characterize the critical points and sharp local minima of F_n in Lemma B.1. Second, in Proposition B.2, we use this characterization to establish the existence of an active manifold at each critical point other than $\pm x^\star$. Third, we analyze the Hessian along these manifolds to prove the active strict saddle property.

B.1 Part 1: Characterization of critical points and sharp local minima

Lemma B.1 (Characterization of critical points and sharp local minima)**.** *The empirical function F_n is $2\left\|\frac{1}{n}\sum_{i=1}^n a_i a_i^\top\right\|_{\mathrm{op}}$-weakly convex. For every $x \in \mathbb{R}^d$, the subdifferential of F_n at x is*

$$\partial F_n(x) = \frac{2}{n}\left(\sum_{i \notin I(x)} \mathrm{sign}\big((a_i^\top x)^2 - (a_i^\top x^\star)^2\big)(a_i^\top x) a_i + \sum_{i \in I(x)} [-1, 1](a_i^\top x) a_i\right).$$

Suppose $n \geq 2d$. Then, almost surely, every $x \in \mathbb{R}^d \setminus \{\pm x^\star\}$ satisfies $|I(x)| \leq d$. Moreover, such a point x is critical if, and only if, there exist coefficients $\lambda_i \in (-1,1)$, $i \in I(x)$, satisfying

$$0 = \sum_{i \notin I(x)} \operatorname{sign}\big((a_i^\top x)^2 - (a_i^\top x^\star)^2\big)(a_i^\top x)a_i + \sum_{i \in I(x)} \lambda_i (a_i^\top x) a_i.$$

Furthermore, such a point x is a sharp local minimum if, and only if, it is critical and $|I(x)| = d$.

Proof. By rotational invariance of the standard Gaussian measurements, we may assume without loss of generality that $x^\star = e$, where e is the first standard basis vector. For every realization of the measurements, F_n is weakly convex. Indeed, the identity $|s-t| + s = \max\{2s - t, t\}$ implies

$$F_n(x) + \frac{1}{n}\sum_{i=1}^n (a_i^\top x)^2 = \frac{1}{n}\sum_{i=1}^n \max\big\{2(a_i^\top x)^2 - (a_i^\top e)^2,\, (a_i^\top e)^2\big\},$$

and the right-hand side is convex. Since $\frac{1}{n}\sum_{i=1}^n a_i a_i^\top \preceq \big\|\frac{1}{n}\sum_{i=1}^n a_i a_i^\top\big\|_{\mathrm{op}} I_d$, it follows that F_n is $2\big\|\frac{1}{n}\sum_{i=1}^n a_i a_i^\top\big\|_{\mathrm{op}}$-weakly convex. The formula for the subdifferential of F_n follows directly from the standard chain rule [56, Theorem 10.6]. For the remainder of the proof, we work on the probability-one event

$$\mathcal{E}_1 = \{a_i^\top e \neq 0 \text{ for all } i \in [n]\} \cap \{(a_i)_{i \in I} \text{ are linearly independent for all } I \subset [n] \text{ with } |I| \leq d\}. \tag{B.1}$$

Bound on active set size. For any fixed $I \subset [n]$ with $|I| = d$ and $s \in \{-1,1\}^I$, let $x_{I,s}$ denote the solution of the equations $a_i^\top x = s_i a_i^\top e$, $i \in I$. On $\mathcal{E}_1$, this solution exists and is unique, and it depends only on $(a_i)_{i \in I}$. Suppose that some $x \neq \pm e$ has more than d active indices. Choose $I \subset I(x)$ with $|I| = d$ and $j \in I(x) \setminus I$. Since every $i \in I$ is active at x, there are signs s_i with $a_i^\top x = s_i a_i^\top e$, so $x = x_{I,s}$. Hence the event $\mathcal{E}_{I,s,j} = \big\{x_{I,s} \neq \pm e \text{ and } j \in I(x_{I,s})\big\}$ occurs. Since $j \notin I$, the vector a_j is stochastically independent of $(a_i)_{i \in I}$ and hence of $x_{I,s}$. Conditional on $(a_i)_{i \in I}$ and on $x_{I,s} \neq \pm e$, the event $j \in I(x_{I,s})$ requires $a_j^\top (x_{I,s} - e) = 0$ or $a_j^\top (x_{I,s} + e) = 0$. Each is a nonzero linear constraint on a standard Gaussian vector, so $\mathbb{P}(\mathcal{E}_{I,s,j}) = 0$. A union bound over the finitely many triples (I, s, j) shows that, almost surely, $|I(x)| \leq d$ for every $x \in \mathbb{R}^d \setminus \{\pm e\}$.

Characterization of criticality. Recall that x is a critical point if, and only if, there exist coefficients $\lambda_i \in [-1,1]$, $i \in I(x)$, such that

$$0 = \sum_{i \notin I(x)} \operatorname{sign}((a_i^\top x)^2 - (a_i^\top e)^2)(a_i^\top x)a_i + \sum_{i \in I(x)} \lambda_i (a_i^\top x) a_i. \tag{B.2}$$

We will show that, almost surely, every critical point $x \neq \pm e$ satisfies this equation with all active coefficients in $(-1,1)$. We work on the intersection of $\mathcal{E}_1$ and the probability-one event that every $x \neq \pm e$ has at most d active indices. For any such point, the vectors a_i, $i \in I(x)$, are linearly independent, and $a_i^\top x = \pm a_i^\top e \neq 0$ for every $i \in I(x)$. Thus, if x is critical, the coefficients λ_i in (B.2) are unique. If $I(x) = \varnothing$, there is nothing to prove. It therefore remains to rule out the existence of a critical point $x \neq \pm e$ for which $|\lambda_k| = 1$ for some $k \in I(x)$. Fix a nonempty index set $I \subset [n]$ with $|I| \leq d$, active measurement signs $s \in \{-1,1\}^I$, and inactive residual signs $\sigma \in \{-1,1\}^{I^c}$. Define the event

$$\mathcal{E}_{I,s,\sigma} := \left\{ \begin{array}{l} \text{there is a critical point } x \neq \pm e \text{ with } I(x) = I, \quad a_i^\top x = s_i a_i^\top e \text{ for all } i \in I, \\ \operatorname{sign}\big((a_j^\top x)^2 - (a_j^\top e)^2\big) = \sigma_j \text{ for all } j \notin I, \quad \text{and} \quad |\lambda_k| = 1 \text{ for some } k \in I \end{array} \right\}.$$

Since there are only finitely many choices of I, s, and σ, it suffices to show that $\mathbb{P}(\mathcal{E}_{I,s,\sigma}) = 0$.

For each $i \in I$, write $a_i = R_i u_i$, where $R_i = \|a_i\|$ and $u_i = a_i / \|a_i\|$. Since the vectors a_i are independent standard Gaussian vectors, the active radii $(R_i)_{i\in I}$ are independent of the active directions $(u_i)_{i\in I}$ and the inactive vectors $(a_j)_{j\notin I}$. Moreover, each R_i^2 has a continuous distribution. The proof proceeds in two steps. First, we show that any critical point with active set I, active signs s, and inactive signs σ must equal a single candidate $\hat{x}$ that is determined without using the active radii R_i. We then condition on the active directions and inactive vectors and show that $|\lambda_k| = 1$ for some $k \in I$ would force one of the continuously distributed radii to take a specific value.

Step 1: critical points are independent of the active radii. The active equations do not depend on the radii and take the form

$$u_i^\top x = s_i u_i^\top e \qquad \text{for all } i \in I.$$

If $|I| = d$, linear independence of the active directions implies that the active equations already determine a unique point $\hat{x}$, which is therefore independent of the radii. For the rest of this step, suppose that $|I| < d$. Choose an orthonormal basis $v_1, \dots, v_{d-|I|}$ of the orthogonal complement of $\operatorname{span}\{u_i\}_{i\in I}$ and let $V = \begin{bmatrix} v_1, \dots, v_{d-|I|} \end{bmatrix}$. Also choose a solution $\bar{x}$ of the active equations. Both V and $\bar{x}$ can be chosen as measurable functions of the active directions and the fixed signs s. Every solution of the active equations then has the form

$$x = \bar{x} + Vz \qquad \text{with } z \in \mathbb{R}^{d-|I|}.$$

If x is a critical point with active set I and signs (s, σ), multiplying (B.2) by $V^\top$ eliminates the active terms and leads to

$$0 = V^\top \left(\sum_{i\notin I} \sigma_i a_i a_i^\top x + \sum_{i\in I} \lambda_i R_i^2 (u_i^\top x) u_i \right) = V^\top \left(\sum_{i\notin I} \sigma_i a_i a_i^\top \right) x. \tag{B.3}$$

Substituting $x = \bar{x} + Vz$ and setting $H = \sum_{i\notin I} \sigma_i a_i a_i^\top$ yields

$$V^\top H V z = -V^\top H \bar{x}.$$

Lemma F.2 implies that

$$\mathbb{P}(\det(V^\top H V) = 0 \mid (u_i)_{i\in I}) = 0 \quad \text{almost surely.}$$

Thus, almost surely, any critical point with the specified active set and signs must equal

$$\hat{x} = \bar{x} - V(V^\top H V)^{-1} V^\top H \bar{x}.$$

In other words, the set of critical points with active set I and signs (s, σ) is contained in $\{\hat{x}\}$, where $\hat{x}$ is a measurable function of $(u_i)_{i\in I}$ and $(a_j)_{j\notin I}$ alone. The candidate $\hat{x}$ need not itself be critical or realize all the specified signs. Even when it does realize them, whether $\hat{x}$ is critical still depends on the active radii through the constraint $|\lambda_i| \le 1$, and Step 2 exploits this dependence.

Step 2: Use the active radii to exclude boundary coefficients. When $|I| < d$, the projected stationarity equation (B.3) shows that $V^\top H \hat{x} = 0$, so $H\hat{x} \in \operatorname{span}\{u_i : i \in I\}$. When $|I| = d$, the same inclusion holds because the active directions span $\mathbb{R}^d$. Moreover, the active equations imply that $u_i^\top \hat{x} = s_i u_i^\top e \neq 0$ for every $i \in I$. By linear independence, there are therefore unique coefficients c_i, $i \in I$, satisfying

$$\sum_{i\notin I} \sigma_i a_i a_i^\top \hat{x} + \sum_{i\in I} c_i (u_i^\top \hat{x}) u_i = 0.$$

These coefficients are determined entirely by the active directions and inactive vectors. In particular, both $\hat{x}$ and the coefficients c_i are fixed once we condition on $(u_i)_{i\in I}$ and $(a_j)_{j\notin I}$.

Now suppose that a critical point x has active set I and signs s, σ. By Step 1, we have $x = \hat{x}$. Substituting $a_i = R_i u_i$ into the stationarity condition and comparing the result with the preceding

equation, we conclude by uniqueness of the coefficients that $\lambda_i R_i^2 = c_i$ for every $i \in I$. Equivalently, the only possible active coefficients are $\lambda_i = c_i/R_i^2$. Since $R_i > 0$, we have

$$|\lambda_i| = 1 \iff R_i^2 = |c_i|\,.$$

Conditional on the active directions and inactive vectors, the value $|c_i|$ is fixed, whereas R_i^2 remains continuously distributed. For each $i \in I$, the equality on the right therefore has conditional probability zero. Consequently, almost surely,

$$\mathbb{P}(\mathcal{E}_{I,s,\sigma} \mid (u_i)_{i\in I}, (a_i)_{i\notin I}) \le \sum_{i\in I} \mathbb{P}(R_i^2 = |c_i| \mid (u_i)_{i\in I}, (a_i)_{i\notin I}) = 0.$$

Taking expectations yields $\mathbb{P}(\mathcal{E}_{I,s,\sigma}) = 0$. Finally, taking a finite union over all choices of I, s, and σ shows that, almost surely, every critical point $x \neq \pm e$ satisfies (B.2) with coefficients $\lambda_i \in (-1, 1)$ for all $i \in I(x)$.

Characterization of sharp local minima. We start by proving the contrapositive. Suppose $x \neq \pm e$ is a critical point with $|I(x)| < d$. Choose a nonzero vector v orthogonal to every a_i with $i \in I(x)$. Along $x + tv$, the active residuals remain zero, while the inactive residuals retain their signs for sufficiently small t. Thus $t \mapsto F_n(x + tv)$ is locally smooth, and (B.2) implies that its derivative at zero vanishes. Consequently, $F_n(x + tv) - F_n(x) = O(t^2)$, which rules out sharpness.

Conversely, suppose x is critical and $|I(x)| = d$. Set $I = I(x)$, let $\lambda_i \in (-1, 1)$ be its active coefficients, and write $b_i(h) = (a_i^\top x)(a_i^\top h)$. Expanding the residuals at x and using (B.2) to cancel the inactive linear terms implies

$$F_n(x + h) - F_n(x) = \frac{2}{n} \sum_{i\in I} \big(|b_i(h)| - \lambda_i b_i(h)\big) + O(\|h\|^2).$$

Let $\eta = 1 - \max_{i\in I} |\lambda_i| > 0$. Each summand is at least $\eta|b_i(h)|$. On the event $\mathcal{E}_1$, the vectors $(a_i^\top x)a_i$, $i \in I$, form a basis, so there is a constant $c_0 > 0$ such that $\sum_{i\in I} |b_i(h)| \ge c_0 \|h\|$ for every h. Hence

$$F_n(x + h) - F_n(x) \ge \frac{2\eta c_0}{n} \|h\| - C \|h\|^2$$

for some $C \ge 0$ and all sufficiently small h. Shrinking the neighborhood if necessary yields the required positive linear lower bound and proves sharpness, which completes the proof of Lemma B.1. □

B.2 Part 2: Existence of active manifolds

Proposition B.2. *Suppose $n \ge 2d$. Then, almost surely, every critical point $x_0 \in \mathbb{R}^d \setminus \{\pm x^\star\}$ of F_n admits the C^∞ active manifold*

$$\mathcal{M}_{x_0} = \big\{y \in \mathbb{R}^d : a_i^\top (y - x_0) = 0 \text{ for all } i \in I(x_0)\big\}.$$

In particular, if $I(x_0) = \varnothing$, then $\mathcal{M}_{x_0} = \mathbb{R}^d$.

Proof. By rotational invariance, we may assume without loss of generality that $x^\star = e$, where e is the first standard basis vector. By Lemma B.1, we may work on a probability-one event on which the conditions in (B.1) hold and every critical point satisfies the stationarity condition (B.2) with coefficients $\lambda_i \in (-1, 1)$. Fix a critical point $x_0 \in \mathbb{R}^d \setminus \{\pm e\}$. We will show that it admits an active manifold. For brevity, write $I = I(x_0)$ and $\mathcal{M} = \mathcal{M}_{x_0}$.

If $I = \varnothing$, then F_n is C^∞ near x_0, so $\mathbb{R}^d$ is an active manifold at x_0. For the remainder of the proof, suppose $|I| \ge 1$. The restriction of F_n to $\mathcal{M}$ is smooth near x_0. It remains to establish sharpness: we must find a neighborhood U of x_0 and a constant $c > 0$ such that

$$\operatorname{dist}(0, \partial F_n(x)) \ge c \quad \text{for all } x \in U \setminus \mathcal{M}.$$

We first introduce two positive constants. Set

$$\eta = \min_{i \in I} \{1 - |\lambda_i|\} > 0.$$

Since the vectors a_i, $i \in I$, are linearly independent and $\left|a_i^\top x_0\right| = \left|a_i^\top e\right| > 0$ for every $i \in I$, the vectors $(a_i^\top x_0)a_i$, $i \in I$, are also linearly independent. Define

$$c_0 = \min_{i \in I} \operatorname{dist}\big((a_i^\top x_0)a_i,\ \operatorname{span}\{(a_j^\top x_0)a_j \colon j \in I \setminus \{i\}\}\big) > 0.$$

We claim that there exists a radius $\delta > 0$ such that $\operatorname{dist}(0, \partial F_n(x))$ is bounded below by a positive constant uniformly over $x \in B_\delta(x_0) \setminus \mathcal{M}$. By continuity, we can choose δ sufficiently small that, for every $i \in I^c$ and $x \in B_\delta(x_0)$, the sign of $(a_i^\top x)^2 - (a_i^\top e)^2$ agrees with its sign at x_0. We may also shrink this neighborhood so that, for every $i \in I$, $a_i^\top x$ and $a_i^\top x_0$ have the same sign. Now fix $x \in B_\delta(x_0) \setminus \mathcal{M}$. Since $x \notin \mathcal{M}$, there is an index $i_x \in I$ such that $a_{i_x}^\top x \neq a_{i_x}^\top x_0$. We consider the case

$$(a_{i_x}^\top x)^2 > (a_{i_x}^\top x_0)^2 = (a_{i_x}^\top e)^2.$$

The case with the reverse inequality is analogous. In the displayed case, every element of $\partial F_n(x)$ can be written as

$$\frac{2}{n}\left(\sum_{i \in I \setminus \{i_x\}} \lambda_i'(a_i^\top x)a_i + (a_{i_x}^\top x)a_{i_x} + \sum_{i \in I^c} \operatorname{sign}((a_i^\top x)^2 - (a_i^\top e)^2)(a_i^\top x)a_i\right),$$

where $\lambda_i' \in [-1, 1]$. Omitting the factor $\frac{2}{n}$ for the moment, we decompose this expression as follows:

$$\begin{aligned}
&\sum_{i \in I \setminus \{i_x\}} \lambda_i'(a_i^\top x)a_i + (a_{i_x}^\top x)a_{i_x} + \sum_{i \in I^c} \operatorname{sign}((a_i^\top x)^2 - (a_i^\top e)^2)(a_i^\top x)a_i \\
&\quad = \underbrace{\sum_{i \in I \setminus \{i_x\}} \lambda_i'(a_i^\top x_0)a_i + (a_{i_x}^\top x_0)a_{i_x} + \sum_{i \in I^c} \operatorname{sign}((a_i^\top x_0)^2 - (a_i^\top e)^2)(a_i^\top x_0)a_i}_{I_1} \\
&\qquad + \underbrace{\sum_{i \in I \setminus \{i_x\}} \lambda_i'(a_i^\top (x - x_0))a_i + \langle a_{i_x}, x - x_0 \rangle\, a_{i_x}}_{I_2} + \underbrace{\sum_{i \in I^c} \operatorname{sign}((a_i^\top x)^2 - (a_i^\top e)^2)(a_i^\top (x - x_0))a_i}_{I_3}.
\end{aligned}$$

By continuity, the norms of I_2 and I_3 can be made arbitrarily small by choosing δ sufficiently small. For I_1, the stationarity condition at x_0 implies that

$$\begin{aligned}
I_1 &= \sum_{i \in I \setminus \{i_x\}} \lambda_i'(a_i^\top x_0)a_i + (a_{i_x}^\top x_0)a_{i_x} + \sum_{i \in I^c} \operatorname{sign}((a_i^\top x_0)^2 - (a_i^\top e)^2)(a_i^\top x_0)a_i \\
&= \sum_{i \in I \setminus \{i_x\}} (\lambda_i' - \lambda_i)(a_i^\top x_0)a_i + (1 - \lambda_{i_x})(a_{i_x}^\top x_0)a_{i_x}.
\end{aligned}$$

Since the first sum lies in $\operatorname{span}\{(a_j^\top x_0)a_j \colon j \in I \setminus \{i_x\}\}$, the definitions of η and c_0 give

$$\|I_1\| \geq (1 - \lambda_{i_x})\operatorname{dist}\big((a_{i_x}^\top x_0)a_{i_x},\ \operatorname{span}\{(a_j^\top x_0)a_j \colon j \in I \setminus \{i_x\}\}\big) \geq \eta c_0 > 0.$$

Consequently, by choosing δ sufficiently small that $\|I_2\| + \|I_3\| \leq \frac{\eta c_0}{2}$ and restoring the factor $2/n$, we establish the desired sharpness with $c = \eta c_0 / n$. □

B.3 Part 3: Active strict saddle property

Proof of Theorem 3.2. By rotational invariance, we may assume without loss of generality that $x^\star = e$, where e is the first standard basis vector. We work on a probability-one event on which

the conclusions of both Lemma B.1 and Proposition B.2 hold. Fix a critical point $x_0 \notin \{\pm e\}$. By Proposition B.2, an active manifold at x_0 is

$$\mathcal{M}_{x_0} = \{y : a_i^\top (y - x_0) = 0, i \in I(x_0)\}.$$

We consider two cases. If $|I(x_0)| = d$, then the last part of Lemma B.1 shows that x_0 is a sharp, and hence strict, local minimizer of F_n. Now, suppose that $|I(x_0)| < d$. For brevity, write $I = I(x_0)$ and $\mathcal{M} = \mathcal{M}_{x_0}$. Points on the manifold can be written as

$$x = x_0 + Vz \qquad \text{with } z \in \mathbb{R}^{d-|I|},$$

where the columns of $V = \begin{bmatrix} v_1, \dots, v_{d-|I|} \end{bmatrix}$ form an orthonormal basis of the tangent space of $\mathcal{M}$. Near x_0, the restriction of F_n to $\mathcal{M}$ is quadratic. Define $H = \frac{2}{n} \sum_{i \notin I} \operatorname{sign}((a_i^\top x_0)^2 - (a_i^\top e)^2) a_i a_i^\top$. In these coordinates, the Riemannian Hessian of the restriction at x_0 is $V^\top H V$. By Lemma F.2, this Riemannian Hessian is almost surely nonsingular. If it is positive definite, then x_0 is a strict local minimizer of F_n on $\mathcal{M}_{x_0}$, and hence also on $\mathbb{R}^d$ by Lemma F.3. Otherwise, nonsingularity implies that it has a negative eigenvalue, so x_0 is an active strict saddle.

Finally, $\pm e$ are strict local minimizers. Indeed, near e, every $a_i^\top x$ has the same sign as $a_i^\top e$. Thus, $F_n(x) = 0$ implies $a_i^\top x = a_i^\top e$ for every i, and hence $x = e$ because the measurement vectors span $\mathbb{R}^d$. The same argument applies at $-e$. This proves that F_n has the active strict saddle property. □

C Proof of Theorem 3.3

We first prove a first-moment result, Theorem C.2 (Appendix C.2), which directly implies part (1) of Theorem 3.3. Then, together with the second-moment result in Theorem C.3 (Appendix C.3), we prove part (2) of Theorem 3.3 (Appendix C.4). We put all supporting tools in Appendices D and E, together with explanations of their roles in the proof. We first fix some common notation in Appendix C.1, which is used consistently in Appendices C, D, and E.

C.1 Common notation

Throughout Appendices C, D, and E, fix an integer $d \geq 4$ and a deterministic unit vector $x^\star \in \mathbb{R}^d$. We start by defining the notation we will use in the proof. Most of the intuition for what we are about to introduce already appeared in Section 4. We adopt the following convention throughout. All our asymptotic symbols of the form $O(\cdot)$ and $o(\cdot)$ refer to limits $n \to \infty$. Sometimes we will add subscripts on $O(\cdot)$ or $o(\cdot)$ to highlight the parameters that are held fixed and on which the implicit constants may depend.

We use the following constant

$$\kappa = 8\tau/[\pi(1+\tau^2)^2] > 0, \tag{C.1}$$

where again τ is the unique solution of the equation $\arctan \tau + \tau/(1+\tau^2) = \pi/4$. Throughout, we fix the following active sign pattern,

$$s = (s_1, \dots, s_d) = (\underbrace{1, \dots, 1}_{d-2}, -1, -1). \tag{C.2}$$

For any realization of $(a_1, \dots, a_n)$, let A denote the matrix whose jth-row corresponds to the vector a_j. For any index set $J = \{j_1 < \cdots < j_r\} \subset [n]$, define the submatrix indexed by J as

$$A_J = \begin{pmatrix} a_{j_1} & \cdots & a_{j_r} \end{pmatrix}^\top.$$

For any $x, y \in \mathbb{R}^d$ and a random vector $a \sim N(0, I_d)$, define the objects

$$Z_x(a) := \operatorname{sign}((a^\top x)^2 - (a^\top x^\star)^2)(a^\top x)a, \tag{C.3}$$

$$\mu(x) := \mathbb{E}[Z_x(a)], \quad \Sigma_x := \operatorname{Cov}(Z_x(a)), \quad \text{and} \quad \Sigma_{x,y} := \operatorname{Cov}\begin{pmatrix} Z_x(a) \\ Z_y(a) \end{pmatrix}. \tag{C.4}$$

For any pair of matrices $B = \begin{pmatrix} b_1 & \cdots & b_d \end{pmatrix}^\top \in \mathbb{R}^{d\times d}$ and $C = \begin{pmatrix} c_1 & \cdots & c_r \end{pmatrix}^\top \in \mathbb{R}^{r\times d}$, define

$$Q(x; B) := \begin{pmatrix} (b_1^\top x)b_1 & \cdots & (b_d^\top x)b_d \end{pmatrix}, \qquad \text{and} \qquad S(x; C) := \sum_{i=1}^{r} Z_x(c_i).$$

For any index set $I = \{i_1 < \cdots < i_d\} \subset [n]$, define

$$X_I(x) := \begin{pmatrix} a_{i_1}^\top (x - s_1 x^\star) \\ \vdots \\ a_{i_d}^\top (x - s_d x^\star) \end{pmatrix} = A_I x - \operatorname{diag}(s) A_I x^\star. \tag{C.5}$$

Almost surely, A_I is invertible, and the solution of $X_I(x) = 0$ is denoted by

$$x_I = A_I^{-1} \operatorname{diag}(s) A_I x^\star. \tag{C.6}$$

By Lemma 4.1, almost surely, x is a sharp spurious local minimum with active index set I and active sign pattern s if, and only if, the following holds:

$$0 \in S(x; A_{I^c}) + Q(x; A_I)(-1, 1)^d, \qquad \text{and} \qquad X_I(x) = 0.$$

Recall the saddle radius τ and the population saddle sphere $\mathcal{S}$ from Proposition 3.1, and the tube T_n from (4.5). We write $\mathbb{S}^{d-2} = \{u\colon \|u\| = 1,\ u \perp x^\star\}$, so that $\mathcal{S} = \tau\,\mathbb{S}^{d-2}$. When $n^{-1/2} < \tau$, every $x \in T_n$ can be written uniquely as $x = \alpha_x x^\star + (\tau + \beta_x)u_x$ with $u_x \in \mathbb{S}^{d-2}$ and $|\alpha_x|, |\beta_x| \le n^{-1/2}$; we call (α_x, β_x, u_x) the tube coordinates of x. We will restrict to well-conditioned active matrices, using the following lemma. We defer its proof to Appendix D.1.

Lemma C.1 (Choice of regularity constants)**.** *Fix a dimension $d \ge 4$. Then, there exist finite deterministic constants $M_d, \lambda_d, \vartheta_d > 0$, depending only on d, with the following properties. For $x \in \mathbb{R}^d$, define the set*

$$\mathcal{K}_x = \left\{ B \in \mathbb{R}^{d\times d} : \max_{1\le j\le d} \|b_j\| \le M_d,\ |\det B| \ge \lambda_d,\ \textit{and}\ |\det Q(x; B)| \ge \lambda_d \right\},$$

and, for $n \ge d$ and $I \subset [n]$ with $|I| = d$, the conditional expectation

$$a(x) = \mathbb{E}\big[|\det A_I|\,|\det Q(x; A_I)|\,1_{\mathcal{K}_x}(A_I) \,\big|\, X_I(x) = 0\big].$$

The function $a(\cdot)$ is independent of the choice of I, is continuous near $\mathcal{S}$, and satisfies

$$a(\tau u) = \vartheta_d \qquad \textit{for every } u \in \mathbb{S}^{d-2}.$$

Moreover, for any fixed d,

$$\sup_{x\in T_n} |a(x) - \vartheta_d| \to 0 \qquad \textit{as } n \to \infty.$$

We fix the constants M_d, λ_d, and ϑ_d from Lemma C.1 for the rest of the proof. We count only those sharp spurious local minimizers that also lie in the tube T_n around the population saddle sphere and whose active matrix satisfies $A_I \in \mathcal{K}_{x_I}$, with $\mathcal{K}_x$ as in Lemma C.1. Formally, for a set $I \subset [n]$ with $|I| = d$, define the random indicator function

$$\begin{aligned} \chi_I^n(a_1, \ldots, a_n) &= 1_{\{x_I \in T_n\}} 1_{\mathcal{K}_{x_I}}(A_I) 1_{\{-S(x_I; A_{I^c}) \in Q(x_I; A_I)(-1,1)^d\}}, \quad \text{and} \\ N_{d,n}(a_1, \ldots, a_n) &= \sum_{|I|=d} \chi_I^n. \end{aligned} \tag{C.7}$$

Then $\chi_I^n = 1$ precisely when the candidate solution x_I is a sharp spurious local minimum lying in T_n and satisfying $A_I \in \mathcal{K}_{x_I}$. Thus $N_{d,n}$ counts the total number of such sharp spurious local minima for a given realization.

C.2 First moment characterization

For the next result, recall that M_d, λ_d, and ϑ_d are the constants provided by Lemma C.1, and Σ_x and $N_{d,n}$ were defined in (C.4) and (C.7), respectively.

Theorem C.2. *For any $d \geq 4$, we have that*

$$n^{-d/2+1}\mathbb{E}[N_{d,n}] \to \gamma_d \mathcal{I}_d \quad \text{as } n \to \infty$$

with

$$\gamma_d = \frac{2^d \tau^{d-2} \vartheta_d}{d![2\pi(1+\tau^2)]^{d/2}}, \qquad \text{and} \qquad \mathcal{I}_d = \int_{\mathbb{S}^{d-2}} \int_{[-1,1]^2} \varphi_{\Sigma_{\tau u}}(-\kappa z_1 x^\star + \kappa z_2 u)\, \mathrm{d}z \mathrm{d}\sigma(u),$$

where $\mathrm{d}\sigma$ is the surface measure on $\mathbb{S}^{d-2}$, and φ_Σ is the density of $N(0,\Sigma)$.

Proof. By rotational invariance, we may assume without loss of generality that $x^\star = e$, where e is the first standard basis vector. Recall that $N_{d,n} = \sum_{|I|=d} \chi_I^n$. Since the Gaussian rows are i.i.d. and every ordered active set uses the same sign pattern, all the indicators χ_I^n have the same expectation. Hence, for any fixed $I \subset [n]$ with $|I| = d$,

$$\mathbb{E}[N_{d,n}] = \sum_{|J|=d} \mathbb{E}[\chi_J^n] = \binom{n}{d} \mathbb{E}[\chi_I^n].$$

Since $\binom{n}{d} = n^d/d!\,(1 + O_d(n^{-1}))$, it suffices to prove

$$\mathbb{E}[\chi_I^n] = d!\,\gamma_d \mathcal{I}_d\, n^{-d/2-1} + o_d(n^{-d/2-1}). \tag{C.8}$$

In what follows we decompose the expectation of χ_I^n into a leading term and a negligible remainder, and then evaluate the leading term in tube coordinates. To do so, we leverage the Kac-Rice formula (Lemma D.4(I)), which gives

$$\begin{aligned}
\mathbb{E}[\chi_I^n] &= \int_{T_n} p_{X_I(x)}(0) \mathbb{E}\big[|\det A_I|\, 1_{\mathcal{K}_x}(A_I) 1_{-S(x;A_{I^c}) \in Q(x;A_I)(-1,1)^d} \;\big|\; X_I(x) = 0\big] \mathrm{d}x \\
&= \int_{T_n} p_{X_I(x)}(0) \mathbb{E}\big[|\det A_I|\, I_1 \;\big|\; X_I(x) = 0\big] \mathrm{d}x,
\end{aligned}$$

where $p_{X_I(x)}$ is the density of the $\mathbb{R}^d$ Gaussian vector $X_I(x)$, and

$$I_1 = \mathbb{E}\left[1_{\mathcal{K}_x}(A_I) 1_{-S(x;A_{I^c}) \in Q(x;A_I)(-1,1)^d} \middle| A_I, X_I(x) = 0\right].$$

We use a local central limit theorem, Lemma D.7(I), to approximate

$$I_1 = \underbrace{\frac{2^d}{(n-d)^{d/2}} |\det Q(x;A_I)|\, 1_{\mathcal{K}_x}(A_I) \varphi_{\Sigma_x}(-\sqrt{n-d}\mu(x))}_{I_2} + \underbrace{\frac{1}{(n-d)^{d/2}} 1_{\mathcal{K}_x}(A_I) r_n(x;A_I)}_{I_3},$$

where φ_Σ is the density of $N(0,\Sigma)$. To justify this approximation, note that $X_I(x)$ depends only on A_I. Since A_{I^c} is independent of A_I, its rows remain independent standard Gaussian vectors under the conditioning in I_1. The remainder I_3 is negligible, in particular, Lemma D.7(I) ensures that the function r_n satisfies the uniform bound

$$\sup_{x \in T_n} \sup_{A \in \mathbb{R}^{d \times d}} 1_{\mathcal{K}_x}(A)\, |r_n(x;A)| \leq \eta_n \tag{C.9}$$

for some deterministic sequence $\eta_n \to 0$, so that $|I_3| \le (n-d)^{-d/2}\eta_n$ for all $x \in T_n$.

Substituting $I_1 = I_2 + I_3$ into the integral above, we split the contribution of this active set into the leading term I_4 and the remainder I_5:

$$\mathbb{E}[\chi_I^n] = \underbrace{\int_{T_n} p_{X_I(x)}(0)\mathbb{E}\big[|\det A_I|\, I_2 \;\big|\; X_I(x)=0\big]\mathrm{d}x}_{I_4} + \underbrace{\int_{T_n} p_{X_I(x)}(0)\mathbb{E}\big[|\det A_I|\, I_3 \;\big|\; X_I(x)=0\big]\mathrm{d}x}_{I_5}.$$

To control the remainder term I_5, we use (C.9) together with the following bounds, valid for $x \in T_n$ and $A_I \in \mathcal{K}_x$:

$$|\det A_I| \le \prod_{i\in I}\|a_i\| \le M_d^d, \qquad |\det Q(x;A_I)| \le \prod_{i\in I}\Big\|(a_i^\top x)a_i\Big\| = O_d(1),$$

$$\mathrm{Vol}(T_n) = O_d(n^{-1}), \qquad \text{and} \qquad p_{X_I(x)}(0) = (2\pi)^{-d/2}\prod_{i=1}^d \frac{1}{\|x - s_i e\|} = O_d(1).$$

The first three bounds follow from our construction. To derive the last relation, i.e., the density formula, write $I = \{i_1 < \cdots < i_d\}$. For fixed x, the coordinates $a_{i_j}^\top(x - s_j e)$ of $X_I(x)$ are independent $N(0, \|x - s_j e\|^2)$ variables, so their densities at zero, $(2\pi)^{-1/2}\|x - s_j e\|^{-1}$, multiply to produce the displayed expression. Moreover, on T_n, we have $\|x - s_j e\| \ge \tau - n^{-1/2} \ge \tau/2$ for sufficiently large n, which implies the uniform $O_d(1)$ bound. Since $(n-d)/n \to 1$ and $\eta_n \to 0$, these bounds imply

$$I_5 = o_d(n^{-d/2-1}).$$

It remains to evaluate the leading term

$$I_4 = 2^d(n-d)^{-d/2}\int_{T_n} p_{X_I(x)}(0)a(x)\varphi_{\Sigma_x}(-\sqrt{n-d}\mu(x))\mathrm{d}x.$$

By Lemma C.1, $a(\tau u) = \vartheta_d > 0$ for every $u \in \mathbb{S}^{d-2}$. In what follows, we do a change of variables to integrate over the tube. For sufficiently large n such that $n^{-1/2} < \tau$, parametrize T_n by

$$x = x_{n,z,u} = \frac{z_1}{\sqrt{n}}e + \left(\tau + \frac{z_2}{\sqrt{n}}\right)u \qquad \text{with } z \in [-1,1]^2, \quad \text{and} \quad u \in \mathbb{S}^{d-2}.$$

Hence $\mathrm{d}x = \frac{1}{n}\left(\tau + \frac{z_2}{\sqrt{n}}\right)^{d-2}\mathrm{d}z\,\mathrm{d}\sigma(u)$, where $\mathrm{d}\sigma$ is surface measure on $\mathbb{S}^{d-2}$. Therefore,

$$I_4 = \int_{\mathbb{S}^{d-2}}\int_{[-1,1]^2} \frac{2^d\big(\tau + z_2/\sqrt{n}\big)^{d-2}}{n(n-d)^{d/2}}\, p_{X_I(x_{n,z,u})}(0)\, a(x_{n,z,u})\, \varphi_{\Sigma_{x_{n,z,u}}}\big(-\sqrt{n-d}\,\mu(x_{n,z,u})\big)\,\mathrm{d}z\,\mathrm{d}\sigma(u).$$

We now let $n \to \infty$. Since $e \perp u$ and $\|e\| = \|u\| = 1$, a direct calculation shows that

$$\|x_{n,z,u} - s_i e\|^2 = 1 + \tau^2 + \frac{2(\tau z_2 - s_i z_1)}{\sqrt{n}} + \frac{z_1^2 + z_2^2}{n} \qquad \text{for all } i \in \{1,\dots,d\} \to 1 + \tau^2$$

uniformly for $z \in [-1,1]^2$ and $u \in \mathbb{S}^{d-2}$. Substituting this into the density yields the uniform limit

$$p_{X_I(x_{n,z,u})}(0) = (2\pi)^{-d/2}\prod_{i=1}^d \frac{1}{\|x_{n,z,u} - s_i e\|} \to [2\pi(1+\tau^2)]^{-d/2}.$$

Lemma C.1 and Lemma D.8 provide the remaining limits, namely

$$a(x_{n,z,u}) \to \vartheta_d, \quad \sqrt{n-d}\mu(x_{n,z,u}) \to -\kappa z_1 e + \kappa z_2 u, \quad \text{and} \quad \big\|\Sigma_{x_{n,z,u}} - \Sigma_{\tau u}\big\|_{\mathrm{op}} \to 0,$$

uniformly for $z \in [-1,1]^2$ and $u \in \mathbb{S}^{d-2}$. Since $\Sigma_{\tau u} = \tau^2(I_d + 2uu^\top) \succeq \tau^2 I_d$, the Gaussian densities also converge uniformly. The radial factor converges uniformly to τ^{d-2}. Passing to the limit on this

fixed integration domain, using $(n-d)/n \to 1$ and the symmetry $\varphi_\Sigma(-v) = \varphi_\Sigma(v)$, we obtain

$$I_4 = \frac{2^d \tau^{d-2} \vartheta_d}{[2\pi(1+\tau^2)]^{d/2}} n^{-d/2-1} \underbrace{\int_{\mathbb{S}^{d-2}} \int_{[-1,1]^2} \varphi_{\Sigma_{\tau u}}(-\kappa z_1 e + \kappa z_2 u) \mathrm{d}z \, \mathrm{d}\sigma(u)}_{\mathcal{I}_d} + o_d(n^{-d/2-1}).$$

Combining this asymptotic formula for I_4 with the bound for I_5 proves (C.8), completing the proof. □

C.3 Second moment characterization

For the following result and its proof, recall that M_d, λ_d, and ϑ_d are the constants provided by Lemma C.1, and the objects $\kappa, \Sigma_{x,y}$ and $N_{d,n}$ were defined in (C.1), (C.4) and (C.7), respectively.

Theorem C.3. *For any $d \geq 4$, we have that*

$$n^{-d+2} \mathbb{E}[N_{d,n}^2] \to \gamma_d^2 \mathcal{J}_d,$$

where

$$\gamma_d = \frac{2^d \tau^{d-2} \vartheta_d}{d! [2\pi(1+\tau^2)]^{d/2}}, \quad \text{and}$$

$$\mathcal{J}_d = \int_{\substack{u,v \in \mathbb{S}^{d-2} \\ u \neq \pm v}} \int_{[-1,1]^4} \varphi_{\Sigma_{\tau u, \tau v}} \left(\begin{pmatrix} -\kappa z_1 x^\star + \kappa z_2 u \\ -\kappa w_1 x^\star + \kappa w_2 v \end{pmatrix} \right) \mathrm{d}z \, \mathrm{d}w \, \mathrm{d}\sigma(u) \, \mathrm{d}\sigma(v).$$

Proof. By rotational invariance of the standard Gaussian measurements, we may assume without loss of generality that $x^\star = e$. Recall that $N_{d,n} = \sum_{|I|=d} \chi_I^n$ and that $(\chi_I^n)^2 = \chi_I^n$. Expanding the square and grouping the off-diagonal terms by the overlap $r = |I \cap J|$, we obtain

$$\mathbb{E}[N_{d,n}^2] = \mathbb{E}[N_{d,n}] + \sum_{r=0}^{d-1} D_{r,n} \qquad \text{where} \qquad D_{r,n} = \sum_{\substack{I,J \subset [n] \\ |I|=|J|=d, \ |I \cap J|=r}} \mathbb{E}[\chi_I^n \chi_J^n].$$

The sum defining $D_{r,n}$ is over ordered pairs, so no factor of two is needed. By Theorem C.2, we have $\mathbb{E}[N_{d,n}] = O_d(n^{d/2-1}) = o_d(n^{d-2})$. It therefore suffices to establish $\mathcal{J}_d < \infty$ and the estimates

$$D_{0,n} = \gamma_d^2 \mathcal{J}_d \, n^{d-2} + o_d(n^{d-2}), \qquad \text{and} \qquad D_{r,n} = o_d(n^{d-2}) \qquad \text{for } 1 \leq r \leq d-1. \tag{C.10}$$

The rest of the proof proceeds in three steps. We first express each $\mathbb{E}[\chi_I^n \chi_J^n]$ as a conditional Kac–Rice integral. We then split the integration domain into three regions and use the estimates from Appendix E to control the integral in each region. Finally, we combine these estimates.

Step 1: Express each pair contribution as a conditional integral. Fix distinct index sets $I, J \subset [n]$ with $|I| = |J| = d$ and $|I \cap J| = r$. Let $U = (I \cup J)^c$ and $\nu = |U| = n - 2d + r$. The sums over the two sets of inactive measurements decompose as

$$S(x; A_{I^c}) = S(x; A_{J \setminus I}) + S(x; A_U), \qquad \text{and} \qquad S(y; A_{J^c}) = S(y; A_{I \setminus J}) + S(y; A_U).$$

Conditioning on $A_{I \cup J}$ fixes the first term in each decomposition. The rows of A_U remain independent standard Gaussian vectors, since they are independent of $A_{I \cup J}$ and the constraints $X_I(x) = X_J(y) = 0$ involve only rows in $I \cup J$. The sums $S(x; A_U)$ and $S(y; A_U)$ depend on the same Gaussian rows and are generally dependent, so the two inclusion events must be treated jointly.

For a deterministic matrix $B \in \mathbb{R}^{(2d-r) \times d}$ whose rows are indexed by $I \cup J$, let $C = (g_1, \ldots, g_\nu)^\top$,

where the g_j are independent $N(0, I_d)$ vectors, and define

$$\pi_{n,I,J}(x, y; B) = \mathbb{P}_{g_1,\dots,g_\nu}\left\{\begin{matrix} -S(x; C) \in Q(x; B_I)(-1, 1)^d + S(x; B_{J\setminus I}), \\ -S(y; C) \in Q(y; B_J)(-1, 1)^d + S(y; B_{I\setminus J}) \end{matrix}\right\}.$$

This is the probability appearing in Lemma D.7(II). Thus, Lemma D.4(II), followed by conditioning on $A_{I\cup J}$, implies

$$\begin{aligned}&\mathbb{E}[\chi_I^n \chi_J^n] \\ &= \int\limits_{T_n^2 \cap \mathcal{R}_{I,J}} p_{X_I(x),X_J(y)}(0,0)\, \mathbb{E}\big[|\det A_I|\,|\det A_J|\, 1_{\mathcal{K}_x}(A_I) 1_{\mathcal{K}_y}(A_J) \pi_{n,I,J}(x, y; A_{I\cup J}) \;\big|\; X_I(x) = X_J(y) = 0\big]\, \mathrm{d}x\, \mathrm{d}y,\end{aligned} \tag{C.11}$$

where $\mathcal{R}_{I,J} = \left\{(x, y)\colon \operatorname{Cov}\begin{pmatrix} X_I(x) \\ X_J(y) \end{pmatrix} \succ 0\right\}$, and $p_{X_I(x),X_J(y)}$ is the density of the Gaussian vector $\begin{pmatrix} X_I(x) \\ X_J(y) \end{pmatrix} \in \mathbb{R}^{2d}$ on this region.

Step 2: Split the integral into three regions. For all sufficiently large n such that $n^{-1/2} \le \tau/2$, every $x \in T_n$ has a unique representation

$$x = \alpha_x e + (\tau + \beta_x) u_x \qquad \text{with } |\alpha_x|, |\beta_x| \le n^{-1/2}, \qquad \text{and} \qquad u_x \in \mathbb{S}^{d-2}.$$

For a pair (x, y), define

$$\delta(x, y) = \min\{\|u_x - u_y\|, \|u_x + u_y\|\}, \text{ and } \rho(x, y) = \sqrt{\big(\alpha_x - \operatorname{sign}(u_x^\top u_y)\alpha_y\big)^2 + (\beta_x - \beta_y)^2}.$$

The quantity δ measures angular separation up to sign. When $\delta < 1$, the quantity ρ measures how much the two points differ in their normal coordinates, after replacing y by $-y$ if u_y is closer to $-u_x$ than to u_x.

Fix $0 < \delta_0 < 1$. Pairs with $\delta \ge \delta_0$ form the separated region, where the local limit estimate in Lemma D.7(II) applies. The limiting covariance $\Sigma_{\tau u_x, \tau u_y}$ becomes singular at $u_x = \pm u_y$, so pairs with $\delta < \delta_0$ require separate estimates. We divide these nearby pairs by comparing their angular separation δ with their normal separation ρ. Formally, we define three regions

$$\begin{aligned}\Omega^{\mathrm{sep}} &= \{(x, y) \in T_n^2 \cap \mathcal{R}_{I,J} : \delta(x, y) \ge \delta_0\}, \\ \Omega^{\mathrm{tan}} &= \{(x, y) \in T_n^2 \cap \mathcal{R}_{I,J} : \delta(x, y) < \delta_0,\ \rho(x, y) \le \delta(x, y)\}, \quad \text{and} \\ \Omega^{\mathrm{nor}} &= \{(x, y) \in T_n^2 \cap \mathcal{R}_{I,J} : \delta(x, y) < \delta_0,\ \rho(x, y) > \delta(x, y)\}.\end{aligned}$$

We call them the separated, tangential, and normal-dominated regions, respectively. As u varies, the base point τu moves along the saddle sphere $\mathcal{S}$, while α and β describe displacements normal to it. In the tangential region, angular separation is at least as large as separation in the normal coordinates; in the normal-dominated region, the latter is larger.

For a fixed pair I, J with overlap r, let $I_{r,n}^{\mathrm{sep}}$ denote the integral in (C.11) restricted to Ω^{sep}. Define $D_{r,n}^{\mathrm{sep}}$ by summing $I_{r,n}^{\mathrm{sep}}$ over all ordered pairs $I, J \subset [n]$ with $|I| = |J| = d$ and $|I \cap J| = r$. Define $D_{r,n}^{\mathrm{tan}}$ and $D_{r,n}^{\mathrm{nor}}$ in the same way, using the restrictions of (C.11) to Ω^{tan} and Ω^{nor}, respectively. We suppress the dependence of the regions and of $I_{r,n}^{\mathrm{sep}}$ on I, J, as well as the dependence of the regions, integrals, and sums on δ_0. Since the three regions partition the integration domain, we have

$$D_{r,n} = D_{r,n}^{\mathrm{sep}} + D_{r,n}^{\mathrm{tan}} + D_{r,n}^{\mathrm{nor}}.$$

Let $\mathcal{J}_d(\delta_0)$ denote the integral defining $\mathcal{J}_d$ in the theorem, with the additional restriction $\min\{\|u - v\|, \|u + v\|\} \ge \delta_0$. For each fixed δ_0, Lemma D.8 implies $\lambda_{\min}(\Sigma_{\tau u, \tau v}) \ge c_{d,\delta_0} > 0$ uniformly over the restricted angular domain. Since $\Sigma_{\tau u, \tau v}$ is a $2d \times 2d$ covariance matrix, the

Gaussian density satisfies

$$\varphi_{\Sigma_{\tau u,\tau v}}(\xi) \le \frac{1}{(2\pi)^d\sqrt{\det \Sigma_{\tau u,\tau v}}} \le \frac{1}{(2\pi c_{d,\delta_0})^d} \qquad \text{for every } \xi \in \mathbb{R}^{2d}.$$

Thus the integrand defining $\mathcal{J}_d(\delta_0)$ is uniformly bounded. Since $[-1,1]^4$ has finite volume and $\mathbb{S}^{d-2}$ has finite surface measure, $\mathcal{J}_d(\delta_0) < \infty$.

Table 1 summarizes the estimates, proved in Appendix E, for the three regional terms. All asymptotics in the table are as $n \to \infty$ with δ_0 fixed. In particular, the separated region supplies the leading term for disjoint pairs, while the tangential error can be made small by subsequently decreasing δ_0.

Term	Asymptotic estimate		Reference
$D_{r,n}^{\mathrm{sep}}$	$r = 0:$	$\gamma_d^2 \mathcal{J}_d(\delta_0) n^{d-2} + o_{d,\delta_0}(n^{d-2})$,	Lemma E.1
	$1 \le r \le d-1:$	$o_{d,\delta_0}(n^{d-2})$	
$D_{r,n}^{\mathrm{tan}}$	$r = 0:$	$\le C_d n^{d-2}\delta_0^{d/2-1} + C_d n^{d/2-1}$,	Lemma E.2
	$1 \le r \le d-1:$	$o_{d,\delta_0}(n^{d-2})$	
$D_{r,n}^{\mathrm{nor}}$	$0 \le r \le d-1:$	$o_{d,\delta_0}(n^{d-2})$	Lemma E.7

Table 1: Estimates for the three regional terms at fixed δ_0. The constant $C_d > 0$ depends only on d, and γ_d is as in Theorem C.2.

Step 3: Combine the estimates and remove the angular cutoff. For $1 \le r \le d-1$, fix any $\delta_0 \in (0,1)$. All three terms in Table 1 are $o_{d,\delta_0}(n^{d-2})$, which proves the second estimate in (C.10).

It remains to treat $r = 0$. Fix $\delta_0 \in (0,1)$ and multiply the estimates in Table 1 by n^{2-d}. As $n \to \infty$ with this cutoff fixed, we have

$$n^{2-d}D_{0,n}^{\mathrm{sep}} \longrightarrow \gamma_d^2 \mathcal{J}_d(\delta_0), \quad 0 \le n^{2-d}D_{0,n}^{\mathrm{tan}} \le C_d\delta_0^{d/2-1} + C_d n^{1-d/2}, \quad \text{and} \quad n^{2-d}D_{0,n}^{\mathrm{nor}} \longrightarrow 0.$$

Since the regional terms are nonnegative and $n^{1-d/2} \to 0$, we have

$$\gamma_d^2 \mathcal{J}_d(\delta_0) \le \liminf_{n\to\infty} n^{2-d}D_{0,n} \le \limsup_{n\to\infty} n^{2-d}D_{0,n} \le \gamma_d^2 \mathcal{J}_d(\delta_0) + C_d\delta_0^{d/2-1}. \tag{C.12}$$

The total $D_{0,n}$ is independent of δ_0. Taking $\delta_0 = 1/2$ in the upper bound of (C.12), and using $\mathcal{J}_d(1/2) < \infty$, shows that

$$L := \limsup_{n\to\infty} n^{2-d}D_{0,n} < \infty.$$

The lower bound in (C.12) then implies $\gamma_d^2 \mathcal{J}_d(\delta_0) \le L$ for every $\delta_0 \in (0,1)$. As $\delta_0 \downarrow 0$, the angular integration domains increase to $\{(u,v) : u \ne \pm v\}$. Since the integrand is nonnegative, monotone convergence yields

$$\lim_{\delta_0 \downarrow 0} \mathcal{J}_d(\delta_0) = \mathcal{J}_d \le L/\gamma_d^2 < \infty.$$

Finally, let $\delta_0 \downarrow 0$ in (C.12). Since $\mathcal{J}_d(\delta_0) \uparrow \mathcal{J}_d$ and $C_d\delta_0^{d/2-1} \to 0$, both the lower and upper bounds tend to $\gamma_d^2 \mathcal{J}_d$. This proves the first estimate in (C.10) and completes the proof. □

C.4 Constant-probability lower bound

For the following result, recall that M_d and λ_d are the constants provided by Lemma C.1, and $N_{d,n}$ was defined in (C.7). We show via numerical computation in Appendix C.5 that we can take $c_\star = 0.136$ in the following theorem.

Theorem C.4. *For any $d \geq 4$, we have*

$$\liminf_{n\to\infty} \mathbb{P}(N_{d,n} > 0) \geq c_\star,$$

where $c_\star > 0$ is a universal constant independent of d.

Proof. By rotational invariance, we may assume without loss of generality that $x^\star = e$, where e is the first standard basis vector. The Paley–Zygmund inequality implies that

$$\mathbb{P}(N_{d,n} > 0) \geq \frac{(\mathbb{E}N_{d,n})^2}{\mathbb{E}N_{d,n}^2}.$$

Theorems C.2 and C.3 then imply, at fixed d,

$$\liminf_{n\to\infty} \mathbb{P}(N_{d,n} > 0) \geq \frac{\mathcal{I}_d^2}{\mathcal{J}_d},$$

where

$$\mathcal{I}_d = \int_{\mathbb{S}^{d-2}} \int_{[-1,1]^2} \varphi_{\Sigma_{\tau u}}(-\kappa z_1 e + \kappa z_2 u)\, \mathrm{d}z \mathrm{d}\sigma(u), \quad \text{and}$$

$$\mathcal{J}_d = \int_{\substack{u,v\in\mathbb{S}^{d-2}\\ u\neq\pm v}} \int_{[-1,1]^4} \varphi_{\Sigma_{\tau u,\tau v}} \left(\begin{pmatrix} -\kappa z_1 e + \kappa z_2 u \\ -\kappa w_1 e + \kappa w_2 v \end{pmatrix} \right) \mathrm{d}z\, \mathrm{d}w\, \mathrm{d}\sigma(u)\, \mathrm{d}\sigma(v).$$

Here $\kappa = 8\tau/[\pi(1+\tau^2)^2]$, and φ_Σ is the density of $N(0,\Sigma)$. By Lemma D.8, we have

$$\Sigma_{\tau u} = \tau^2(I_d + 2uu^\top), \quad \text{and} \tag{C.13}$$

$$\Sigma_{\tau u,\tau v} = \begin{pmatrix} \tau^2(I_d + 2uu^\top) & C(u,v) \\ C(u,v)^\top & \tau^2(I_d + 2vv^\top) \end{pmatrix}, \tag{C.14}$$

where

$$C(u,v) = \tau^2 \mathbb{E}\left[(a^\top u)(a^\top v) \mathrm{sign}\left(\tau^2(a^\top u)^2 - (a^\top e)^2\right) \mathrm{sign}\left(\tau^2(a^\top v)^2 - (a^\top e)^2\right) aa^\top \right].$$

We claim that

$$\mathcal{I}_d = \kappa^{-2} p_\tau \Lambda_d, \qquad \text{and} \qquad \mathcal{J}_d \leq \kappa^{-4} C \Lambda_d^2, \tag{C.15}$$

where

$$\Lambda_d = \sigma(\mathbb{S}^{d-2})(2\pi\tau^2)^{-(d-2)/2}, \quad \text{and} \quad p_\tau = \left(2\Phi\left(\frac{\kappa}{\tau}\right) - 1\right)\left(2\Phi\left(\frac{\kappa}{\sqrt{3}\,\tau}\right) - 1\right) > 0, \tag{C.16}$$

where Φ denotes the standard Gaussian distribution function, and $C < \infty$ is a numerical constant. Since τ and κ are numerical constants independent of d, these bounds will imply the theorem.

The roadmap for the rest of the proof goes as follows. We first characterize $\mathcal{I}_d$, then bound $\mathcal{J}_d$ in terms of the joint tangential density. Finally, we show that the constant in the latter bound can be chosen independently of d.

Step 1: Characterizing $\mathcal{I}_d$. For each $u, v \in \mathbb{S}^{d-2}$, let (G_u, G_v) be an $\mathbb{R}^{2d}$ Gaussian vector with mean zero and covariance $\Sigma_{\tau u,\tau v}$.

For each u, choose an orthonormal basis $b_1, \ldots, b_{d-2}$ of $\{e,u\}^\perp$, and set

$$H_u = (b_1^\top G_u, \ldots, b_{d-2}^\top G_u)^\top, \quad \text{and} \quad N_u = (e^\top G_u, u^\top G_u).$$

Equation (C.13) implies that H_u and N_u are independent, with

$$H_u \sim N(0, \tau^2 I_{d-2}), \qquad \text{and} \qquad N_u \sim N\big(0, \operatorname{diag}(\tau^2, 3\tau^2)\big).$$

We note that the law of H_u is independent of the choice of orthonormal basis due to rotational invariance. Fix $u \in \mathbb{S}^{d-2}$. The vector $-\kappa z_1 e + \kappa z_2 u$ has zero component in $\{e, u\}^\perp$, and its normal coordinates are $(-\kappa z_1, \kappa z_2)$. Set $(s,t) = (-\kappa z_1, \kappa z_2)$, so that $\mathrm{d}z = \kappa^{-2}\mathrm{d}s\,\mathrm{d}t$ and $(s,t) \in [-\kappa,\kappa]^2$. Then, we have

$$\begin{aligned}
\int_{[-1,1]^2} \varphi_{\Sigma_{\tau u}}(-\kappa z_1 e + \kappa z_2 u)\,\mathrm{d}z &= p_{H_u}(0) \int_{[-1,1]^2} p_{N_u}(-\kappa z_1, \kappa z_2)\,\mathrm{d}z \\
&= \kappa^{-2} p_{H_u}(0) \int_{[-\kappa,\kappa]^2} p_{N_u}(s,t)\,\mathrm{d}s\,\mathrm{d}t \\
&= \kappa^{-2}(2\pi\tau^2)^{-(d-2)/2} \left(2\Phi\left(\frac{\kappa}{\tau}\right) - 1\right)\left(2\Phi\left(\frac{\kappa}{\sqrt{3}\,\tau}\right) - 1\right),
\end{aligned}$$

where the last equality uses $p_{H_u}(0) = (2\pi\tau^2)^{-(d-2)/2}$ and the fact that the two coordinates of N_u are independent with laws $N(0,\tau^2)$ and $N(0,3\tau^2)$, respectively. The integral above is constant in $u \in \mathbb{S}^{d-2}$. Integrating over u, we have

$$\mathcal{I}_d = \frac{\sigma(\mathbb{S}^{d-2})}{\kappa^2(2\pi\tau^2)^{(d-2)/2}} \left(2\Phi\left(\frac{\kappa}{\tau}\right) - 1\right)\left(2\Phi\left(\frac{\kappa}{\sqrt{3}\,\tau}\right) - 1\right).$$

This establishes the first equality in (C.15).

Step 2: Bounding $\mathcal{J}_d$. For $u \neq \pm v$, Lemma D.8(III) ensures that the joint density exists. Set $\xi(z,w) = (-\kappa z_1, \kappa z_2, -\kappa w_1, \kappa w_2)$. Since the coordinate transformation from (G_u, G_v) to (H_u, H_v, N_u, N_v) is orthogonal,

$$\begin{aligned}
p_{G_u,G_v}(-\kappa z_1 e + \kappa z_2 u,\, -\kappa w_1 e + \kappa w_2 v) &= p_{H_u,H_v,N_u,N_v}(0,0,\xi(z,w)) \\
&= p_{H_u,H_v}(0,0)\, p_{(N_u,N_v)|H_u=H_v=0}(\xi(z,w)),
\end{aligned}$$

where the second factor is the Gaussian density of (N_u, N_v) conditional on $(H_u, H_v) = (0,0)$. The change of variables $(z,w) \mapsto \xi(z,w)$ has absolute Jacobian κ^4. Hence, for fixed $u, v \in \mathbb{S}^{d-2}$ with $u \neq \pm v$,

$$\begin{aligned}
\int_{[-1,1]^4} p_{(N_u,N_v)|H_u=H_v=0}(\xi(z,w))\,\mathrm{d}z\,\mathrm{d}w &= \kappa^{-4} \int_{[-\kappa,\kappa]^4} p_{(N_u,N_v)|H_u=H_v=0}(\xi)\,\mathrm{d}\xi \\
&= \kappa^{-4}\mathbb{P}\big((N_u,N_v) \in [-\kappa,\kappa]^4 \mid H_u = H_v = 0\big).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathcal{J}_d &= \kappa^{-4} \int_{\substack{u,v\in\mathbb{S}^{d-2} \\ u\neq\pm v}} p_{H_u,H_v}(0,0)\mathbb{P}\big((N_u,N_v) \in [-\kappa,\kappa]^4 \mid H_u = H_v = 0\big)\,\mathrm{d}\sigma(u)\,\mathrm{d}\sigma(v) \\
&\leq \kappa^{-4} \int_{\substack{u,v\in\mathbb{S}^{d-2} \\ u\neq\pm v}} p_{H_u,H_v}(0,0)\,\mathrm{d}\sigma(u)\,\mathrm{d}\sigma(v) =: \kappa^{-4} Q_d.
\end{aligned}$$

It remains to prove $Q_d \leq C\sigma(\mathbb{S}^{d-2})^2(2\pi\tau^2)^{-(d-2)}$ uniformly for $d \geq 4$.

Changing the orthonormal bases of $\{e,u\}^\perp$ and $\{e,v\}^\perp$ leaves $p_{H_u,H_v}(0,0)$ unchanged. We therefore choose the following bases to simplify its computation. Fix $u, v \in \mathbb{S}^{d-2}$ with $u \neq \pm v$, put $c = u^\top v$ and $s = \sqrt{1-c^2}$, and choose an orthonormal basis $t_1, \ldots, t_{d-3}$ of $\{e,u,v\}^\perp$. Define

$$h_u = \frac{v - cu}{s}, \qquad \text{and} \qquad h_v = \frac{cv - u}{s}.$$

The tuples $(h_u, t_1, \ldots, t_{d-3})$ and $(h_v, t_1, \ldots, t_{d-3})$ are orthonormal bases of $\{e, u\}^\perp$ and $\{e, v\}^\perp$, respectively. Under this choice of bases,

$$H_u = \begin{pmatrix} h_u^\top G_u \\ t_1^\top G_u \\ \vdots \\ t_{d-3}^\top G_u \end{pmatrix}, \qquad \text{and} \qquad H_v = \begin{pmatrix} h_v^\top G_v \\ t_1^\top G_v \\ \vdots \\ t_{d-3}^\top G_v \end{pmatrix}.$$

Equation (C.13) implies $\mathrm{Cov}(H_u) = \mathrm{Cov}(H_v) = \tau^2 I_{d-2}$. For a standard Gaussian vector $a \in \mathbb{R}^d$, define the following real standard Gaussian variables:

$$P = a^\top e, \;\; X = a^\top u, \;\; Y = a^\top v, \;\; Z = a^\top h_u, \;\; T = a^\top h_v, \quad \text{and} \quad L_j = a^\top t_j \;\; (1 \le j \le d-3).$$

The variables $P, X, Z, L_1, \ldots, L_{d-3}$ are independent, while $Y = cX + sZ$ and $T = -sX + cZ$. Also define

$$S_X = \mathrm{sign}(\tau^2 X^2 - P^2), \qquad \text{and} \qquad S_Y = \mathrm{sign}(\tau^2 Y^2 - P^2).$$

By the definition of $C(u, v)$ in (C.14),

$$\begin{aligned} \mathrm{Cov}(h_u^\top G_u, h_v^\top G_v) &= h_u^\top C(u, v) h_v \\ &= \tau^2 \mathbb{E}[(a^\top u)(a^\top v) S_X S_Y (a^\top h_u)(a^\top h_v)] \\ &= \tau^2 \mathbb{E}[XY S_X S_Y Z T]. \end{aligned}$$

Likewise, for $1 \le i, j \le d-3$, independence of the L_j from (P, X, Z) implies

$$\begin{aligned} \mathrm{Cov}(t_i^\top G_u, t_j^\top G_v) &= t_i^\top C(u, v) t_j = \tau^2 \mathbb{E}[XY S_X S_Y L_i L_j] = \tau^2 \mathbb{E}[XY S_X S_Y] \delta_{ij}, \\ \mathrm{Cov}(h_u^\top G_u, t_j^\top G_v) &= h_u^\top C(u, v) t_j = \tau^2 \mathbb{E}[XY S_X S_Y Z L_j] = 0, \quad \text{and} \\ \mathrm{Cov}(t_i^\top G_u, h_v^\top G_v) &= t_i^\top C(u, v) h_v = \tau^2 \mathbb{E}[XY S_X S_Y L_i T] = 0. \end{aligned}$$

Since P, X, Z are independent standard Gaussians and $Y = cX + sZ$, $T = -sX + cZ$, these covariances depend on u, v only through their inner product c. We write

$$r(c) = \mathbb{E}[XZYT S_X S_Y], \qquad \text{and} \qquad q(c) = \mathbb{E}[XY S_X S_Y].$$

Thus

$$\mathrm{Cov} \begin{pmatrix} H_u \\ H_v \end{pmatrix} = \tau^2 \begin{pmatrix} I_{d-2} & D \\ D & I_{d-2} \end{pmatrix}, \qquad \text{and} \qquad D = \mathrm{diag}(r(c), q(c), \ldots, q(c)).$$

By the Schur complement,

$$\begin{aligned} \det \mathrm{Cov} \begin{pmatrix} H_u \\ H_v \end{pmatrix} &= \tau^{4(d-2)} \det \begin{pmatrix} I_{d-2} & D \\ D & I_{d-2} \end{pmatrix} \\ &= \tau^{4(d-2)} \det(I_{d-2} - D^2) = \tau^{4(d-2)} (1 - r(c)^2)(1 - q(c)^2)^{d-3}. \end{aligned}$$

Since (H_u, H_v) is a centered Gaussian vector in $\mathbb{R}^{2(d-2)}$, its density at zero is

$$p_{H_u, H_v}(0, 0) = (2\pi\tau^2)^{-(d-2)} (1 - r(c)^2)^{-1/2} (1 - q(c)^2)^{-(d-3)/2}. \tag{C.17}$$

Step 3: Uniform bound in dimension. We verify the scalar hypotheses of Lemma F.5 numerically using interval arithmetic in Section C.5. Thus Lemmas F.4 and F.5 give, with $b = \gamma > 0$,

$$1 - q(c)^2 \ge (1 - c^2)(1 + \gamma c^2), \qquad \text{and} \qquad 1 - r(c)^2 \ge b\sqrt{1 - c^2}.$$

The density formula (C.17) implies

$$\begin{aligned} Q_d &= (2\pi\tau^2)^{-(d-2)} \int_{\substack{u,v\in\mathbb{S}^{d-2}\\ u\neq\pm v}} (1-r(u^\top v)^2)^{-1/2}(1-q(u^\top v)^2)^{-(d-3)/2}\,\mathrm{d}\sigma(u)\,\mathrm{d}\sigma(v) \\ &= \sigma(\mathbb{S}^{d-2})^2(2\pi\tau^2)^{-(d-2)} \int_{\substack{u,v\in\mathbb{S}^{d-2}\\ u\neq\pm v}} (1-r(u^\top v)^2)^{-1/2}(1-q(u^\top v)^2)^{-(d-3)/2}\,\mathrm{d}u\,\mathrm{d}v \\ &\le b^{-1/2}\sigma(\mathbb{S}^{d-2})^2(2\pi\tau^2)^{-(d-2)} A_d \int_{-1}^{1} (1-c^2)^{-3/4}(1+\gamma c^2)^{-(d-3)/2}\,\mathrm{d}c. \end{aligned}$$

Here $\mathrm{d}u$ and $\mathrm{d}v$ denote the uniform probability measures on $\mathbb{S}^{d-2}$. Under their product measure, $c = u^\top v$ has density

$$A_d(1-c^2)^{(d-4)/2} \qquad \text{where } A_d = \frac{\Gamma((d-1)/2)}{\sqrt{\pi}\,\Gamma((d-2)/2)} \le \frac{\sqrt{d-2}}{\sqrt{2\pi}}.$$

The last inequality follows directly from Gautschi's inequality. In the following estimates, $C > 0$ denotes a constant independent of d that may change from line to line. Since $\log(1+\gamma c^2) \ge \gamma c^2/(1+\gamma/4)$ on $|c| \le 1/2$, we have, for $d \ge 4$,

$$\begin{aligned} \int_{-1}^{1} (1-c^2)^{-3/4}(1+\gamma c^2)^{-(d-3)/2}\,\mathrm{d}c &\le C\int_{-1/2}^{1/2} e^{-\gamma(d-3)c^2/[2(1+\gamma/4)]}\,\mathrm{d}c \\ &\quad + (1+\gamma/4)^{-(d-3)/2}\int_{1/2<|c|<1} (1-c^2)^{-3/4}\,\mathrm{d}c \\ &\le C(d-3)^{-1/2} + C(1+\gamma/4)^{-(d-3)/2} \le C(d-2)^{-1/2}. \end{aligned}$$

Since $A_d \le \sqrt{(d-2)/(2\pi)}$, we obtain $Q_d \le C\sigma(\mathbb{S}^{d-2})^2(2\pi\tau^2)^{-(d-2)}$ uniformly in d. This proves the second bound in (C.15).

Finally, (C.15) implies

$$\frac{\mathcal{I}_d^2}{\mathcal{J}_d} \ge \frac{p_\tau^2}{C} =: c_\star > 0 \qquad \text{for } d \ge 4.$$

Combining this bound with the earlier Paley–Zygmund estimate proves the theorem. □

C.5 Numerical certification of the probability bound

In this appendix, we make the probability bound in Theorem C.4 explicit. The preceding proof gives, for every fixed $d \ge 4$,

$$\liminf_{n\to\infty} \mathbb{P}(N_{d,n} > 0) \ge \frac{\mathcal{I}_d^2}{\mathcal{J}_d} \ge \frac{p_\tau^2}{C}, \tag{C.18}$$

where C is any constant satisfying (C.15) uniformly in d. Combining analytic estimates with certified interval computations, we show that $C = 3.054$ is admissible and that $p_\tau^2/3.054 > 0.1361$. This establishes the choice $c_\star = 0.136$. We defer the details for the reproduction of these results to the end of this appendix.

Scalar parameters and correlation bounds. Recall that τ is the positive root of

$$\arctan t + \frac{t}{1+t^2} - \frac{\pi}{4} = 0.$$

The derivative of the left-hand side is $2/(1+t^2)^2 > 0$. We enclose τ between rational endpoints by verifying, with interval arithmetic, that the left-hand side has opposite signs at the endpoints. For

the coefficient w_0 in the proof of Lemma F.4, we use the identity

$$w_0 = 1 - \frac{8}{\pi}\arctan\frac{1}{\tau} + \frac{8}{\pi}\arcsin\frac{1}{1+\tau^2} + \frac{8\tau}{\pi}\left(\frac{1}{1+\tau^2} - \frac{2}{(1+\tau^2)\sqrt{\tau^2+2}} + \frac{1}{(\tau^2+2)^{3/2}}\right). \tag{C.19}$$

Evaluating this expression and $\gamma = 1 - w_0^2$ on the certified interval for τ gives

$$\begin{aligned} &0.44161079 < \tau < 0.44161080, \\ &0.60010962 < w_0 < 0.60010963, \quad \text{and} \\ &0.63986844 < \gamma < 0.63986845. \end{aligned}$$

The displayed enclosures are rounded outward from the narrower intervals used in the computations. In particular,

$$\frac{2}{5} \le \tau \le \frac{9}{20}, \qquad w_0 \ge \frac{3}{5}, \qquad \text{and} \qquad \frac{4}{47} < \gamma < \frac{16}{25}. \tag{C.20}$$

Lemma F.4 gives

$$1 - q(c)^2 \ge (1-c^2)(1+\gamma c^2) \qquad \text{for } -1 \le c \le 1. \tag{C.21}$$

The first two bounds in (C.20) allow us to apply Lemma F.5, giving

$$1 - r(c)^2 \ge \gamma\sqrt{1-c^2} \qquad \text{for } -1 < c < 1, \tag{C.22}$$

with equality if, and only if, $c = 0$. Thus we may take $b = \gamma$ in the preceding proof.

Uniform bound for the angular integral. Combining the correlation bounds (C.21) and (C.22) with the density formula (C.17) shows that (C.15) holds whenever

$$C \ge \gamma^{-1/2} \sup_{d \ge 4} H_d(\gamma), \tag{C.23}$$

where

$$H_d(\gamma) = A_d \int_{-1}^{1} (1-c^2)^{-3/4}(1+\gamma c^2)^{-(d-3)/2}\,\mathrm{d}c, \quad \text{and} \quad A_d = \frac{\Gamma((d-1)/2)}{\sqrt{\pi}\,\Gamma((d-2)/2)}.$$

The weight $(1-c^2)^{-3/4}$ is obtained by bounding the inverse covariance factors and multiplying by the spherical correlation density $A_d(1-c^2)^{(d-4)/2}$. Next, we consider two cases.

Case 1: $4 \le d \le 49$. The substitution $u = c^2$ and the binomial expansion give

$$H_d(\gamma) = A_d B(1/2, 1/4) \sum_{n=0}^{\infty} (-1)^n a_n,$$

$$a_0 = 1, \qquad \text{and} \qquad a_{n+1} = a_n \gamma \frac{((d-3)/2+n)(1/2+n)}{(3/4+n)(n+1)},$$

where $B(1/2,1/4) = \Gamma(1/2)\Gamma(1/4)/\Gamma(3/4)$. For $0 \le \gamma \le 16/25$, the nonnegative terms satisfy $a_{n+1} \le (4/5)a_n$ for all $n \ge 100$. Consequently, for every even $N \ge 100$, the partial sum $S_N = \sum_{n=0}^{N}(-1)^n a_n$ satisfies

$$A_d B(1/2,1/4)(S_N - a_{N+1}) \le H_d(\gamma) \le A_d B(1/2,1/4) S_N. \tag{C.24}$$

We evaluate the partial sums and the first omitted term using interval arithmetic. Together with (C.24), this accounts for both rounding and truncation errors and yields

$$\begin{aligned} &H_d(\gamma) < H_6(\gamma) \quad \text{for } 4 \le d \le 49 \text{ with } d \ne 6, \quad \text{and} \\ &2.44237353 < H_6(\gamma) < 2.44237354. \end{aligned} \tag{C.25}$$

For each comparison, the upper endpoint of the interval for $H_d(\gamma)$ lies below the lower endpoint of the interval for $H_6(\gamma)$.

Case 2: $d \geq 50$. We use the analytic estimate

$$H_d(\gamma) \leq \left(\frac{4}{3}\right)^{3/4} \sqrt{\frac{1+\gamma/4}{\gamma}} \sqrt{\frac{48}{47}} + B(1/2, 1/4) \sqrt{\frac{24}{\pi}} \exp\left(-\frac{47}{2}\log(1+\gamma/4)\right). \tag{C.26}$$

This bound holds when $\gamma > 0$ and $48\log(1+\gamma/4) \geq 1$, as ensured by (C.20). Its proof splits the integral at $|c| = 1/2$, applies a Gaussian bound on the middle interval and a beta-integral bound on the remainder, and uses monotonicity in d. Interval evaluation of the right-hand side gives a value less than 2.131647, below the lower bound for $H_6(\gamma)$ in (C.25).

Combining the two cases, we obtain

$$\sup_{d \geq 4} H_d(\gamma) = H_6(\gamma).$$

Since $\gamma > 0.639868$, the displayed enclosures imply

$$\gamma^{-1/2} \sup_{d \geq 4} H_d(\gamma) = \frac{H_6(\gamma)}{\sqrt{\gamma}} < \frac{2.44237354}{\sqrt{0.639868}} < 3.054.$$

Thus $C = 3.054$ satisfies (C.23).

Evaluation of the probability bound. By (C.16),

$$p_\tau = \operatorname{erf}\left(\frac{\kappa}{\sqrt{2}\,\tau}\right) \operatorname{erf}\left(\frac{\kappa}{\sqrt{6}\,\tau}\right), \qquad \text{and} \qquad \kappa = \frac{8\tau}{\pi(1+\tau^2)^2},$$

where $\operatorname{erf}(x) = 2\Phi(\sqrt{2}x) - 1$. Both arguments are nonnegative and have squares less than 3, as verified by interval arithmetic. In this range, the terms of the alternating series for erf decrease in absolute value, giving

$$\left|\operatorname{erf}(x) - \frac{2}{\sqrt{\pi}} \sum_{n=0}^{N} \frac{(-1)^n x^{2n+1}}{n!(2n+1)}\right| \leq \frac{2}{\sqrt{\pi}} \frac{x^{2N+3}}{(N+1)!(2N+3)} \qquad \text{for } N \geq 0.$$

Interval evaluation with this remainder gives

$$0.64480748 < p_\tau < 0.64480749.$$

Combining this enclosure with (C.18), we conclude that, for every fixed $d \geq 4$,

$$\liminf_{n \to \infty} \mathbb{P}(N_{d,n} > 0) \geq \frac{p_\tau^2}{3.054} > \frac{(0.64480748)^2}{3.054} > 0.1361.$$

Since $N_{d,n} \leq M_{d,n}$, this establishes the choice $c_\star = 0.136$ in Theorems C.4 and 3.3.

Verification and reproduction. The accompanying Lean formalization verifies the identity (C.19), the correlation bounds, the angular series and its remainder estimate, and the uniform bound (C.26). The moment theorems and covariance reduction are proved in the paper but are not formalized in Lean. The numerical enclosures are computed separately, outside Lean, using interval arithmetic with outward rounding and the explicit remainder bounds above; the error-function evaluation is also outside the formalization. All displayed terminating decimals denote exact rational numbers. Supplementary quadrature checks are not used to certify the inequalities.

For the source code, full-precision certificates, and reproduction instructions, see `https://github.`

com/PenghaoWang1225/Phase-Retrieval. The formal proofs are in `04-Lean`, and the numerical computations are in `01-numerical-lower-bound`. The repository documentation records the formal declarations and computational details. In the code and formal proofs, the dimension parameter is $d-2$ rather than d.

D Probabilistic tools for counting minima

This appendix collects the probabilistic tools used to compute the first and second moments of $N_{d,n}$ in Appendix C, using the notation from Appendix C.1. We prove Lemma C.1 and cover joint density bounds, Kac–Rice formulas, uniform local limit approximations, and mean and covariance estimates near the population saddle sphere.

D.1 Proof of Lemma C.1

Proof of Lemma C.1. By rotational invariance of the standard Gaussian measurements, we may assume without loss of generality that $x^\star = e$, where e is the first standard basis vector. Fix $I = \{i_1 < \cdots < i_d\}$. Since all the vectors a_i are i.i.d. and every ordered active set uses the same sign pattern, the conditional expectation $a(x)$ is independent of the choice of I.

We first show that $a(\tau u) > 0$. Fix $u \in \mathbb{S}^{d-2}$. Conditional on $X_I(\tau u) = 0$, the vectors a_{i_j} are independent standard Gaussian vectors on the hyperplanes $(\tau u - s_j e)^\perp$, for $j = 1, \dots, d$. We will construct a deterministic matrix whose rows lie in these hyperplanes and which satisfies the inequalities defining $\mathcal{K}_{\tau u}$ strictly. Choose an orthonormal basis $v_1, \dots, v_{d-2}$ of $\{e, u\}^\perp$, and let B_0 be the matrix with rows $b_j^\top$, where

$$\begin{aligned} b_1 &= \tau e + u, & b_j &= \tau e + u + v_{j-1} \quad \text{for } 2 \le j \le d-2, \\ b_{d-1} &= -\tau e + u, \quad \text{and} & b_d &= -\tau e + u + v_{d-2}. \end{aligned}$$

Each vector b_j satisfies $b_j^\top(\tau u - s_j e) = 0$ and $b_j^\top(\tau u) = \tau$. To compute the determinant, subtract row 1 from rows $2, \dots, d-2$ and row $d-1$ from row d. Reorder the resulting rows as $b_1^\top, b_{d-1}^\top, v_1^\top, \dots, v_{d-2}^\top$, and express them in coordinates relative to the orthonormal basis $(e, u, v_1, \dots, v_{d-2})$. These operations preserve the absolute determinant, so

$$|\det B_0| = \left| \det \begin{pmatrix} \tau & 1 & 0 \\ -\tau & 1 & 0 \\ 0 & 0 & I_{d-2} \end{pmatrix} \right| = \left| \det \begin{pmatrix} \tau & 1 \\ -\tau & 1 \end{pmatrix} \right| = 2\tau.$$

Also, $b_j^\top(\tau u) = \tau$ implies $Q(\tau u; B_0) = \tau B_0^\top$, and hence $|\det Q(\tau u; B_0)| = 2\tau^{d+1}$.

Since $\max_j \|b_j\| = \sqrt{\tau^2 + 2}$ and $0 < \tau < 1$, choose $M_d > \sqrt{\tau^2 + 2}$ and $0 < \lambda_d < 2\tau^{d+1}$. All three inequalities defining $\mathcal{K}_{\tau u}$ are then strict at B_0. By continuity, they also hold for all matrices sufficiently close to B_0 whose rows remain in the conditioning hyperplanes. The joint conditional Gaussian density is positive on the product of these hyperplanes, so this neighborhood has positive conditional probability. Thus

$$\mathbb{P}\left(A_I \in \mathcal{K}_{\tau u} \mid X_I(\tau u) = 0\right) > 0.$$

On $\mathcal{K}_{\tau u}$, both absolute determinants in the definition of a are at least λ_d, so

$$a(\tau u) \ge \lambda_d^2 \, \mathbb{P}\left(A_I \in \mathcal{K}_{\tau u} \mid X_I(\tau u) = 0\right) > 0.$$

It remains to show that $a(\tau u)$ does not depend on u. For any $v \in \mathbb{S}^{d-2}$, choose an orthogonal map R with $Re = e$ and $Ru = v$. This map sends each hyperplane $(\tau u - s_j e)^\perp$ to $(\tau v - s_j e)^\perp$ and preserves the standard Gaussian law on it. Therefore, the law of $A_I R^\top$ conditional on $X_I(\tau u) = 0$

is the law of A_I conditional on $X_I(\tau v) = 0$. Rotation preserves the row norms and the absolute determinant of A_I, and

$$Q(\tau v; A_I R^\top) = RQ(\tau u; A_I).$$

Consequently, $A_I \in \mathcal{K}_{\tau u}$ if, and only if, $A_I R^\top \in \mathcal{K}_{\tau v}$. Since R also preserves the two absolute determinant factors in the definition of a, applying this rotation to the conditional expectation shows that $a(\tau u) = a(\tau v)$ for all $u, v \in \mathbb{S}^{d-2}$. Denote this common positive value by ϑ_d.

We next prove continuity near $\mathcal{S}$. Fix an integer n_0 with $n_0^{-1/2} < \tau$. The set T_{n_0} is compact, its interior contains $\mathcal{S}$, and $x \neq \pm e$ for every $x \in T_{n_0}$. To compare the conditional laws at nearby points, let $g_1, \dots, g_d$ be independent $N(0, I_d)$ vectors and use these same vectors to define, for every $x \in T_{n_0}$,

$$\widehat{a}_j(x) = g_j - \frac{g_j^\top (x - s_j e)}{\|x - s_j e\|^2}(x - s_j e) \qquad \text{for } j = 1, \dots, d,$$

and let $\widehat{A}(x)$ be the matrix with rows $\widehat{a}_j(x)^\top$. The vector $\widehat{a}_j(x)$ is the orthogonal projection of g_j onto $(x - s_j e)^\perp$. Thus $\widehat{A}(x)$ has the law of A_I conditional on $X_I(x) = 0$, and

$$a(x) = \mathbb{E}_{g_1,\dots,g_d}\Big[\Big|\det \widehat{A}(x)\Big| \Big|\det Q(x; \widehat{A}(x))\Big| \, 1_{\mathcal{K}_x}\big(\widehat{A}(x)\big)\Big].$$

Fix $x \in T_{n_0}$ and a sequence $x_j \in T_{n_0}$ with $x_j \to x$. The projection formula implies $\widehat{A}(x_j) \to \widehat{A}(x)$ almost surely, so both determinant factors converge almost surely. To handle the indicator, we must also exclude equality in the inequalities defining $\mathcal{K}_x$. For this fixed x, each such equality requires one of the polynomials

$$\|\widehat{a}_j(x)\|^2 - M_d^2, \qquad \big(\det \widehat{A}(x)\big)^2 - \lambda_d^2, \qquad \text{and} \qquad \big(\det Q(x; \widehat{A}(x))\big)^2 - \lambda_d^2$$

in the components of $(g_1, \dots, g_d)$ to vanish. None is identically zero: when all $g_j = 0$, their values are $-M_d^2$ or $-\lambda_d^2$. Lemma F.1 therefore shows that these equalities occur with probability zero. Since all the quantities defining $\mathcal{K}_x$ depend continuously on x and the matrix entries, it follows that

$$1_{\mathcal{K}_{x_j}}\big(\widehat{A}(x_j)\big) \longrightarrow 1_{\mathcal{K}_x}\big(\widehat{A}(x)\big) \qquad \text{almost surely.}$$

Hence the full integrand defining $a(x_j)$ converges almost surely to that defining $a(x)$. To pass to expectations, note that whenever $\widehat{A}(x) \in \mathcal{K}_x$, Hadamard's inequality and the row bounds imply

$$\Big|\det \widehat{A}(x)\Big| \le M_d^d, \qquad \text{and} \qquad \Big|\det Q(x; \widehat{A}(x))\Big| \le \prod_{j=1}^{d} \Big(\Big|\widehat{a}_j(x)^\top x\Big| \, \|\widehat{a}_j(x)\|\Big) \le \|x\|^d \, M_d^{2d}.$$

Since $\|x\| \le \tau + 2n_0^{-1/2}$ on T_{n_0}, the integrand is bounded by the deterministic constant $(\tau + 2n_0^{-1/2})^d M_d^{3d}$ for every $x \in T_{n_0}$. Dominated convergence now implies $a(x_j) \to a(x)$, proving continuity on T_{n_0} and hence near $\mathcal{S}$.

Finally, continuity on the compact set T_{n_0} implies uniform continuity there. For $n \ge n_0$ and $x = \alpha e + (\tau + \beta)u \in T_n$, both x and τu lie in T_{n_0}, and

$$\|x - \tau u\| = \sqrt{\alpha^2 + \beta^2} \le \sqrt{2} n^{-1/2}.$$

This bound holds uniformly over $x \in T_n$. Since $a(\tau u) = \vartheta_d$, uniform continuity implies

$$\sup_{x \in T_n} |a(x) - \vartheta_d| \to 0 \qquad \text{as } n \to \infty.$$

□

D.2 Joint density bounds

Recall the active equations X_I and their solutions x_I from (C.5) and (C.6). For an active set $I = \{i_1 < \cdots < i_d\}$, write

$$s^I_{i_k} = s_k \qquad \text{for } 1 \le k \le d$$

for the row signs induced by the fixed pattern (C.2). For distinct active sets I, J and $x, y \in \mathbb{R}^d$, define

$$\Gamma_{I,J}(x,y) = \operatorname{Cov}\begin{pmatrix} X_I(x) \\ X_J(y) \end{pmatrix}.$$

For $v, w \in \mathbb{R}^d$, set

$$\Delta(v,w) = \sqrt{\|v\|^2 \|w\|^2 - (v^\top w)^2}.$$

Lemma D.1. *Fix integers $d \ge 4$ and $n \ge d$, and a deterministic vector $x^\star \in \mathbb{R}^d$ with $\|x^\star\| = 1$. Let $a_1, \dots, a_n$ be independent $N(0, I_d)$ vectors. Then, for every pair of distinct index sets $I, J \subset [n]$ with $|I| = |J| = d$ and every $x, y \in \mathbb{R}^d$,*

$$\det \Gamma_{I,J}(x,y) = \prod_{i \in I \setminus J} \left\|x - s^I_i x^\star\right\|^2 \prod_{j \in J \setminus I} \left\|y - s^J_j x^\star\right\|^2 \prod_{\ell \in I \cap J} \Delta^2(x - s^I_\ell x^\star, y - s^J_\ell x^\star). \tag{D.1}$$

Moreover, almost surely, this covariance map is positive definite at the random points (x_I, x_J) for every such pair I, J:

$$\Gamma_{I,J}(x_I, x_J) \succ 0.$$

Proof. By rotational invariance of the standard Gaussian measurements, we may assume without loss of generality that $x^\star = e$, where e is the first standard basis vector. Fix $I \ne J$. For $i \in I \setminus J$, the coordinate $a_i^\top(x - s^I_i e)$ is independent of all other coordinates and has variance $\left\|x - s^I_i e\right\|^2$; the same holds for $j \in J \setminus I$. For each $\ell \in I \cap J$, the two coordinates involving a_ℓ have covariance matrix

$$\begin{pmatrix} \left\|x - s^I_\ell e\right\|^2 & (x - s^I_\ell e)^\top (y - s^J_\ell e) \\ (x - s^I_\ell e)^\top (y - s^J_\ell e) & \left\|y - s^J_\ell e\right\|^2 \end{pmatrix},$$

whose determinant is $\Delta^2(x - s^I_\ell e, y - s^J_\ell e)$. Therefore,

$$\det \Gamma_{I,J}(x,y) = \prod_{i \in I \setminus J} \left\|x - s^I_i e\right\|^2 \prod_{j \in J \setminus I} \left\|y - s^J_j e\right\|^2 \prod_{\ell \in I \cap J} \Delta^2(x - s^I_\ell e, y - s^J_\ell e).$$

Almost surely, every A_I is invertible and $a_i^\top e \ne 0$ for all i. On this event, no x_I is parallel to e. Indeed, since the sign pattern s contains both 1 and -1, if $x_I = \alpha e$, then the active equations $X_I(x) = 0$ would require α to equal both 1 and -1, which is impossible. Therefore, almost surely, we have

$$\prod_{i \in I \setminus J} \left\|x_I - s^I_i e\right\|^2 \prod_{j \in J \setminus I} \left\|x_J - s^J_j e\right\|^2 > 0.$$

Since $\Gamma_{I,J}(x_I, x_J)$ is positive semidefinite, the determinant identity (D.1) shows that it remains to prove, almost surely,

$$\Delta(x_I - s^I_\ell e, x_J - s^J_\ell e) > 0 \qquad \text{for every } \ell \in I \cap J.$$

There is nothing further to check if $I \cap J = \varnothing$. Otherwise, fix $\ell \in I \cap J$. We will show that the two vectors in this expression are linearly dependent with probability zero.

Condition on the rows $(a_i)_{i \in I}$, and set $x = x_I$, $\sigma = s^I_\ell$, and $\sigma' = s^J_\ell$. Then x is fixed and is not parallel to e, while the vectors $(a_j)_{j \in J \setminus I}$ remain independent $N(0, I_d)$ vectors. Since $I \ne J$ and $|I| = |J|$, the set $J \setminus I$ is nonempty. Both vectors associated with ℓ are nonzero, so if they are linearly

dependent, there is a scalar $t \neq 0$ such that

$$x_J = \sigma' e + t(x - \sigma e).$$

For every $j \in J$, substituting this expression into $a_j^\top (x_J - s_j^J e) = 0$ implies

$$t\, a_j^\top (x - \sigma e) + (\sigma' - s_j^J) a_j^\top e = 0. \tag{D.2}$$

First, suppose some $j \in J \setminus I$ has $s_j^J = \sigma'$. Since $t \neq 0$, equation (D.2) requires $a_j^\top (x - \sigma e) = 0$. Conditional on the rows in I, this has probability zero because a_j is a standard Gaussian vector independent of those rows and $x - \sigma e \neq 0$.

It remains to consider the case in which every $j \in J \setminus I$ has sign $-\sigma'$. Suppose first that $J \setminus I$ contains two distinct indices j, k. Conditional on the rows in I, we have $a_j^\top (x - \sigma e) \neq 0$ almost surely, so equation (D.2) for row j determines

$$t = -\frac{2\sigma' a_j^\top e}{a_j^\top (x - \sigma e)}.$$

After also conditioning on a_j, this value of t is fixed, and the equation for row k requires

$$a_k^\top \big(t(x - \sigma e) + 2\sigma' e\big) = 0.$$

The vector in parentheses is nonzero because x is not parallel to e. Since a_k remains an independent standard Gaussian vector, this equation has conditional probability zero.

The only remaining case is $J \setminus I = \{k\}$ with $s_k^J = -\sigma'$. Each sign occurs at least twice in the fixed pattern because $d \geq 4$, so there is a shared row $j \in I \cap J$ with $s_j^J = -\sigma'$. Since $a_j^\top x = s_j^I a_j^\top e$, equation (D.2) for this row becomes

$$\big[t(s_j^I - \sigma) + 2\sigma'\big] a_j^\top e = 0.$$

We have $a_j^\top e \neq 0$, so $s_j^I = \sigma$ would make this equality impossible. Thus $s_j^I = -\sigma$, which forces $t = \sigma'/\sigma$. Substituting this value into the equation for row k implies

$$a_k^\top \big((\sigma'/\sigma)x + \sigma' e\big) = 0.$$

The vector in parentheses is fixed under the conditioning and is nonzero because x is not parallel to e. Since a_k is a standard Gaussian vector independent of the rows in I, this equation also has conditional probability zero.

These cases exhaust all possibilities for $J \setminus I$. Averaging over the rows in I shows that the two vectors associated with ℓ are linearly dependent with probability zero. Taking a finite union over I, J and $\ell \in I \cap J$, we conclude that all factors in (D.1) are positive almost surely at (x_I, x_J). Thus $\det \Gamma_{I,J}(x_I, x_J) > 0$, and positive semidefiniteness implies $\Gamma_{I,J}(x_I, x_J) \succ 0$ for every such pair I, J, almost surely. □

For distinct active sets I, J, let

$$\mathcal{R}_{I,J} = \{(x, y) \in \mathbb{R}^{2d} : \Gamma_{I,J}(x, y) \succ 0\},$$

and denote the joint Gaussian density on this region by $p_{X_I(x), X_J(y)}$. We use the angular separation $\delta(x, y)$ from the proof of Theorem C.3. For $x \in \mathbb{R}^d$, write $v_\pm = x \mp x^\star$, and let $P_{v^\perp}$ denote orthogonal projection onto $v^\perp$. In part (II) below, the shared-row counts are

$$r_\pm = \#\{i \in I \cap J : s_i^I = s_i^J = \pm 1\}.$$

Lemma D.2 (Joint density and angular integral bounds)**.** *Fix integers $d \geq 4$ and $0 \leq r \leq d-1$, and index sets $I, J \subset [n]$ with $|I| = |J| = d$ and $|I \cap J| = r$. The following estimates hold.*

(I) *There exist constants n_0 and $C_d > 0$, depending only on d, such that for every $n \geq n_0$ and every $(x, y) \in T_n^2 \cap \mathcal{R}_{I,J}$ with $\delta(x, y) > 0$, we have*

$$p_{X_I(x), X_J(y)}(0, 0) \leq C_d \delta(x, y)^{-r}.$$

(II) *There exist constants n_0 and $C_d > 0$, depending only on d, such that for every $n \geq n_0$ and every $(x, y) \in T_n^2 \cap \mathcal{R}_{I,J}$ with $\delta(x, y) > 0$ and $u_x^\top u_y > 0$, we have*

$$p_{X_I(x), X_J(y)}(0, 0) \leq \frac{C_d}{\left\|P_{v_+^\perp}(y - x)\right\|^{r_+} \left\|P_{v_-^\perp}(y - x)\right\|^{r_-}}. \tag{D.3}$$

(III) *For any integers $0 \leq r_+, r_- \leq d-2$, there exist constants n_0 and $C_d > 0$, depending only on d, such that for every $n \geq n_0$,*

$$\sup_{x \in T_n} \int_{\mathbb{S}^{d-1}} \frac{\mathrm{d}\sigma(\omega)}{\left\|P_{v_+^\perp}\omega\right\|^{r_+} \left\|P_{v_-^\perp}\omega\right\|^{r_-}} \leq C_d. \tag{D.4}$$

Proof. By rotational invariance of the standard Gaussian measurements, we may assume without loss of generality that $x^\star = e$, where e is the first standard basis vector. Recall the notation $\Delta(a, b)$ from Lemma D.1.

First, by Lemma D.1, we have

$$\det\left(\mathrm{Cov}\begin{pmatrix} X_I(x) \\ X_J(y) \end{pmatrix}\right) = \prod_{i \in I \setminus J} \left\|x - s_i^I e\right\|^2 \prod_{j \in J \setminus I} \left\|y - s_j^J e\right\|^2 \prod_{i \in I \cap J} \Delta^2(x - s_i^I e, y - s_i^J e).$$

Indeed, for each $i \in I \setminus J$, the entry $a_i^\top(x - s_i^I e)$ is independent of the others and has variance $\left\|x - s_i^I e\right\|^2$. The same holds for $j \in J \setminus I$. For each $\ell \in I \cap J$, the pair of Gaussian variables $a_\ell^\top(x - s_\ell^I e)$ and $a_\ell^\top(y - s_\ell^J e)$ has mean zero and covariance matrix

$$\begin{pmatrix} \left\|x - s_\ell^I e\right\|^2 & (x - s_\ell^I e)^\top (y - s_\ell^J e) \\ (x - s_\ell^I e)^\top (y - s_\ell^J e) & \left\|y - s_\ell^J e\right\|^2 \end{pmatrix},$$

whose determinant is $\Delta^2(x - s_\ell^I e, y - s_\ell^J e)$. Therefore

$$\begin{aligned} p_{X_I(x), X_J(y)}(0, 0) &= (2\pi)^{-d} \left(\det\left(\mathrm{Cov}\begin{pmatrix} X_I(x) \\ X_J(y) \end{pmatrix}\right)\right)^{-1/2} \\ &= (2\pi)^{-d} \prod_{i \in I \setminus J} \left\|x - s_i^I e\right\|^{-1} \prod_{j \in J \setminus I} \left\|y - s_j^J e\right\|^{-1} \prod_{i \in I \cap J} \Delta^{-1}(x - s_i^I e, y - s_i^J e). \end{aligned} \tag{D.5}$$

When n is large enough so that $n^{-1/2} \leq \tau/2$, for any $x, y \in T_n$ and any sign $\sigma \in \{-1, 1\}$, we have $\|x - \sigma e\|, \|y - \sigma e\| \geq \tau/2$, so there exists some constant $0 < C_d < \infty$ such that

$$\prod_{i \in I \setminus J} \left\|x - s_i^I e\right\|^{-1} \prod_{j \in J \setminus I} \left\|y - s_j^J e\right\|^{-1} \leq C_d. \tag{D.6}$$

For $x \in T_n$, we can write its tube parametrization

$$x = \alpha_x e + (\tau + \beta_x) u_x \qquad \text{with } u_x \in \mathbb{S}^{d-2}, \quad \text{and } |\alpha_x|, |\beta_x| \leq \frac{\tau}{2}.$$

Then for any $\ell \in I \cap J$, a direct expansion implies

$$\begin{aligned}
\Delta^2(x - s_\ell^I e, y - s_\ell^J e) &= (\tau + \beta_x)^2(\tau + \beta_y)^2 \left[1 - (u_x^\top u_y)^2\right] \\
&\quad + \left\|(\alpha_x - s_\ell^I)(\tau + \beta_y)u_y - (\alpha_y - s_\ell^J)(\tau + \beta_x)u_x\right\|^2 \\
&\geq (\tau + \beta_x)^2(\tau + \beta_y)^2 \left[1 - (u_x^\top u_y)^2\right] \\
&\geq \frac{\tau^4}{16}\left[1 - (u_x^\top u_y)^2\right] \\
&= \frac{\tau^4}{16}\left(1 - \left(1 - \frac{\delta(x,y)^2}{2}\right)^2\right) \qquad \text{(D.7)} \\
&\geq \frac{\tau^4}{32}\delta(x,y)^2, \qquad \text{(D.8)}
\end{aligned}$$

where the equality (D.7) follows from

$$\delta(x,y)^2 = \min\{\|u_x - u_y\|^2, \|u_x + u_y\|^2\} = \min\{2 - 2u_x^\top u_y, 2 + 2u_x^\top u_y\} = 2(1 - \left|u_x^\top u_y\right|).$$

Substituting (D.6) and (D.8) into the density formula (D.5) proves statement (I).

A second estimate of $\Delta^2(x - s_\ell^I e, y - s_\ell^J e)$ yields (D.3). For matching shared signs $s_i^I = s_i^J = \sigma$, we have

$$\Delta(x - \sigma e, y - \sigma e) = \Delta(x - \sigma e, y - x) = \|x - \sigma e\| \left\|P_{(x-\sigma e)^\perp}(y - x)\right\|. \qquad \text{(D.9)}$$

For opposite shared signs, first consider the corresponding points on the saddle sphere $\mathcal{S}$. With $t = u_x^\top u_y > 0$, a direct expansion implies

$$\Delta^2(\tau u_x - \sigma e, \tau u_y + \sigma e) = (1 + \tau^2)^2 - (\tau^2 t - 1)^2 = \tau^2(1 + t)\big[2 + \tau^2(1 - t)\big] \geq 2\tau^2.$$

Since $\|x - \tau u_x\| + \|y - \tau u_y\| \leq C_d n^{-1/2}$ for some constant C_d, by uniform continuity of $(x,y) \mapsto \Delta^2(x - \sigma e, y + \sigma e)$ near $\mathcal{S}$, there exists n_0, depending only on d, such that when $n \geq n_0$,

$$\Delta^2(x - \sigma e, y + \sigma e) \geq \tau^2. \qquad \text{(D.10)}$$

Thus opposite-sign factors are uniformly bounded away from zero. Substituting (D.6), (D.9), and (D.10) into the density formula (D.5) proves statement (II).

To prove (III), fix $x \in T_n$ and recall that $v_\pm = x \mp e$. For $\omega \in \mathbb{S}^{d-1}$, the formula for orthogonal projection implies

$$\left\|P_{v_\pm^\perp}\omega\right\|^2 = \left\|\omega - \frac{v_\pm^\top \omega}{\|v_\pm\|^2}v_\pm\right\|^2 = 1 - \frac{(v_\pm^\top \omega)^2}{\|v_\pm\|^2} = \sin^2\theta_\pm(\omega),$$

where

$$\theta_\pm(\omega) = \arccos\left(\frac{v_\pm^\top \omega}{\|v_\pm\|}\right) \in [0, \pi]$$

is the angle between ω and $v_\pm$. Since sine is nonnegative on $[0,\pi]$, we have

$$\left\|P_{v_\pm^\perp}\omega\right\| = \sin\theta_\pm(\omega).$$

Thus a denominator factor becomes small only when its angle is close to 0 or π. We divide the integral according to which of these two factors is small.

First, the directions of v_+ and v_- remain separated uniformly in x. Indeed, using $x = \alpha_x e + (\tau + \beta_x)u_x$, we obtain

$$\Delta(v_+, v_-)^2 = 4\big(\|x\|^2 - (x^\top e)^2\big) = 4(\tau + \beta_x)^2.$$

Since $\tau + \beta_x \geq \tau/2$ for sufficiently large n, and the norms $\|x \mp e\|$ are uniformly bounded on T_n, there is a fixed $c_0 \in (0,1)$, depending only on d, such that

$$\frac{\Delta(v_+, v_-)}{\|v_+\| \, \|v_-\|} = \frac{2(\tau + \beta_x)}{\|x - e\| \, \|x + e\|} \geq c_0.$$

The left side is the sine of the angle between the two vectors. Hence the angle between either signed direction of v_+ and either signed direction of v_- is at least $\arcsin c_0$.

Choose a fixed $0 < \theta_0 < \frac{1}{3} \arcsin c_0$ and define

$$\begin{aligned} E_+ &= \{\omega \in \mathbb{S}^{d-1} : \theta_+(\omega) < \theta_0 \text{ or } \theta_+(\omega) > \pi - \theta_0\}, \\ E_- &= \{\omega \in \mathbb{S}^{d-1} : \theta_-(\omega) < \theta_0 \text{ or } \theta_-(\omega) > \pi - \theta_0\}, \quad \text{and} \\ E_0 &= \mathbb{S}^{d-1} \setminus (E_+ \cup E_-). \end{aligned}$$

The separation bound implies $E_+ \cap E_- = \varnothing$. Thus these three regions partition the sphere: E_+ contains directions close to the line spanned by v_+, E_- contains directions close to the line spanned by v_-, and E_0 contains the remaining directions. Consequently,

$$\int_{\mathbb{S}^{d-1}} \frac{\mathrm{d}\sigma(\omega)}{\left\|P_{v_+^\perp}\omega\right\|^{r_+} \left\|P_{v_-^\perp}\omega\right\|^{r_-}} = \left(\int_{E_+} + \int_{E_-} + \int_{E_0}\right) \frac{\mathrm{d}\sigma(\omega)}{(\sin\theta_+(\omega))^{r_+}(\sin\theta_-(\omega))^{r_-}}.$$

On E_+, disjointness from E_- implies $\theta_-(\omega) \in [\theta_0, \pi - \theta_0]$, so $\sin\theta_-(\omega) \geq \sin\theta_0$. To integrate the other factor, use spherical coordinates around v_+:

$$\omega = \cos\theta \, \frac{v_+}{\|v_+\|} + \sin\theta \, w \qquad \text{with } 0 < \theta < \pi, \quad \text{and } w \in \mathbb{S}^{d-1} \cap v_+^\perp.$$

Here $\theta = \theta_+(\omega)$, the sphere containing w is identified with $\mathbb{S}^{d-2}$, and

$$\mathrm{d}\sigma(\omega) = (\sin\theta)^{d-2} \mathrm{d}\theta \, \mathrm{d}\sigma(w).$$

The region E_+ corresponds to $\theta \in [0, \theta_0) \cup (\pi - \theta_0, \pi]$. Since $\sin(\pi - \theta) = \sin\theta$, the two angular intervals contribute equally. Therefore,

$$\begin{aligned} \int_{E_+} \frac{\mathrm{d}\sigma(\omega)}{(\sin\theta_+(\omega))^{r_+}(\sin\theta_-(\omega))^{r_-}} &\leq (\sin\theta_0)^{-r_-} \int_{E_+} (\sin\theta_+(\omega))^{-r_+} \mathrm{d}\sigma(\omega) \\ &= 2(\sin\theta_0)^{-r_-} \int_0^{\theta_0} \int_{\mathbb{S}^{d-2}} (\sin\theta)^{d-2-r_+} \mathrm{d}\sigma(w) \, \mathrm{d}\theta \\ &= 2(\sin\theta_0)^{-r_-} \left|\mathbb{S}^{d-2}\right| \int_0^{\theta_0} (\sin\theta)^{d-2-r_+} \mathrm{d}\theta \leq C_d. \end{aligned}$$

The last inequality uses $d - 2 - r_+ \geq 0$ and $r_- \leq d - 2$. Interchanging $+$ and $-$ implies the same bound for the integral over E_-.

Finally, on E_0, both angles belong to $[\theta_0, \pi - \theta_0]$. Hence

$$\int_{E_0} \frac{\mathrm{d}\sigma(\omega)}{(\sin\theta_+(\omega))^{r_+}(\sin\theta_-(\omega))^{r_-}} \leq \left|\mathbb{S}^{d-1}\right| (\sin\theta_0)^{-(r_+ + r_-)} \leq C_d.$$

Adding the three bounds proves (D.4). The choice of θ_0 and all constants are independent of x and n for sufficiently large n, which proves (III). □

D.3 Kac–Rice formulas

We first state a weighted Kac–Rice formula for Gaussian random fields and then apply it to the active equations defining candidate minima.

Lemma D.3 (Weighted Kac–Rice formula [2, Theorem 6.4])**.** *Let $V \subset \mathbb{R}^p$ be open, let $Z : V \to \mathbb{R}^p$ be a Gaussian random field with almost surely C^1 sample paths, and let $\Theta \in \mathbb{R}^q$ be a random vector. Suppose the following conditions hold.*

(K1) *The random vector Θ and the field Z are jointly Gaussian, i.e., for every $n \geq 1$ and every $x_1, \dots, x_n \in V$, the vector $\big(\Theta, Z(x_1), \dots, Z(x_n)\big) \in \mathbb{R}^{q+np}$ is Gaussian.*

(K2) *For every $x \in V$, we have $\operatorname{Cov}(Z(x)) \succ 0$.*

(K3) *We have $\mathbb{P}(\text{There exists an } x \in V \text{ such that } Z(x) = 0 \text{ and } \det DZ(x) = 0) = 0$.*

Then, for every Borel set $B \subset V$ and every nonnegative Borel function $H : V \times \mathbb{R}^q \to [0, \infty]$, we have

$$\mathbb{E}\left[\sum_{\substack{x \in B \\ Z(x)=0}} H(x, \Theta)\right] = \int_B p_{Z(x)}(0)\mathbb{E}\big[|\det DZ(x)|\, H(x, \Theta) \;\big|\; Z(x) = 0\big]\, \mathrm{d}x,$$

where the conditional expectation uses the usual Gaussian conditional law.

Proof. The original theorem [2, Theorem 6.4] holds for every compact set B and bounded continuous function H under the same conditions. We first fix a compact set B; a standard monotone-class argument and the monotone convergence theorem then extend the identity to nonnegative Borel functions.

To extend the identity to a Borel set $C \subset V$, choose compact sets $K_n \subset V$ with $K_n \uparrow V$. By (K3), the Jacobian $DZ(x)$ is invertible at every zero of Z, almost surely. Since the sample paths are C^1, the inverse function theorem implies that every zero is isolated. The zero set is also closed in V by continuity, so each compact set K_n contains only finitely many zeros. Since the sets K_n exhaust V, the full zero set is at most countable. The sums in the formula are taken over these zeros. As K_n increases, more zeros are included, and the nonnegativity of H implies that, almost surely,

$$\sum_{\substack{x \in K_n \\ Z(x)=0}} 1_{x \in C} H(x, \Theta) \to \sum_{\substack{x \in C \\ Z(x)=0}} H(x, \Theta).$$

Apply the compact-set Kac–Rice identity on K_n to the nonnegative Borel function $(x, \theta) \mapsto 1_{x \in C} H(x, \theta)$. Monotone convergence allows us to pass to the limit both in the expectation of the sum and in the integral on the right-hand side, establishing the formula for C. □

For an active set $I \subset [n]$ with $|I| = d$, define

$$H_I(x; A) = 1_{\mathcal{K}_x}(A_I) 1_{-S(x;A_{I^c}) \in Q(x;A_I)(-1,1)^d},$$

using the notation from Appendix C.1. For $x \neq \pm x^\star$, let $p_{X_I(x)}$ denote the Gaussian density of $X_I(x)$; the joint density and its domain $\mathcal{R}_{I,J}$ were defined above.

Lemma D.4 (Kac–Rice formulae)**.** *Fix integers $d \geq 4$ and $n \geq d$, and a deterministic unit vector $x^\star \in \mathbb{R}^d$. Let $a_1, \dots, a_n$ be independent $N(0, I_d)$ vectors, and suppose $n^{-1/2} < \tau$. Then the following identities hold.*

(I) *For any active set I with $|I| = d$, we have*

$$\mathbb{E}[\chi_I^n] = \int_{T_n} p_{X_I(x)}(0)\mathbb{E}\big[|\det A_I|\, H_I(x; A) \;\big|\; X_I(x) = 0\big]\mathrm{d}x.$$

(II) *For any active sets $I \neq J$ with $|I| = |J| = d$, we have*

$$\mathbb{E}[\chi_I^n \chi_J^n] = \int_{T_n^2 \cap \mathcal{R}_{I,J}} p_{X_I(x), X_J(y)}(0,0)\, \mathbb{E}\big[|\det A_I|\, |\det A_J|\, H_I(x;A) H_J(y;A) \;\big|\; X_I(x) = X_J(y) = 0\big] \mathrm{d}x\, \mathrm{d}y.$$

Proof. By rotational invariance of the standard Gaussian measurements, we may assume without loss of generality that $x^\star = e$, where e is the first standard basis vector. We apply the weighted Kac–Rice formula in Lemma D.3, first to X_I and then to the joint field (X_I, X_J). In both applications, take Θ to be the vector of all entries of A. For each fixed realization of A, the maps $x \mapsto X_I(x)$ and $(x, y) \mapsto (X_I(x), X_J(y))$ are affine, so their sample paths are smooth. At fixed inputs, their values are linear in Θ, so both fields are jointly Gaussian with Θ. It is also easy to verify that H_I is a nonnegative Borel function.

For statement (I), consider X_I on the open set $V = \mathbb{R}^d \setminus \{\pm e\}$. Write $I = \{i_1 < \cdots < i_d\}$. The coordinates of $X_I(x)$ are independent Gaussian variables with mean zero, and hence

$$\mathrm{Cov}(X_I(x)) = \mathrm{diag}\big(\|x - s_1 e\|^2, \ldots, \|x - s_d e\|^2\big) \succ 0 \qquad \text{for } x \in V.$$

Moreover, $DX_I(x) = A_I$, which is almost surely nonsingular. Thus all three hypotheses of Lemma D.3 hold. The assumption $n^{-1/2} < \tau$ ensures that $T_n \subset V$. Since X_I has the unique zero x_I, we have

$$\sum_{\substack{x \in T_n \\ X_I(x) = 0}} H_I(x;A) = 1_{\{x_I \in T_n\}} H_I(x_I;A) = \chi_I^n.$$

Applying the formula with the Borel set T_n proves statement (I).

For statement (II), consider the joint field on $\mathcal{R}_{I,J}$. Since covariance matrices are positive semidefinite, $\mathcal{R}_{I,J}$ is precisely the set where the covariance determinant is strictly positive. This determinant depends continuously on (x, y), so $\mathcal{R}_{I,J}$ is open. The derivative of the joint field is

$$D_{(x,y)} \begin{pmatrix} X_I(x) \\ X_J(y) \end{pmatrix} = \begin{pmatrix} A_I & 0 \\ 0 & A_J \end{pmatrix}.$$

Both active matrices are almost surely nonsingular, so the Jacobian is nonsingular and its absolute determinant is $|\det A_I|\, |\det A_J|$. Apply Lemma D.3 with the weight $H_I(x;A) H_J(y;A)$ and the Borel set $T_n^2 \cap \mathcal{R}_{I,J}$. The right-hand side is exactly the integral in statement (II). Since the joint field has the unique zero (x_I, x_J), the left-hand side is

$$\mathbb{E}\left[\chi_I^n \chi_J^n 1_{\{(x_I, x_J) \in \mathcal{R}_{I,J}\}}\right].$$

Lemma D.1 implies $(x_I, x_J) \in \mathcal{R}_{I,J}$ almost surely. Removing this indicator proves statement (II). □

D.4 Uniform local limit theorems

We state a local limit theorem for parallelepipeds, derive a version uniform over a compact parameter set, and apply it to the probabilities appearing in the first- and second-moment formulas. The following is a direct consequence of [7, Theorem 1.I]. We state it without proof.

Lemma D.5 (Parallelepiped version of [7, Theorem 1.I])**.** *Let $Y_{n,1}, \ldots, Y_{n,n}$ be independent copies of Y_n, with $\mathbb{E}Y_n = 0$, and $\mathrm{Cov}(Y_n) = I_p$. Define $T_n = \sum_{j=1}^n Y_{n,j}$, and suppose the following holds.*

(L) *For every $\epsilon > 0$, we have $\lim_{n\to\infty} \mathbb{E}\left[\|Y_n\|^2 1_{\{\|Y_n\| > \epsilon\sqrt{n}\}}\right] = 0$.*

(R) *For every $0 < r < R < \infty$, we have $\limsup_{n\to\infty} \sup_{r \le \|\xi\| \le R} \left|\mathbb{E} e^{i\xi^\top Y_n}\right| < 1$.*

Then, for every compact set $\mathcal{B} \subset GL_p(\mathbb{R})$, we have

$$\lim_{n\to\infty} \sup_{\substack{a\in\mathbb{R}^p \\ B\in\mathcal{B}}} \left| \frac{n^{p/2}}{2^p |\det B|} \mathbb{P}\left(T_n \in a + B(-1,1)^p\right) - \varphi_{I_p}\left(\frac{a}{\sqrt{n}}\right) \right| = 0,$$

where φ_{I_p} is the density of $N(0, I_p)$.

Lemma D.6 (Uniform local limit theorem over a compact parameter set)**.** *Let $K \subset \mathbb{R}^q$ be a nonempty compact set, and let $\{W_t : t \in K\}$ be a family of random vectors in $\mathbb{R}^p$. For each $t \in K$, let $W_{t,1}, W_{t,2}, \ldots$ be independent copies of W_t, and define*

$$\mu_t = \mathbb{E}[W_t], \qquad \Sigma_t = \mathrm{Cov}(W_t) \qquad \text{and} \qquad S_{n,t} = \sum_{j=1}^{n} W_{t,j}.$$

Suppose that the following hold true.

(H1) *The family is weakly continuous, i.e., for every sequence $t_n \to t$ in K, we have $W_{t_n} \xrightarrow{d} W_t$.*

(H2) *The third moments are uniformly bounded, i.e., $\sup_{t\in K} \mathbb{E}[\|W_t\|^3] < \infty$.*

(H3) *The covariance matrices are uniformly positive definite, i.e., $\inf_{t\in K} \lambda_{\min}(\Sigma_t) > 0$.*

(H4) *For every $t \in K$ and every $\xi \in \mathbb{R}^p \setminus \{0\}$, we have $\left|\mathbb{E}[e^{i\xi^\top W_t}]\right| < 1$.*

Then, for every compact set $\mathcal{B} \subset GL_p(\mathbb{R})$, we have

$$\lim_{n\to\infty} \sup_{\substack{t\in K, a\in\mathbb{R}^p \\ B\in\mathcal{B}}} \left| \frac{n^{p/2}}{2^p |\det B|} \mathbb{P}(S_{n,t} \in a + B(-1,1)^p) - \varphi_{\Sigma_t}\left(\frac{a - n\mu_t}{\sqrt{n}}\right) \right| = 0, \tag{D.11}$$

where φ_{Σ_t} is the density of $N(0, \Sigma_t)$.

Proof. Suppose, for contradiction, that the limit (D.11) is false. Then, there exist $\epsilon > 0$ and sequences $n_k \uparrow \infty$, $t_k \in K$, $a_k \in \mathbb{R}^p$, and $B_k \in \mathcal{B}$ such that

$$\left| \frac{n_k^{p/2}}{2^p |\det B_k|} \mathbb{P}(S_{n_k,t_k} \in a_k + B_k(-1,1)^p) - \varphi_{\Sigma_{t_k}}\left(\frac{a_k - n_k\mu_{t_k}}{\sqrt{n_k}}\right) \right| \geq \epsilon. \tag{D.12}$$

Because K is compact, passing to a subsequence, we may assume $t_k \to t_\star$ for some $t_\star \in K$. Condition (H2) implies uniform integrability of both $\{W_t\}_{t\in K}$ and $\{\|W_t\|^2\}_{t\in K}$. Combining uniform integrability with (H1) yields $\mu_{t_k} \to \mu_{t_\star}$ and $\Sigma_{t_k} \to \Sigma_{t_\star}$. In particular, the map $t \mapsto \Sigma_t$ is continuous on K. Since K is compact, we have

$$\sup_{t\in K} \lambda_{\max}(\Sigma_t) < \infty.$$

Together with (H3), this yields uniform bounds on the spectrum of Σ_t. In turn, the map $t \mapsto \Sigma_t^{-1/2}$ is continuous and uniformly bounded on K.

Define the random variables

$$Y_k = \Sigma_{t_k}^{-1/2}(W_{t_k} - \mu_{t_k}), \qquad \text{and} \qquad T_k = \sum_{j=1}^{n_k} Y_{k,j}.$$

Clearly $\mathbb{E}Y_k = 0$ and $\mathrm{Cov}(Y_k) = I_p$. The uniform bound on $\Sigma_t^{-1/2}$ and condition (H2) imply $\sup_{t\in K} \mathbb{E}\left\|\Sigma_t^{-1/2}(W_t - \mu_t)\right\|^3 < \infty$. Hence, for every $\epsilon > 0$,

$$\mathbb{E}\Big[\|Y_k\|^2 1_{\{\|Y_k\| > \epsilon\sqrt{n_k}\}}\Big] \leq \frac{\sup_{t\in K} \mathbb{E}\|\Sigma_t^{-1/2}(W_t - \mu_t)\|^3}{\epsilon\sqrt{n_k}} \longrightarrow 0,$$

which verifies condition (L) at the sample sizes n_k. For (R), condition (H1) and the continuity of μ_t and Σ_t imply that

$$(t,\xi) \longmapsto \left|\mathbb{E}e^{i\xi^\top \Sigma_t^{-1/2}(W_t-\mu_t)}\right|$$

is continuous. By condition (H4) and the invertibility of Σ_t, this function is strictly less than 1 whenever $\xi \neq 0$. Therefore, for every $0 < r < R < \infty$,

$$\sup_k \sup_{r\le\|\xi\|\le R} \left|\mathbb{E}e^{i\xi^\top Y_k}\right| \le \max_{\substack{t\in K\\ r\le\|\xi\|\le R}} \left|\mathbb{E}e^{i\xi^\top \Sigma_t^{-1/2}(W_t-\mu_t)}\right| < 1.$$

These estimates hold uniformly over $t \in K$, so Lemma D.5 applies along the sample sizes n_k. Finally, the set $\mathcal{C} = \left\{\Sigma_t^{-1/2}B : t \in K, B \in \mathcal{B}\right\}$ is a compact subset of $GL_p(\mathbb{R})$. Define

$$b_k := \Sigma_{t_k}^{-1/2}(a_k - n_k\mu_{t_k}), \qquad \text{and} \qquad C_k := \Sigma_{t_k}^{-1/2}B_k \in \mathcal{C}.$$

Applying Lemma D.5 to Y_k, T_k and the compact set $\mathcal{C}$, and using the relation

$$\mathbb{P}(S_{n_k,t_k} \in a_k + B_k(-1,1)^p) = \mathbb{P}(T_k \in b_k + C_k(-1,1)^p),$$

we have

$$\lim_{k\to\infty} \left| \frac{n_k^{p/2}}{2^p|\det B_k|}\mathbb{P}(S_{n_k,t_k} \in a_k + B_k(-1,1)^p) - \varphi_{\Sigma_{t_k}}\left(\frac{a_k - n_k\mu_{t_k}}{\sqrt{n_k}}\right)\right| = 0,$$

contradicting (D.12). Hence the result follows. □

We use the notation from Appendix C.1, and write φ_Σ for the density of $N(0,\Sigma)$. For the one-point probability, let $n > d$ and $C = (g_1,\dots,g_{n-d})^\top$, where the rows are independent $N(0,I_d)$ vectors. For $x \in \mathbb{R}^d$ and a deterministic matrix $B \in \mathbb{R}^{d\times d}$, define

$$\pi_n(x;B) = \mathbb{P}_{g_1,\dots,g_{n-d}}(-S(x;C) \in Q(x;B)(-1,1)^d).$$

For the joint probability, take distinct active sets $I, J \subset [n]$ with $|I| = |J| = d$, and write $r = |I \cap J|$ and $\nu = n - 2d + r$. When $\nu \ge 1$, let $C = (g_1,\dots,g_\nu)^\top$ have independent $N(0,I_d)$ rows. For a deterministic matrix $B \in \mathbb{R}^{(2d-r)\times d}$ whose rows are indexed by $I \cup J$, write B_L for its submatrix with row labels in L. For $x, y \in \mathbb{R}^d$, define

$$\pi_{n,I,J}(x,y;B) = \mathbb{P}_{g_1,\dots,g_\nu}\left\{\begin{matrix} -S(x;C) \in Q(x;B_I)(-1,1)^d + S(x;B_{J\setminus I}),\\ -S(y;C) \in Q(y;B_J)(-1,1)^d + S(y;B_{I\setminus J})\end{matrix}\right\}.$$

We use the angular separation $\delta(x,y)$ from the proof of Theorem C.3.

Lemma D.7 (Local limit theorem)**.** *Fix an integer $d \ge 4$ and a deterministic unit vector $x^\star \in \mathbb{R}^d$. The following limits hold.*

(I) *We have*

$$\lim_{n\to\infty} \sup_{x\in T_n} \sup_{B\in\mathcal{K}_x} \left|(n-d)^{d/2}\pi_n(x;B) - 2^d\left|\det Q(x;B)\right|\varphi_{\Sigma_x}(-\sqrt{n-d}\mu(x))\right| = 0.$$

(II) *Fix $0 < \delta_0 < 1$ and an integer $0 \le r \le d-1$. For index sets $I, J \subset [n]$ with $|I| = |J| = d$ and $|I \cap J| = r$, we have*

$$\lim_{n\to\infty} \sup_{\substack{x,y\in T_n\\ \delta(x,y)\ge\delta_0}} \sup_{\substack{B:\ B_I\in\mathcal{K}_x\\ B_J\in\mathcal{K}_y}} \left|\begin{matrix}\nu^d\pi_{n,I,J}(x,y;B)\\ -2^{2d}\left|\det Q(x;B_I)\right|\left|\det Q(y;B_J)\right|\varphi_{\Sigma_{x,y}}\left(-\sqrt{\nu}\begin{pmatrix}\mu(x)\\ \mu(y)\end{pmatrix}\right)\end{matrix}\right| = 0.$$

Proof. By rotational invariance, we may assume without loss of generality that $x^\star = e$, where e is the first standard basis vector. Recall that

$$\Sigma_x = \mathrm{Cov}(Z_x(a)), \qquad \text{and} \qquad \Sigma_{x,y} = \mathrm{Cov}\begin{pmatrix} Z_x(a) \\ Z_y(a) \end{pmatrix},$$

where both components of the joint vector use the same $a \sim N(0, I_d)$. We use the uniform local limit theorem in Lemma D.6.

To prove (I), choose a fixed integer n_0 sufficiently large that $n_0^{-1/2} < \tau/2$ and the bound in Lemma D.8(II) holds on T_{n_0}, and set $K = T_{n_0}$. This compact set contains T_n for all $n \geq n_0$ and excludes $\pm e$. In Lemma D.6, we take $W_x = Z_x(a) \in \mathbb{R}^d$ for $x \in K$. For any sequence $x_j \to x$ in K, the sign in $Z_{x_j}(a)$ eventually agrees with that in $Z_x(a)$ unless $a^\top(x - e) = 0$ or $a^\top(x + e) = 0$. Since $x \neq \pm e$, the Gaussian vector a lies on one of these hyperplanes with probability zero. Thus $Z_{x_j}(a) \to Z_x(a)$ almost surely and hence in distribution, verifying condition (H1). Moreover, since K is bounded,

$$\|Z_x(a)\| \leq \|x\| \, \|a\|^2 \leq C_d \, \|a\|^2 \qquad \text{for } x \in K.$$

Since $\mathbb{E}\,\|a\|^6 < \infty$, this bound implies $\sup_{x \in K} \mathbb{E}\,\|Z_x(a)\|^3 < \infty$, verifying condition (H2). Lemma D.8(II) implies $\inf_{x \in K} \lambda_{\min}(\Sigma_x) \geq c_d > 0$, verifying condition (H3).

We next verify condition (H4). Seeking a contradiction, suppose that $|\mathbb{E} \exp(i\xi^\top Z_x(a))| = 1$ for some $\xi \neq 0$. Since $e^{i\xi^\top Z_x(a)}$ has modulus one, equality in the triangle inequality implies that it equals $e^{i\theta}$ almost surely for some fixed $\theta \in \mathbb{R}$. Choose an open ball U that does not intersect either hyperplane $a^\top(x - e) = 0$ or $a^\top(x + e) = 0$. On U, the sign in $Z_x(a)$ is fixed, so

$$\xi^\top Z_x(a) = \sigma(a^\top x)(a^\top \xi) \qquad \text{for } a \in U \text{ and some fixed } \sigma \in \{-1, 1\}.$$

The Gaussian density is positive everywhere, and $e^{i\xi^\top Z_x(a)}$ is continuous on U. Thus its almost sure equality to $e^{i\theta}$ implies equality at every point of U. Consequently,

$$\sigma(a^\top x)(a^\top \xi) \in \theta + 2\pi\mathbb{Z} \qquad \text{for all } a \in U.$$

Since this polynomial is continuous and U is connected, it cannot move between these discrete values and must be constant on U. A polynomial that is constant on an open ball is constant on all of $\mathbb{R}^d$. Evaluating at $a = 0$ shows that this constant is zero, so $(a^\top x)(a^\top \xi) = 0$ for every $a \in \mathbb{R}^d$. This is impossible when both x and ξ are nonzero: one can choose a outside the two hyperplanes $a^\top x = 0$ and $a^\top \xi = 0$. This contradiction proves (H4). Thus all four conditions of Lemma D.6 hold on K.

To apply the conclusion of Lemma D.6 to $\pi_n(x; B)$ uniformly over $B \in \mathcal{K}_x$, we need the matrices $Q(x; B)$ to lie in a fixed compact subset of $GL_d(\mathbb{R})$. Consider

$$\mathcal{B}_1 = \{Q(x; B) : x \in K,\ B \in \mathcal{K}_x\}.$$

The pairs (x, B) in this definition form a compact set, since K is compact, the rows of B are bounded, and the inequalities defining $\mathcal{K}_x$ are nonstrict and depend continuously on (x, B). The map $(x, B) \mapsto Q(x; B)$ is continuous, and $|\det Q(x; B)| \geq \lambda_d$ for $B \in \mathcal{K}_x$. Hence $\mathcal{B}_1$ is a compact subset of $GL_d(\mathbb{R})$; in particular, its determinants are uniformly bounded.

The set $Q(x; B)(-1, 1)^d$ is symmetric, so

$$\pi_n(x; B) = \mathbb{P}\big(S(x; C) \in Q(x; B)(-1, 1)^d\big).$$

Applying Lemma D.6 with translation zero, and multiplying its error by the uniformly bounded factor $2^d \, |\det Q(x; B)|$, yields

$$(n - d)^{d/2} \pi_n(x; B) = 2^d \, |\det Q(x; B)| \, \varphi_{\Sigma_x}(-\sqrt{n - d}\, \mu(x)) + o(1),$$

uniformly for $x \in K$ and $B \in \mathcal{K}_x$. Since $T_n \subset K$ for $n \geq n_0$ and $n - d \to \infty$, this proves (I).

For (II), fix δ_0 and r as in the statement. Choose $n_1 \geq n_0$ sufficiently large, as specified below,

and consider

$$K_{\delta_0} = \{(x, y) \in T_{n_1} \times T_{n_1} : \delta(x, y) \geq \delta_0\}.$$

For $x \in K$, the orthogonal projection of x onto $e^\perp$ is nonzero, so its normalization u_x depends continuously on x. Hence $\delta(x, y)$ is continuous on $K \times K$, and K_{δ_0} is a closed subset of the compact set $K \times K$. For this application of Lemma D.6, we take

$$W_{x,y} = \begin{pmatrix} Z_x(a) \\ Z_y(a) \end{pmatrix} \in \mathbb{R}^{2d} \qquad \text{for } (x, y) \in K_{\delta_0}.$$

The same convergence and domination arguments verify conditions (H1) and (H2). The set $\{(u, v) \in (\mathbb{S}^{d-2})^2 : \min\{\|u - v\|, \|u + v\|\} \geq \delta_0\}$ is compact. Lemma D.8(III) therefore allows n_1 to be chosen so that

$$\inf_{(x,y)\in K_{\delta_0}} \lambda_{\min}(\Sigma_{x,y}) \geq c_{d,\delta_0} > 0.$$

This verifies condition (H3).

To verify condition (H4), fix $(x, y) \in K_{\delta_0}$ and suppose

$$\left|\mathbb{E}e^{i[\xi^\top Z_x(a)+\eta^\top Z_y(a)]}\right| = 1$$

for some $(\xi, \eta) \neq (0, 0)$. As in (I), the random variable $e^{i[\xi^\top Z_x(a)+\eta^\top Z_y(a)]}$ must be almost surely constant. Since $\delta(x, y) > 0$, we have $u_y \neq \pm u_x$, so e, x, y are linearly independent. The linear map $a \mapsto (a^\top e, a^\top x, a^\top y)$ is therefore onto $\mathbb{R}^3$: we can prescribe these three inner products independently. In particular, we can choose a so that

$$(a^\top e, a^\top x, a^\top y) = (1, 2, 2) \qquad \text{or} \qquad (1, 2, 0).$$

The residual pair $\big((a^\top x)^2 - (a^\top e)^2, (a^\top y)^2 - (a^\top e)^2\big)$ is $(3, 3)$ for the first choice and $(3, -1)$ for the second. By continuity, the signs of its entries remain $(1, 1)$ on an open ball around the first choice of a, and $(1, -1)$ on an open ball around the second. On each ball, $\xi^\top Z_x(a) + \eta^\top Z_y(a)$ agrees with a quadratic polynomial. The argument from (I) shows that this polynomial is constant on the ball, hence constant on all of $\mathbb{R}^d$. Its value at $a = 0$ is zero, so the two sign patterns imply

$$(a^\top x)(a^\top \xi) + (a^\top y)(a^\top \eta) = 0, \quad \text{and} \quad (a^\top x)(a^\top \xi) - (a^\top y)(a^\top \eta) = 0 \qquad \text{for every } a \in \mathbb{R}^d.$$

Adding and subtracting these identities shows that each product vanishes for every $a \in \mathbb{R}^d$. Since x and y are nonzero, this forces $\xi = \eta = 0$, a contradiction. This proves (H4). Thus all four conditions of Lemma D.6 hold for $W_{x,y}$ on K_{δ_0}.

We now apply that lemma to $\pi_{n,I,J}(x, y; B)$. Recall that $\nu = n - 2d + r = n - |I \cup J|$ is the number of measurements outside the two active sets. The rows $g_1, \ldots, g_\nu$ of C are independent, so the event defining $\pi_{n,I,J}$ involves a sum of ν independent copies of $W_{x,y}$. By symmetry of the cube, we can write

$$\pi_{n,I,J}(x, y; B) = \mathbb{P}\left\{\sum_{\ell=1}^{\nu} \begin{pmatrix} Z_x(g_\ell) \\ Z_y(g_\ell) \end{pmatrix} \in -b + D(-1, 1)^{2d}\right\},$$

where

$$b = \begin{pmatrix} S(x; B_{J\setminus I}) \\ S(y; B_{I\setminus J}) \end{pmatrix}, \qquad \text{and} \qquad D = \begin{pmatrix} Q(x; B_I) & 0 \\ 0 & Q(y; B_J) \end{pmatrix}.$$

When $B_I \in \mathcal{K}_x$ and $B_J \in \mathcal{K}_y$, both diagonal blocks of D belong to $\mathcal{B}_1$. Thus D lies in a fixed compact subset of $GL_{2d}(\mathbb{R})$, as required for uniformity over the matrices in Lemma D.6. We apply that lemma in dimension $2d$, with ν summands, translation $-b$, and matrix D. Multiplying its error

by the uniformly bounded factor $2^{2d}\,|\det D|$ yields

$$\nu^d \pi_{n,I,J}(x,y;B) = 2^{2d}\,|\det Q(x;B_I)|\,|\det Q(y;B_J)| \times \varphi_{\Sigma_{x,y}}\left(-\sqrt{\nu}\begin{pmatrix}\mu(x)\\ \mu(y)\end{pmatrix} - \frac{b}{\sqrt{\nu}}\right) + o(1),$$

uniformly for $(x,y) \in K_{\delta_0}$ and matrices B satisfying $B_I \in \mathcal{K}_x$ and $B_J \in \mathcal{K}_y$.

To obtain the expression in (II), we remove the shift $b/\sqrt{\nu}$ from the Gaussian argument. Every row of B has norm at most M_d, and each sum defining b contains at most d terms. The bound $\|Z_x(a)\| \le C_d\,\|a\|^2$ therefore implies $\|b\| \le C_d$, uniformly over the index sets and matrices B satisfying $B_I \in \mathcal{K}_x$ and $B_J \in \mathcal{K}_y$. Moreover, the uniform covariance lower bound and compactness imply

$$\sup_{\substack{(x,y)\in K_{\delta_0}\\ z\in\mathbb{R}^{2d}}} \left\|\nabla \varphi_{\Sigma_{x,y}}(z)\right\| < \infty.$$

Since $\|b\|/\sqrt{\nu} \le C_d/\sqrt{\nu}$ and the determinant factors are uniformly bounded, removing $b/\sqrt{\nu}$ from the Gaussian argument changes the displayed expression by $o(1)$, uniformly. Restricting to $x, y \in T_n$ and using $\nu = n - 2d + r \to \infty$ proves (II). □

D.5 Estimates in the shrinking tube

Recall $Z_x(a)$ from (C.3) and its mean $\mu(x)$ and covariance matrices Σ_x and $\Sigma_{x,y}$ from (C.4). The following lemma gives the uniform estimates near the saddle sphere needed for the local limit approximations. Recall that we set $\kappa = 8\tau/[\pi(1+\tau^2)^2]$.

Lemma D.8 (Mean and covariance near the saddle sphere)**.** *For any $n > \max\{d, \tau^{-2}\}$, write*

$$x_{n,z,u} = \frac{z_1}{\sqrt{n}}x^\star + \left(\tau + \frac{z_2}{\sqrt{n}}\right)u, \qquad (z,u) \in [-1,1]^2 \times \mathbb{S}^{d-2}.$$

Then the following hold.

(I) *The mean satisfies*

$$\sup_{\substack{z\in[-1,1]^2\\ u\in\mathbb{S}^{d-2}}} \left\|\sqrt{n}\,\mu(x_{n,z,u}) + \kappa z_1 x^\star - \kappa z_2 u\right\| \to 0 \qquad \text{as } n \to \infty.$$

(II) *The map $x \mapsto \Sigma_x$ is continuous near $\mathcal{S}$, and for every $x \in T_n$,*

$$\Sigma_x = \|x\|^2 I_d + 2xx^\top - \mu(x)\mu(x)^\top. \tag{D.13}$$

In particular, $\Sigma_{\tau u} = \tau^2(I_d + 2uu^\top) \succ 0$ for $u \in \mathbb{S}^{d-2}$. Consequently,

$$\inf_{x\in T_n} \lambda_{\min}(\Sigma_x) \ge c_d > 0$$

for all sufficiently large n, with the constant and threshold depending only on d.

(III) *The map $(x,y) \mapsto \Sigma_{x,y}$ is continuous near $\mathcal{S}\times\mathcal{S}$, and $\Sigma_{\tau u,\tau v} \succ 0$ whenever $u, v \in \mathbb{S}^{d-2}$ and $u \ne \pm v$. Consequently, for every compact set $K \subset \{(u,v) \in (\mathbb{S}^{d-2})^2 : u \ne \pm v\}$,*

$$\inf_{\substack{z,w\in[-1,1]^2\\ (u,v)\in K}} \lambda_{\min}\left(\Sigma_{x_{n,z,u},x_{n,w,v}}\right) \ge c_{d,K} > 0$$

for all sufficiently large n, with the constant and threshold depending only on d and K.

Proof. By rotational invariance, we may assume without loss of generality that $x^\star = e$, where e is the first standard basis vector. For (I), recall that the population function is

$$F(x) = \mathbb{E}\left|(a^\top x)^2 - (a^\top e)^2\right|.$$

For each fixed x near $\mathcal{S}$, the random function $x \mapsto \left|(a^\top x)^2 - (a^\top e)^2\right|$ is differentiable at x almost surely, with gradient $2Z_x(a)$. For $\|h\| = 1$ and $0 < |t| \le 1$, its difference quotient satisfies

$$\left|\frac{\left|(a^\top(x+th))^2 - (a^\top e)^2\right| - \left|(a^\top x)^2 - (a^\top e)^2\right|}{t}\right| \le 2\left|a^\top x\right|\left|a^\top h\right| + |t|\,(a^\top h)^2$$
$$\le (2\|x\| + 1)\|a\|^2 \le C_d\|a\|^2,$$

uniformly for x in a fixed bounded neighborhood of $\mathcal{S}$. Since $\mathbb{E}[\|a\|^2] = d < \infty$, dominated convergence implies

$$\nabla F(x) = 2\mathbb{E}[Z_x(a)] = 2\mu(x).$$

Proposition 3.1 therefore implies

$$\mu(\tau u) = 0, \qquad \text{and} \qquad D\mu(\tau u) = \kappa(uu^\top - ee^\top).$$

Since $D\mu$ is continuous near the compact set $\mathcal{S}$, Taylor expansion implies

$$\mu(x_{n,z,u}) = \frac{\kappa}{\sqrt{n}}(-z_1 e + z_2 u) + o(n^{-1/2}),$$

uniformly for $z \in [-1,1]^2$ and $u \in \mathbb{S}^{d-2}$. Multiplying by $\sqrt{n}$ proves (I).

For (II), recall that

$$\Sigma_x = \mathbb{E}[Z_x(a)Z_x(a)^\top] - \mu(x)\mu(x)^\top.$$

For x near $\mathcal{S}$, the sign factor in $Z_x(a)$ squares to one almost surely, and a direct computation gives

$$\mathbb{E}[Z_x(a)Z_x(a)^\top] = \mathbb{E}[(a^\top x)^2 aa^\top] = \|x\|^2 I_d + 2xx^\top.$$

Combining the two displayed identities proves the formula for Σ_x in (D.13). Since $\mu = \nabla F/2$ is continuous near $\mathcal{S}$, this formula shows that Σ_x, and hence $\lambda_{\min}(\Sigma_x)$, is continuous near $\mathcal{S}$. In particular, setting $x = \tau u$ and using $\mu(\tau u) = 0$, we obtain

$$\Sigma_{\tau u} = \tau^2(I_d + 2uu^\top) \succeq \tau^2 I_d.$$

Fix n_0 with $n_0^{-1/2} < \tau$. Since T_{n_0} is compact and $T_n \subset T_{n_0}$ for $n \ge n_0$, uniform continuity and $\sup_{z,u}\|x_{n,z,u} - \tau u\| \le \sqrt{2}n^{-1/2} \to 0$ imply that, for all sufficiently large n,

$$\inf_{\substack{z\in[-1,1]^2\\ u\in\mathbb{S}^{d-2}}} \lambda_{\min}\left(\Sigma_{x_{n,z,u}}\right) \ge c_d := \tau^2/2.$$

For (III), choose a compact set $U \subset \mathbb{R}^d$ whose interior contains $\mathcal{S}$ and which contains neither e nor $-e$. Set $W_{x,y}(a) = \begin{pmatrix} Z_x(a) \\ Z_y(a) \end{pmatrix}$. For any sequence $(x_j, y_j) \to (x, y)$ in $U \times U$, we have $W_{x_j,y_j}(a) \to W_{x,y}(a)$ almost surely. Moreover, for $x, y \in U$,

$$\|W_{x,y}(a)\|^2 = \|Z_x(a)\|^2 + \|Z_y(a)\|^2 \le (\|x\|^2 + \|y\|^2)\|a\|^4 \le 2R^2\|a\|^4,$$

where $R = \max_{w\in U}\|w\|$. Since $\mathbb{E}[\|a\|^4] < \infty$, dominated convergence applied to each entry of the mean vector and second-moment matrix implies $\Sigma_{x_j,y_j} \to \Sigma_{x,y}$.

We next prove that $\Sigma_{\tau u,\tau v} \succ 0$ for $u, v \in \mathbb{S}^{d-2}$ with $u \ne \pm v$. Fix such u, v, and set $x = \tau u$ and $y = \tau v$. For $\xi, \eta \in \mathbb{R}^d$, suppose

$$\mathrm{Var}\left(\xi^\top Z_x(a) + \eta^\top Z_y(a)\right) = 0.$$

Then $\xi^\top Z_x(a) + \eta^\top Z_y(a)$ is almost surely constant. Since e, x, y are linearly independent, the same polynomial argument used to verify condition (H4) in the proof of Lemma D.7(II) implies $\xi = \eta = 0$. Thus $\Sigma_{\tau u,\tau v} \succ 0$.

For any compact set $K \subset \{(u, v) \in \mathbb{S}^{d-2} \times \mathbb{S}^{d-2} : u \ne \pm v\}$, the function $(u, v) \mapsto \lambda_{\min}(\Sigma_{\tau u,\tau v})$

is continuous and positive on K. Hence $\inf_{(u,v)\in K} \lambda_{\min}(\Sigma_{\tau u,\tau v}) > 0$. The same uniform-continuity argument as in (II), using $x_{n,z,u} \to \tau u$ and $x_{n,w,v} \to \tau v$ uniformly, implies that for all sufficiently large n,

$$\inf_{\substack{z,w\in[-1,1]^2 \\ (u,v)\in K}} \lambda_{\min}\big(\Sigma_{x_{n,z,u},x_{n,w,v}}\big) \geq c_{d,K} > 0.$$

□

E Three ingredients for second moment

This section contains the supporting lemmas used in the proof of Theorem C.3.

In the proofs below, $C_d, c_d > 0$ denote constants depending only on d, used for upper and lower bounds, respectively. Their values may change from line to line; additional dependencies are indicated by subscripts.

E.1 Separated region

We use the separated-region contribution $D_{r,n}^{\mathrm{sep}}$ and the truncated angular integral $\mathcal{J}_d(\delta_0)$ defined in the proof of Theorem C.3, and the constant γ_d from Theorem C.2.

Lemma E.1 (Separated region)**.** *Fix $0 < \delta_0 < 1$. As $n \to \infty$, we have $D_{0,n}^{\mathrm{sep}} = \gamma_d^2 \mathcal{J}_d(\delta_0) n^{d-2}(1 + o_{d,\delta_0}(1))$ and*

$$D_{r,n}^{\mathrm{sep}} = O_{d,\delta_0}(n^{d-2-r}) = o_{d,\delta_0}(n^{d-2}) \qquad \text{for } 1 \leq r \leq d-1.$$

Proof. The quantity $D_{r,n}^{\mathrm{sep}}$ is a sum over ordered pairs of active sets with overlap r. We first estimate the contribution of one such pair, and then sum over the pairs.

Fix $I \neq J$ with $|I| = |J| = d$ and $|I \cap J| = r$. Recall from the proof of Theorem C.3 that

$$I_{r,n}^{\mathrm{sep}} = \int_{\Omega^{\mathrm{sep}}} p_{X_I(x),X_J(y)}(0,0) \times \mathbb{E}\left[\begin{array}{c} |\det A_I|\,|\det A_J|\,1_{\mathcal{K}_x}(A_I)1_{\mathcal{K}_y}(A_J) \\ \times\, \pi_{n,I,J}(x,y;A_{I\cup J}) \end{array} \,\middle|\, X_I(x) = X_J(y) = 0 \right] \mathrm{d}x\,\mathrm{d}y.$$

We first state the bounds used in both cases. For all sufficiently large n, Lemmas D.2(I) and D.8(III) imply, uniformly over Ω^{sep},

$$p_{X_I(x),X_J(y)}(0,0) \leq C_d \delta(x,y)^{-r}, \qquad \text{and} \qquad \lambda_{\min}(\Sigma_{x,y}) \geq c_{d,\delta_0} > 0. \tag{E.1}$$

The covariance bound follows from the continuity and positive definiteness of $(x,y) \mapsto \Sigma_{x,y}$ on the compact set of limiting pairs $(\tau u_x, \tau u_y)$ with $\delta(x,y) \geq \delta_0$. In particular,

$$\sup_{v\in\mathbb{R}^{2d}} \varphi_{\Sigma_{x,y}}(v) = \frac{1}{(2\pi)^d \sqrt{\det \Sigma_{x,y}}} \leq C_{d,\delta_0}. \tag{E.2}$$

Also, the row norm bounds defining $\mathcal{K}_x$ and $\mathcal{K}_y$, together with boundedness of T_n, imply

$$|\det A_I|\,|\det A_J| \leq M_d^{2d}, \qquad \text{and} \qquad |\det Q(x;A_I)|\,|\det Q(y;A_J)| \leq C_d \tag{E.3}$$

whenever $x, y \in T_n$, $A_I \in \mathcal{K}_x$, and $A_J \in \mathcal{K}_y$.

Let $\nu = n - 2d + r = n - |I \cup J|$, the number of rows outside the two active sets. For $A_I \in \mathcal{K}_x$

and $A_J \in \mathcal{K}_y$, Lemma D.7(II) implies

$$\pi_{n,I,J}(x,y;A_{I\cup J}) = \frac{2^{2d}}{\nu^d}\,|\det Q(x;A_I)|\,|\det Q(y;A_J)|\,\varphi_{\Sigma_{x,y}}\left(-\sqrt{\nu}\begin{pmatrix}\mu(x)\\ \mu(y)\end{pmatrix}\right) + \nu^{-d} r_n(x,y;A_{I\cup J}), \tag{E.4}$$

where, for some deterministic sequence $\eta_n \to 0$,

$$\sup_{(x,y)\in\Omega^{\mathrm{sep}}} \sup_{\substack{B\in\mathbb{R}^{(2d-r)\times d}\\ B_I\in\mathcal{K}_x,\ B_J\in\mathcal{K}_y}} |r_n(x,y;B)| \le \eta_n. \tag{E.5}$$

We now use (E.4) to bound the terms with $r \ge 1$ and evaluate the leading term with $r = 0$.

Case 1: $1 \le r \le d-1$. Combining (E.2), (E.3), (E.4), and (E.5), we obtain

$$\sup_{\substack{B:B_I\in\mathcal{K}_x\\ B_J\in\mathcal{K}_y}} \pi_{n,I,J}(x,y;B) \le \nu^{-d}(C_{d,\delta_0}+\eta_n) \le C_{d,\delta_0}\nu^{-d}. \tag{E.6}$$

Substituting (E.6) into the definition of $I^{\mathrm{sep}}_{r,n}$ and using (E.1), we obtain

$$I^{\mathrm{sep}}_{r,n} \le C_{d,\delta_0}\nu^{-d}\int_{\Omega^{\mathrm{sep}}} p_{X_I(x),X_J(y)}(0,0)\mathrm{d}x\,\mathrm{d}y \le C_{d,\delta_0}\nu^{-d}\delta_0^{-r}\,\mathrm{Vol}(T_n)^2 \le C_{d,\delta_0}n^{-d-2}, \tag{E.7}$$

since $\nu/n \to 1$ and $\mathrm{Vol}(T_n) = O_d(n^{-1})$. The bound (E.7) is uniform over all pairs with overlap r. There are $\binom{n}{d}\binom{d}{r}\binom{n-d}{d-r} = O_d(n^{2d-r})$ such ordered pairs, so

$$D^{\mathrm{sep}}_{r,n} \le C_{d,\delta_0}n^{2d-r}n^{-d-2} = C_{d,\delta_0}n^{d-2-r} = o_{d,\delta_0}(n^{d-2}).$$

Case 2: $r = 0$. Now $I\cap J = \varnothing$ and $\nu = n-2d$. Write $I = \{i_1 < \cdots < i_d\}$. For fixed $x \in T_n$ and each $k = 1,\dots,d$, the k-th coordinate of $X_I(x)$ is $a_{i_k}^\top(x - s_k x^\star)$, which is Gaussian with mean zero and variance $\|x - s_k x^\star\|^2$. These coordinates are independent, so

$$p_{X_I(x)}(0) = (2\pi)^{-d/2}\prod_{k=1}^{d}\frac{1}{\|x - s_k x^\star\|}. \tag{E.8}$$

For sufficiently large n, these variances are bounded away from zero on T_n, since $\|x - s_k x^\star\| \ge \tau - n^{-1/2} \ge \tau/2$. The formula (E.8) also holds for $X_J(y)$. Since the two active sets are disjoint, $X_I(x)$ and $X_J(y)$ are independent; hence $T_n \times T_n \subset \mathcal{R}_{I,J}$ and

$$p_{X_I(x),X_J(y)}(0,0) = p_{X_I(x)}(0)p_{X_J(y)}(0). \tag{E.9}$$

Define

$$a(x) = \mathbb{E}\big[|\det A_I|\,|\det Q(x;A_I)|\,1_{\mathcal{K}_x}(A_I)\ \big|\ X_I(x) = 0\big],$$

as in Lemma C.1. This function is independent of I, since every ordered active set uses the same sign pattern. Moreover, A_I and A_J remain independent under the conditioning $X_I(x) = X_J(y) = 0$, because each constraint concerns only its own block. Consequently,

$$\mathbb{E}\left[\begin{array}{l}|\det A_I|\,|\det Q(x;A_I)|\,1_{\mathcal{K}_x}(A_I)\\ \quad\times|\det A_J|\,|\det Q(y;A_J)|\,1_{\mathcal{K}_y}(A_J)\end{array}\ \middle|\ X_I(x) = X_J(y) = 0\right] = a(x)a(y). \tag{E.10}$$

We now substitute the expansion (E.4) for $\pi_{n,I,J}(x,y;A_{I\cup J})$ in the integral defining $I^{\mathrm{sep}}_{0,n}$. For fixed x,y, the Gaussian density in that expansion is deterministic, so it can be taken outside the conditional expectation. The remaining expectation in the leading term factorizes as in (E.10).

Together with (E.9), this yields $I_{0,n}^{\mathrm{sep}} = I_1 + I_2$, where

$$I_1 = \frac{2^{2d}}{\nu^d} \int_{\Omega^{\mathrm{sep}}} p_{X_I(x)}(0) p_{X_J(y)}(0) a(x) a(y) \varphi_{\Sigma_{x,y}} \left(-\sqrt{\nu} \begin{pmatrix} \mu(x) \\ \mu(y) \end{pmatrix} \right) \mathrm{d}x \, \mathrm{d}y,$$

and I_2 contains the remainder r_n. By (E.5), (E.8), and (E.3),

$$|I_2| \le C_d \eta_n \nu^{-d} \operatorname{Vol}(T_n)^2 = o_{d,\delta_0}(n^{-d-2}). \tag{E.11}$$

It remains to evaluate I_1.

Define the parametrization

$$x_n(z,u) = \frac{z_1}{\sqrt{n}} x^\star + \left(\tau + \frac{z_2}{\sqrt{n}} \right) u \qquad \text{for } z \in [-1,1]^2 \text{ and } u \in \mathbb{S}^{d-2}.$$

We use this map for both integration variables, writing $x = x_n(z,u)$ and $y = x_n(w,v)$, where $z, w \in [-1,1]^2$ and $u, v \in \mathbb{S}^{d-2}$. For each of x and y, the two normal coordinates each contribute a factor $n^{-1/2}$, and polar coordinates in $(x^\star)^\perp$ contribute the radial factor $(\tau + z_2/\sqrt{n})^{d-2}$. Consequently,

$$\mathrm{d}x \, \mathrm{d}y = n^{-2} \left(\tau + \frac{z_2}{\sqrt{n}} \right)^{d-2} \left(\tau + \frac{w_2}{\sqrt{n}} \right)^{d-2} \mathrm{d}z \, \mathrm{d}w \, \mathrm{d}\sigma(u) \, \mathrm{d}\sigma(v).$$

The restriction $\delta(x,y) \ge \delta_0$ becomes $\min\{\|u - v\|, \|u + v\|\} \ge \delta_0$, so the transformed integration domain is compact and independent of n.

We now let $n \to \infty$. Since $\|\tau u - s_k x^\star\|^2 = 1 + \tau^2$, equation (E.8) and Lemma C.1 imply

$$\begin{aligned} p_{X_I(x_n(z,u))}(0) &\to [2\pi(1+\tau^2)]^{-d/2}, \qquad a(x_n(z,u)) \to \vartheta_d, \\ p_{X_J(x_n(w,v))}(0) &\to [2\pi(1+\tau^2)]^{-d/2}, \qquad \text{and} \qquad a(x_n(w,v)) \to \vartheta_d. \end{aligned} \tag{E.12}$$

Also, Lemma D.8(I) and (III), together with $\nu/(n-d) \to 1$, implies

$$\begin{aligned} \sqrt{\nu}\, \mu(x_n(z,u)) &\to -\kappa z_1 x^\star + \kappa z_2 u, \\ \sqrt{\nu}\, \mu(x_n(w,v)) &\to -\kappa w_1 x^\star + \kappa w_2 v, \quad \text{and} \\ \left\| \Sigma_{x_n(z,u), x_n(w,v)} - \Sigma_{\tau u, \tau v} \right\|_{\mathrm{op}} &\to 0. \end{aligned} \tag{E.13}$$

The limits in (E.12) and (E.13) are uniform over the transformed integration domain. By (E.1), the covariance matrices stay uniformly positive definite, so the Gaussian densities also converge uniformly. We may therefore pass to the limit in the integral. Using $\varphi_\Sigma(-v) = \varphi_\Sigma(v)$, we obtain

$$n^2 \nu^d I_1 \longrightarrow \frac{2^{2d} \tau^{2d-4} \vartheta_d^2}{[2\pi(1+\tau^2)]^d} \mathcal{J}_d(\delta_0) = (d!)^2 \gamma_d^2 \mathcal{J}_d(\delta_0). \tag{E.14}$$

Combining (E.14), (E.11), and $\nu/n \to 1$, we obtain

$$I_{0,n}^{\mathrm{sep}} = (d!)^2 \gamma_d^2 \mathcal{J}_d(\delta_0) n^{-d-2} (1 + o_{d,\delta_0}(1)). \tag{E.15}$$

Relabeling the Gaussian rows shows that all disjoint pairs have the same integral. Multiplying (E.15) by their number, $\binom{n}{d}\binom{n-d}{d} = n^{2d}(1 + O_d(n^{-1}))/(d!)^2$, proves the asserted formula for $D_{0,n}^{\mathrm{sep}}$. □

E.2 Tangential region

We continue to use the notation $\Omega^{\tan}$ and $D_{r,n}^{\tan}$ from the proof of Theorem C.3.

Lemma E.2 (Tangential region)**.** *Fix $d \ge 4$ and $0 < \delta_0 < 1$. For all sufficiently large n, we have*

$$D_{0,n}^{\tan} \le C_d n^{d-2} \delta_0^{d/2-1} + C_d n^{d/2-1}.$$

For $1 \le r \le d-1$,

$$D_{r,n}^{\tan} = o_d(n^{d-2}).$$

The constants C_d are independent of δ_0.

Proof. Since there are a total of $\binom{n}{d}\binom{d}{r}\binom{n-d}{d-r} = O_d(n^{2d-r})$ ordered pairs (I,J) with overlap r, it suffices to show that, for any fixed $I \neq J$ with $|I| = |J| = d$ and $|I \cap J| = r$, the integral

$$I_{r,n}^{\tan} := \int_{\Omega^{\tan}} p_{X_I(x),X_J(y)}(0,0) \times \mathbb{E}\left[\begin{array}{l} |\det A_I|\,|\det A_J|\,1_{\mathcal{K}_x}(A_I)1_{\mathcal{K}_y}(A_J) \\ \quad \times \pi_{n,I,J}(x,y;A_{I\cup J}) \end{array} \middle| X_I(x) = X_J(y) = 0 \right] \mathrm{d}x\,\mathrm{d}y,$$

satisfies

$$I_{0,n}^{\tan} \le C_d n^{-3d/2-1} + C_d n^{-d-2}\delta_0^{d/2-1}, \tag{E.16}$$

and

$$I_{r,n}^{\tan} = o_d(n^{-d+r-2}) \qquad \text{for } 1 \le r \le d-1. \tag{E.17}$$

Since the set $\{\delta(x,y) = 0\}$ has Lebesgue measure zero, we first estimate the integral over

$$\Omega_+^{\tan} = \Omega^{\tan} \cap \{(x,y) : u_x^\top u_y > 0,\ \delta(x,y) > 0\}.$$

For sufficiently large n, we have $n^{-1} < \delta_0$. We split at $\delta(x,y) = n^{-1}$, where the prefactor $n^{-d}\delta(x,y)^{-d/2}$ in the joint probability estimate equals the marginal scale $n^{-d/2}$. For larger separations, we use the joint estimate and its Gaussian decay in the normal displacement; for smaller separations, we use the marginal estimate and the small volume imposed by $\rho(x,y) \le \delta(x,y)$. Define

$$\Omega_+^{(1)} = \Omega_+^{\tan} \cap \{(x,y) : n^{-1} \le \delta(x,y) \le \delta_0\} \quad \text{and} \quad \Omega_+^{(2)} = \Omega_+^{\tan} \cap \{(x,y) : 0 < \delta(x,y) < n^{-1}\},$$

and denote by I_1 and I_2 the integrals over $\Omega_+^{(1)}$ and $\Omega_+^{(2)}$, respectively.

Separations $\delta(x,y) \ge n^{-1}$. We claim that, uniformly on $\Omega_+^{\tan}$,

$$\sup_{\substack{B:B_I\in\mathcal{K}_x \\ B_J\in\mathcal{K}_y}} \pi_{n,I,J}(x,y;B) \le C_d n^{-d}\delta(x,y)^{-d/2} \exp\left[-c_d n\left(\alpha_x^2 + \beta_x^2 + \frac{\rho(x,y)^2}{\delta(x,y)}\right)\right]. \tag{E.18}$$

We defer the proof of (E.18) to Lemma E.4. By (E.18), Lemma D.2(I), and (E.3), we obtain

$$I_1 \le C_d n^{-d} \int_{\substack{(x,y)\in\Omega_+^{\tan} \\ n^{-1}\le\delta(x,y)<\delta_0}} \delta(x,y)^{-r-d/2} \exp\left[-c_d n\left(\alpha_x^2 + \beta_x^2 + \frac{\rho(x,y)^2}{\delta(x,y)}\right)\right] \mathrm{d}x\,\mathrm{d}y. \tag{E.19}$$

We use tube coordinates, in which

$$\mathrm{d}x\,\mathrm{d}y = (\tau+\beta_x)^{d-2}(\tau+\beta_y)^{d-2}\mathrm{d}\alpha_x\,\mathrm{d}\beta_x\,\mathrm{d}\alpha_y\,\mathrm{d}\beta_y\,\mathrm{d}\sigma(u_x)\,\mathrm{d}\sigma(u_y).$$

Since $\delta(x,y)$ depends only on u_x and u_y, (E.19) implies

$$I_1 \le C_d n^{-d} \int_{\Omega_u} \delta^{-r-d/2} \underbrace{\int_{\Omega_{\alpha,\beta}(\delta)} e^{-c_d n(\alpha_x^2+\beta_x^2+\rho^2/\delta)} \prod_{q\in\{x,y\}} \left[(\tau+\beta_q)^{d-2}\,\mathrm{d}\alpha_q\,\mathrm{d}\beta_q\right]}_{I_3} \mathrm{d}\sigma(u_x)\,\mathrm{d}\sigma(u_y),$$

where $\delta = \delta(x,y)$, $\rho = \rho(x,y)$, and, with $u_x, u_y \in \mathbb{S}^{d-2}$,

$$\Omega_u = \{(u_x,u_y) : u_x^\top u_y > 0,\ n^{-1} \le \delta < \delta_0\}, \quad \text{and} \quad \Omega_{\alpha,\beta}(\delta) = [-n^{-1/2}, n^{-1/2}]^4 \cap \{\rho \le \delta\}.$$

The change of variables $\zeta = \alpha_x - \alpha_y$ and $\eta = \beta_x - \beta_y$, under which $\rho(x,y)^2 = \zeta^2 + \eta^2$, has Jacobian of absolute value one. The tube Jacobian factors are bounded, so after enlarging $\Omega_{\alpha,\beta}(\delta)$, we obtain

$$I_3 \le C_d \int_{\mathbb{R}^2} e^{-c_d n(\alpha_x^2+\beta_x^2)}\mathrm{d}\alpha_x\,\mathrm{d}\beta_x \int_{\mathbb{R}^2} e^{-c_d n(\zeta^2+\eta^2)/\delta}\mathrm{d}\zeta\,\mathrm{d}\eta \le C_d n^{-1}\frac{\delta}{n} = C_d\delta n^{-2}. \tag{E.20}$$

Substituting (E.20) and then using Lemma F.6, we obtain

$$I_1 \le C_d n^{-d-2} \int_{\substack{u_x,u_y\in\mathbb{S}^{d-2}\\ n^{-1}\le\delta(u_x,u_y)<\delta_0}} \delta(u_x,u_y)^{1-r-d/2}\mathrm{d}\sigma(u_x)\,\mathrm{d}\sigma(u_y) \le C_d n^{-d-2}\int_{1/n}^{\delta_0} t^{d/2-2-r}\mathrm{d}t. \quad \text{(E.21)}$$

Evaluating the integral in (E.21) gives

$$I_1 \le \begin{cases} C_d n^{-d-2}\delta_0^{d/2-1-r}, & r < d/2-1,\\ C_d n^{-d-2}\log n, & r = d/2-1,\\ C_d n^{-3d/2-1+r}, & r > d/2-1. \end{cases} \quad \text{(E.22)}$$

Separations $0<\delta(x,y)<n^{-1}$. Condition on $A_{I\cup J}=B$, and let C_ν contain the $\nu=n-2d+r$ remaining independent Gaussian rows. The constraint at x becomes

$$S(x;C_\nu)\in -Q(x;B_I)(-1,1)^d - S(x;B_{J\setminus I}).$$

When $B_I\in\mathcal{K}_x$ and $B_J\in\mathcal{K}_y$, the target set on the right-hand side lies in $B(0,R_d)$ for a fixed R_d. Applying the local limit argument for $S(x;C_\nu)$ from the proof of Lemma D.7(I) to a fixed cube containing this ball and using the uniform covariance lower bound, we obtain

$$\sup_{\substack{B:B_I\in\mathcal{K}_x\\ B_J\in\mathcal{K}_y}} \pi_{n,I,J}(x,y;B) \le C_d\nu^{-d/2} \le C_d n^{-d/2}.$$

Using Lemma D.2(I) and the bounded tube Jacobian factors, we obtain

$$I_2 \le C_d n^{-d/2}\int_{\substack{u_x,u_y\in\mathbb{S}^{d-2}\\ 0<\delta(u_x,u_y)<1/n}} \delta^{-r}\underbrace{\int_{\Omega_{\alpha,\beta}(\delta)} \mathrm{d}\alpha_x\,\mathrm{d}\beta_x\,\mathrm{d}\alpha_y\,\mathrm{d}\beta_y}_{I_4}\mathrm{d}\sigma(u_x)\,\mathrm{d}\sigma(u_y).$$

The change of variables $\zeta=\alpha_x-\alpha_y$ and $\eta=\beta_x-\beta_y$ has Jacobian of absolute value one and transforms $\rho\le\delta$ into $\zeta^2+\eta^2\le\delta^2$. After enlarging the image of $\Omega_{\alpha,\beta}(\delta)$ under this change of variables, we obtain

$$I_4 \le \int_{[-n^{-1/2},n^{-1/2}]^2}\mathrm{d}\alpha_x\,\mathrm{d}\beta_x\int_{\zeta^2+\eta^2\le\delta^2}\mathrm{d}\zeta\,\mathrm{d}\eta = \frac{4\pi\delta^2}{n}.$$

Consequently, Lemma F.6 yields

$$\begin{aligned} I_2 &\le C_d n^{-d/2-1}\int_{\substack{u_x,u_y\in\mathbb{S}^{d-2}\\ 0<\delta(u_x,u_y)<1/n}} \delta(u_x,u_y)^{2-r}\mathrm{d}\sigma(u_x)\,\mathrm{d}\sigma(u_y)\\ &\le C_d n^{-d/2-1}\int_0^{1/n} t^{d-1-r}\mathrm{d}t \le C_d n^{-3d/2-1+r}, \end{aligned} \quad \text{(E.23)}$$

where the last integral is finite because $r\le d-1$.

The antipodal contribution. For pairs with $u_x^\top u_y<0$, replace y by $y'=-y$ and reverse all active signs assigned to J. Each active equation for J is then multiplied by -1. Moreover,

$$Z_{y'}(a)=-Z_y(a), \qquad Q(y';A_J)=-Q(y;A_J), \qquad \text{and} \qquad \mathcal{K}_{y'}=\mathcal{K}_y.$$

Thus the joint density at zero, the conditional law of the active rows, and the determinant factors are unchanged, while the second score constraint is multiplied by -1. The antipodal contribution is therefore the corresponding integral in the positive chart with reversed signs for J. Indeed, y' has coordinates $(-\alpha_y,\beta_y,-u_y)$ and $\mathrm{d}y'=\mathrm{d}y$. The two terms defining $\delta(x,y)$ are interchanged, and, since $u_x^\top u_y<0$, both $\rho(x,y)^2$ and $\rho(x,y')^2$ equal $(\alpha_x+\alpha_y)^2+(\beta_x-\beta_y)^2$. The estimates above are

uniform under this sign reversal. The probability bounds for $\pi_{n,I,J}$ use only the boundedness of the target sets when $B_I \in \mathcal{K}_x$ and $B_J \in \mathcal{K}_y$. The density bounds in Lemma D.2(I)–(II) follow from the same rowwise calculation for the reversed signs. For the angular bound in Lemma D.2(III), the signs on I remain fixed, so the numbers of shared rows with matching signs satisfy $r_+ \le d-2$ and $r_- \le 2 \le d-2$, even after reversing the signs on J.

Combining (E.22) and (E.23) with the reflected antipodal contribution proves (E.16) for $r=0$ and (E.17) for $1 \le r \le d-1$, uniformly over I, J. This completes the proof. □

Lemma E.3 (Distance identity)**.** *For all sufficiently large n and every $x, y \in T_n$ with $u_x^\top u_y > 0$,*

$$\|y-x\|^2 = \rho(x,y)^2 + (\tau+\beta_x)(\tau+\beta_y)\delta(x,y)^2.$$

Proof. Since $x^\star \perp u_x, u_y$ and $u_x^\top u_y > 0$, we have $\delta(x,y) = \|u_x - u_y\|$ and $\rho(x,y)^2 = (\alpha_y-\alpha_x)^2 + (\beta_y-\beta_x)^2$. Therefore

$$\begin{aligned}\|y-x\|^2 &= (\alpha_y-\alpha_x)^2 + \|(\tau+\beta_y)u_y - (\tau+\beta_x)u_x\|^2 \\ &= (\alpha_y-\alpha_x)^2 + (\beta_y-\beta_x)^2 + (\tau+\beta_x)(\tau+\beta_y)\,\|u_y-u_x\|^2 \\ &= \rho(x,y)^2 + (\tau+\beta_x)(\tau+\beta_y)\delta(x,y)^2.\end{aligned}$$

□

Lemma E.4 (Tangential-region probability estimate)**.** *Fix $d \ge 4$ and active sets $I \ne J$ with $|I| = |J| = d$. There are constants $C_d, c_d > 0$ such that, for all sufficiently large n and every $x, y \in T_n$ satisfying $u_x^\top u_y > 0$, $0 < \delta(x,y) < 1$, and $\rho(x,y) \le \delta(x,y)$,*

$$\sup_{\substack{B: B_I \in \mathcal{K}_x \\ B_J \in \mathcal{K}_y}} \pi_{n,I,J}(x,y;B) \le C_d n^{-d} \delta(x,y)^{-d/2} \exp\left[-c_d n\left(\alpha_x^2 + \beta_x^2 + \frac{\rho(x,y)^2}{\delta(x,y)}\right)\right].$$

Proof. By rotational invariance of the standard Gaussian measurements, we may assume without loss of generality that $x^\star = e$, where e is the first standard basis vector. Let $r = |I \cap J|$ and $\nu = n - 2d + r$. Put $\delta = \delta(x,y)$, $h = y - x$, and $t = \|h\|$. Lemma E.3 and $\rho(x,y) \le \delta$ imply

$$t^2 = \rho(x,y)^2 + (\tau+\beta_x)(\tau+\beta_y)\delta^2, \qquad \text{and} \qquad c_d\delta \le t \le C_d\delta.$$

Choose $0 < \delta_* < 1$, depending only on d, sufficiently small that $C_d\delta_* \le t_1$, where C_d is the constant in $t \le C_d\delta$ and t_1 is as in Lemma E.6.

First consider $0 < \delta \le \delta_*$. Then $t \le t_1$, and the tangential-region assumptions of Lemma E.6 hold, so it applies with constants depending only on d. Conditional on the active rows $A_{I\cup J} = B$, relabel the remaining independent standard Gaussian rows as $g_1, \dots, g_\nu$. The target sets for the sums at x and y are

$$\mathcal{B}_x = -Q(x; B_I)(-1,1)^d - S(x; B_{J\setminus I}), \qquad \text{and} \qquad \mathcal{B}_y = -Q(y; B_J)(-1,1)^d - S(y; B_{I\setminus J}).$$

When $B_I \in \mathcal{K}_x$ and $B_J \in \mathcal{K}_y$, both targets lie in $B(0, R_d)$ for a fixed R_d. To combine the exponential estimate in Lemma E.5(II) with the concentration estimate in Lemma E.6(II), we use two independent groups of rows. Take the first $\nu_1 = \lfloor \nu/2 \rfloor$ rows for C_1 and the rest for C_2:

$$C_1 = (g_1, \dots, g_{\nu_1})^\top, \qquad \text{and} \qquad C_2 = (g_{\nu_1+1}, \dots, g_\nu)^\top.$$

Write $\nu_2 = \nu - \nu_1$; both groups have $n/2 + O_d(1)$ rows. Define

$$U_j = S(x; C_j), \qquad \text{and} \qquad V_j = S(y; C_j) - S(x; C_j) \qquad \text{for } j = 1, 2.$$

Thus $\pi_{n,I,J}(x,y;B) = \mathbb{P}(F)$, where

$$F = \{U_1 + U_2 \in \mathcal{B}_x, \quad U_1 + U_2 + V_1 + V_2 \in \mathcal{B}_y\}.$$

To apply Lemma E.5(II) to each group, let

$$H_1 = \sum_{\ell=1}^{\nu_1} L(g_\ell), \qquad \text{and} \qquad H_2 = \sum_{\ell=\nu_1+1}^{\nu} L(g_\ell),$$

where L is defined in that lemma. Thus

$$H_j = \mu(x)^\top U_j + \frac{[\mu(y)-\mu(x)]^\top V_j}{t} \qquad \text{for } j = 1, 2.$$

Lemma E.5(I) implies

$$\|\mu(y)-\mu(x)\| = \|\mathbb{E}[Z_y(a) - Z_x(a)]\| \le C_d t.$$

Also, $\|\mu(x)\| \le C_d$ on T_n. On the event F, the bounds $\mathcal{B}_x, \mathcal{B}_y \subset B(0, R_d)$ imply $\|U_1 + U_2\| \le R_d$ and $\|V_1 + V_2\| \le 2R_d$. Hence

$$H_1 + H_2 \le R_d \|\mu(x)\| + 2R_d \frac{\|\mu(y)-\mu(x)\|}{t} \le R'_d$$

for a fixed R'_d. Thus at least one of the events $E_j = \{H_j \le R'_d/2\}$ occurs. Lemma E.5(II) implies

$$\mathbb{P}(E_j) \le e^{\lambda R'_d/2 - c_d \nu_j q} \le C_d e^{-c_d n q},$$

where $q = \|\mu(x)\|^2 + \|\mu(y)-\mu(x)\|^2/t$.

To apply Lemma E.6(II) to the other group, define

$$\mathcal{T} = \{(u, v) : u \in \mathcal{B}_x,\ u + v \in \mathcal{B}_y\} \subset B(0, \sqrt{5}R_d).$$

For $j \in \{1, 2\}$, the event E_j is determined by C_j. Since C_{3-j} is independent of C_j, Lemma E.6(II) applies uniformly to the translated set and implies

$$\begin{aligned}
\mathbb{P}(F \cap E_j) &= \mathbb{E}\big[1_{E_j} \mathbb{P}(F \mid C_j)\big] \\
&= \mathbb{E}\big[1_{E_j} \mathbb{P}((U_{3-j}, V_{3-j}) \in \mathcal{T} - (U_j, V_j) \mid C_j)\big] \\
&\le C_d \nu_{3-j}^{-d} t^{-d/2} \mathbb{P}(E_j) \le C_d n^{-d} t^{-d/2} \mathbb{P}(E_j).
\end{aligned}$$

Summing over $j = 1, 2$ proves

$$\pi_{n,I,J}(x, y; B) \le C_d n^{-d} t^{-d/2} \exp\left[-c_d n \left(\|\mu(x)\|^2 + \frac{\|\mu(y)-\mu(x)\|^2}{t}\right)\right]. \tag{E.24}$$

We express the exponent in tube coordinates. Put $\widetilde{y} = \alpha_y e + (\tau + \beta_y) u_x$, so $\|\widetilde{y} - x\| = \rho(x, y)$. By Proposition 3.1, $\mu(\tau u_x) = 0$ and

$$D\mu(\tau u_x) = \kappa(u_x u_x^\top - e e^\top), \qquad \text{and} \qquad \kappa = \frac{8\tau}{\pi(1+\tau^2)^2} > 0.$$

The map $u_x u_x^\top - e e^\top$ is an isometry on $\operatorname{span}\{e, u_x\}$. Continuity of $D\mu$ near the compact saddle sphere therefore implies, for sufficiently large n,

$$\|\mu(x)\| \ge c_d \|z_x\|, \qquad \text{and} \qquad \|\mu(\widetilde{y}) - \mu(x)\| \ge c_d \rho(x, y), \tag{E.25}$$

where $z_x = (\alpha_x, \beta_x)$ and $z_y = (\alpha_y, \beta_y)$. For every orthogonal map R fixing e, rotational invariance implies $\mu(Rz) = R\mu(z)$. For any unit vector $v \perp e, u_x$, the reflection across $v^\perp$ fixes $\widetilde{y}$, so $v^\top \mu(\widetilde{y}) = 0$. Thus $\mu(\widetilde{y}) \in \operatorname{span}\{e, u_x\}$. Choosing R with $Re = e$ and $Ru_x = u_y$ then yields

$$\mu(y) - \mu(\widetilde{y}) = (u_x^\top \mu(\widetilde{y}))(u_y - u_x).$$

Since $\mu(\tau u_x) = 0$ and $D\mu$ is uniformly bounded near $\mathcal{S}$, we have $\|\mu(\widetilde{y})\| \le C_d \|z_y\|$. Hence

$$\|\mu(y) - \mu(\widetilde{y})\| \le \|\mu(\widetilde{y})\| \, \delta \le C_d \|z_y\| \, \delta. \tag{E.26}$$

Combining (E.25) and (E.26) implies

$$\|\mu(y)-\mu(x)\|^2 \geq c_d\rho(x,y)^2 - C_d\|z_y\|^2\delta^2.$$

Since $\|z_y\|^2 \leq 2\|z_x\|^2 + 2\rho(x,y)^2$ and $t \asymp \delta$, after reducing δ_* if necessary, we obtain

$$q \geq c_d\left(\alpha_x^2+\beta_x^2+\frac{\rho(x,y)^2}{\delta}\right) \qquad \text{for } 0<\delta\leq\delta_*.$$

Together with (E.24) and $t \asymp \delta$, this lower bound proves (E.18) for $0<\delta\leq\delta_*$.

For $\delta_* \leq \delta < 1$, Lemma D.7(II) implies $\pi_{n,I,J}(x,y;B) \leq C_d n^{-d}$ uniformly for B satisfying $B_I \in \mathcal{K}_x$ and $B_J \in \mathcal{K}_y$. In this range, $n(\alpha_x^2+\beta_x^2+\rho(x,y)^2/\delta)$ is bounded by a constant depending only on d. Enlarging C_d proves (E.18) in this range. □

Lemma E.5. *Fix $d \geq 4$. There exist an integer n_0 and constants $t_0, c_d, C_d > 0$, all depending only on d, such that the following holds for every $x, y \in T_{n_0}$ with $0 < t = \|y-x\| \leq t_0$. For $a \sim N(0, I_d)$, let*

$$V = Z_y(a) - Z_x(a), \qquad \text{and} \qquad W = \begin{pmatrix} Z_x(a) \\ V \end{pmatrix}.$$

(I) *We have*

$$\mathbb{E}\|V\|^j \leq C_d t \qquad \text{for } j = 1,2,3.$$

In particular, $\|\mathbb{E}V\| \leq C_d t$.

(II) *There is a constant $\lambda > 0$, depending only on d, such that, for every $k \geq 1$, independent copies $a_1, \ldots, a_k$ of a, and every $R \in \mathbb{R}$,*

$$\mathbb{P}\left(\sum_{\ell=1}^k L(a_\ell) \leq R\right) \leq e^{\lambda R - c_d k q},$$

where

$$L(a) = \mu(x)^\top Z_x(a) + \frac{[\mu(y)-\mu(x)]^\top V}{t}, \quad \text{and} \quad q = \mathbb{E}L(a) = \|\mu(x)\|^2 + \frac{\|\mu(y)-\mu(x)\|^2}{t}.$$

Proof. By rotational invariance of the standard Gaussian measurements, we may assume without loss of generality that $x^\star = e$, where e is the first standard basis vector.

For (I), put $w = (y-x)/t$. Let $s_x(a) = \operatorname{sign}((a^\top x)^2 - (a^\top e)^2)$, and similarly define s_y. Then

$$V = t s_y(a)(a^\top w)a + [s_y(a) - s_x(a)](a^\top x)a. \tag{E.27}$$

Thus

$$\|Z_x(a)\| \leq C_d\|a\|^2, \qquad \text{and} \qquad \|V\| \leq C_d\|a\|^2\left(t + 1_{\{s_x \neq s_y\}}\right). \tag{E.28}$$

To bound the sign-change term in (E.28), for $\sigma \in \{-1,1\}$ define

$$A_{\sigma,t} = \{a : (a^\top b_\sigma)(a^\top b_\sigma + t a^\top w) < 0\}, \qquad \text{and} \qquad b_\sigma = x - \sigma e.$$

The factorization

$$(a^\top x)^2 - (a^\top e)^2 = \left(a^\top(x-e)\right)\left(a^\top(x+e)\right)$$

and $y = x + tw$ imply that, if $s_x(a) \neq s_y(a)$, at least one of these two factors changes sign. Thus $\{s_x \neq s_y\} \subseteq A_{+,t} \cup A_{-,t}$ up to Gaussian null sets.

If $a \in A_{\sigma,t}$, continuity along the segment from b_σ to $b_\sigma + tw$ implies $a^\top(b_\sigma + stw) = 0$ for some $s \in [0,1]$. Writing $a = \|a\|\omega$, we obtain $\left|\omega^\top b_\sigma\right| \leq t$. Since $\|b_\sigma\|$ is bounded below uniformly on T_{n_0}, a direct computation of the surface measure implies $\mathbb{P}(A_{\sigma,t}) \leq C_d t$. Therefore,

$$\mathbb{P}(s_x(a) \neq s_y(a)) \leq \mathbb{P}(A_{+,t}) + \mathbb{P}(A_{-,t}) \leq C_d t.$$

The events $A_{\sigma,t}$ depend only on ω, which is independent of $\|a\|$. Taking $t_0 \le 1$ and using finite Gaussian moments, we obtain from (E.28), for $j = 1, 2, 3$,

$$\mathbb{E}\,\|V\|^j \le C_d \left(t^j \mathbb{E}\,\|a\|^{2j} + \mathbb{E}\big[\|a\|^{2j}\, 1_{A_{+,t} \cup A_{-,t}}\big] \right) \le C_d t. \tag{E.29}$$

In particular, $\|\mathbb{E}V\| \le \mathbb{E}\,\|V\| \le C_d t$, which proves (I).

To prove (II), we use $e^{-s} \le 1 - s + s^2 e^{|s|}/2$ to estimate $\mathbb{E}e^{-\lambda L(a)}$, so we first bound $\mathbb{E}[L(a)^2 e^{\lambda_0 |L(a)|}]$ for some fixed $\lambda_0 > 0$. Compactness of the tube and (E.29) imply $\|\mu(x)\| + \|\mu(y) - \mu(x)\|\,/t \le C_d$. The coefficient bound implies

$$\begin{aligned} |L(a)| &\le \|\mu(x)\|\,\|Z_x(a)\| + \frac{\|\mu(y) - \mu(x)\|}{t}\,\|V\| \\ &\le \left(\|\mu(x)\| + \frac{\|\mu(y) - \mu(x)\|}{t} \right) \|W\| \le C_d\,\|W\|\,. \end{aligned}$$

For the exponential weight, the independence of ω and $\|a\|$, together with the χ^2_d distribution of $\|a\|^2$, implies

$$\mathbb{E}\left[(1 + \|a\|)^6 e^{\|a\|^2/4} 1_{A_{\sigma,t}} \right] \le C_d t. \tag{E.30}$$

Choose λ_0 small enough that $\lambda_0\,\|W\| \le \|a\|^2/4$. Then (E.28) and (E.30) imply

$$\mathbb{E}\left[\|Z_x(a)\|^2\, e^{\lambda_0 \|W\|} \right] \le C_d, \qquad \text{and} \qquad \mathbb{E}\left[\|V\|^2\, e^{\lambda_0 \|W\|} \right] \le C_d t. \tag{E.31}$$

After reducing λ_0 if necessary, (E.31) remains valid with the weight $e^{\lambda_0 |L(a)|}$. Using $|a+b|^2 \le 2(|a|^2 + |b|^2)$, we obtain

$$\begin{aligned} \mathbb{E}\left[L(a)^2 e^{\lambda_0 |L(a)|} \right] &\le 2\,\|\mu(x)\|^2\, \mathbb{E}\left[\|Z_x(a)\|^2\, e^{\lambda_0 |L(a)|} \right] \\ &\qquad + \frac{2\,\|\mu(y) - \mu(x)\|^2}{t^2} \mathbb{E}\left[\|V\|^2\, e^{\lambda_0 |L(a)|} \right] \\ &\le C_d\,\|\mu(x)\|^2 + C_d \frac{\|\mu(y) - \mu(x)\|^2}{t^2}\, t \\ &= C_d \left(\|\mu(x)\|^2 + \frac{\|\mu(y) - \mu(x)\|^2}{t} \right) = C_d q. \end{aligned}$$

Since $\mathbb{E}L(a) = q$, for a sufficiently small fixed $0 < \lambda \le \lambda_0$, we obtain

$$\mathbb{E}e^{-\lambda L(a)} \le 1 - \lambda q + C_d \lambda^2 q \le e^{-c_d q}.$$

By Markov's inequality, we have

$$\begin{aligned} \mathbb{P}\left(\sum_{\ell=1}^k L(a_\ell) \le R \right) &\le e^{\lambda R} \mathbb{E} \exp\left(-\lambda \sum_{\ell=1}^k L(a_\ell) \right) \\ &= e^{\lambda R} \prod_{\ell=1}^k \mathbb{E}e^{-\lambda L(a_\ell)} \le e^{\lambda R - c_d k q}, \end{aligned}$$

which proves (II). □

Lemma E.6 (Covariance and concentration in the tangential region)**.** *Fix $d \ge 4$, and let n_0, t_0 be as in Lemma E.5. There exists $0 < t_1 \le t_0$, depending only on d, such that, for every $x, y \in T_{n_0}$ satisfying $\|y - x\| \le t_1$ and*

$$u_x^\top u_y > 0, \qquad 0 < \delta(x,y) < 1, \qquad \textit{and} \qquad \rho(x,y) \le \delta(x,y),$$

the following conclusions hold, where $t = \|y - x\|$.

(I) *For* $a \sim N(0, I_d)$, *let* $V = Z_y(a) - Z_x(a)$ *and* $W = (Z_x(a)^\top, V^\top)^\top$. *After increasing* n_0 *if necessary, there is a constant* $c_d > 0$, *depending only on* d, *such that*

$$\mathrm{Cov}(W) \succeq c_d \begin{pmatrix} I_d & 0 \\ 0 & tI_d \end{pmatrix}. \tag{E.32}$$

(II) *Let* C_k *contain* k *independent standard Gaussian rows. After increasing* n_0 *if necessary, for every fixed* $R > 0$ *we have*

$$\sup_{z \in \mathbb{R}^{2d}} \mathbb{P}\left(\begin{pmatrix} S(x; C_k) \\ S(y; C_k) - S(x; C_k) \end{pmatrix} \in z + B(0, R) \right) \le C_{d,R} k^{-d} t^{-d/2} \qquad \text{for all } k \ge 1.$$

The constant $C_{d,R}$ *is uniform over* x, y, k *satisfying these assumptions.*

Proof. By rotational invariance of the standard Gaussian measurements, we may assume without loss of generality that $x^\star = e$, where e is the first standard basis vector.

To prove the covariance bound (E.32) in (I), it suffices to show that, for every $\xi, \eta \in \mathbb{R}^d$,

$$\mathrm{Var}\big(\xi^\top Z_x(a) + \eta^\top V\big) \ge c_d\big(\|\xi\|^2 + t\,\|\eta\|^2\big).$$

Expanding the variance, we have

$$\mathrm{Var}\big(\xi^\top Z_x(a) + \eta^\top V\big) = \xi^\top \Sigma_x \xi + \eta^\top \mathrm{Cov}(V)\eta + 2\xi^\top \mathrm{Cov}(Z_x(a), V)\eta. \tag{E.33}$$

By continuity of Σ_x near the compact saddle sphere and (D.13) with $x = \tau u$, choosing n_0 sufficiently large ensures $c_d I_d \preceq \Sigma_x \preceq C_d I_d$ uniformly on T_{n_0}. By Young's inequality, it therefore suffices to prove

$$\mathrm{Cov}(V) \succeq c_d t I_d, \qquad \text{and} \qquad \|\mathrm{Cov}(Z_x(a), V)\|_{\mathrm{op}} \le C_d t, \tag{E.34}$$

and then choose t_1 sufficiently small.

The first estimate gives a lower bound of order t for the variance of V in every unit direction. The second controls its interaction with $Z_x(a)$, allowing us to establish (I). We then combine this bound with the moment estimates of Lemma E.5 to prove the concentration bound in (II).

Increment covariance. Lemma E.5(I) implies

$$\mathrm{Var}(\zeta^\top V) = \mathbb{E}(\zeta^\top V)^2 - (\zeta^\top \mathbb{E}V)^2 \ge \mathbb{E}(\zeta^\top V)^2 - C_d t^2 \qquad \text{for } \|\zeta\| = 1.$$

Thus it suffices to find a uniform positive lower bound for $t^{-1}\mathbb{E}(\zeta^\top V)^2$ over unit vectors ζ. The contribution of order t comes from individual sign crossings; the noncrossing and simultaneous-crossing terms are of order t^2. To decompose this second moment, put $w = (y - x)/t$, which is a unit vector, let $s_x(a) = \mathrm{sign}((a^\top x)^2 - (a^\top e)^2)$, and similarly define s_y. By (E.27), on $\{s_x = s_y\}$ we have $V = t s_x(a)(a^\top w)a$, while on $\{s_x \ne s_y\}$ we have $V = -s_x(a)(2a^\top x + ta^\top w)a$.

For $\sigma \in \{-1, 1\}$, recall the events from the proof of Lemma E.5:

$$A_{\sigma,t} = \{a : (a^\top b_\sigma)(a^\top b_\sigma + ta^\top w) < 0\}, \qquad \text{and} \qquad b_\sigma = x - \sigma e.$$

The factorization

$$(a^\top x)^2 - (a^\top e)^2 = \big(a^\top (x - e)\big)\big(a^\top (x + e)\big)$$

and $y = x + tw$ imply that, if $s_x(a) \ne s_y(a)$, at least one of these two factors changes sign. Thus $\{s_x \ne s_y\} \subseteq A_{+,t} \cup A_{-,t}$ up to Gaussian null sets. If both factors change sign, their product keeps

its sign. Hence $\{s_x \neq s_y\}$ occurs if and only if exactly one of $A_{+,t}$ and $A_{-,t}$ occurs. Therefore,

$$\begin{aligned}\mathbb{E}(\zeta^\top V)^2 = t^2\mathbb{E}\left[(a^\top w)^2(a^\top\zeta)^2 1_{\{s_x=s_y\}}\right] \\ + \sum_{\sigma=\pm1}\mathbb{E}\left[(2a^\top x + ta^\top w)^2(a^\top\zeta)^2 1_{A_{\sigma,t}}\right] \\ - 2\mathbb{E}\left[(2a^\top x + ta^\top w)^2(a^\top\zeta)^2 1_{A_{+,t}\cap A_{-,t}}\right].\end{aligned} \tag{E.35}$$

The first term in (E.35) is $O_d(t^2)$ because Gaussian fourth moments are finite.

To bound the intersection term, we first check the geometry of b_+ and b_-. Write $x = \alpha e + (\tau+\beta)u$ and $H_\sigma = b_\sigma^\perp$. A direct computation shows

$$\|P_{H_\sigma}x\| = \frac{\tau+\beta}{\|b_\sigma\|}, \qquad \text{and} \qquad \|P_{H_\sigma}b_{-\sigma}\| = 2\|P_{H_\sigma}x\|. \tag{E.36}$$

Thus the normal vectors $b_+ = x - e$ and $b_- = x + e$ of H_+ and H_-, respectively, are uniformly nonparallel, their lengths are bounded above and below, and $\|P_{H_\sigma}x\|$ is bounded below.

To bound simultaneous crossings, write $a = \|a\|\omega$ and choose orthonormal vectors $v_1 = b_+/\|b_+\|$ and $v_2 = P_{H_+}b_-/\left\|P_{H_+}b_-\right\|$. On $A_{+,t}\cap A_{-,t}$, the defining inequalities of $A_{+,t}$ and $A_{-,t}$ imply

$$\left|\omega^\top v_1\right| \le C_d t, \qquad \text{and} \qquad \left|\omega^\top v_2\right| = \frac{\left|\omega^\top b_- - (b_-^\top v_1)(\omega^\top v_1)\right|}{\left\|P_{H_+}b_-\right\|} \le C_d t,$$

where the bounds on $\|b_+\|$ and $\left\|P_{H_+}b_-\right\|$ follow from (E.36). The pair $(\omega^\top v_1, \omega^\top v_2)$ has density proportional to $(1-z_1^2-z_2^2)^{(d-4)/2}$ on the unit disk. Since $d \ge 4$, this density is bounded, and hence $\mathbb{P}(A_{+,t}\cap A_{-,t}) \le C_d t^2$. Independence of ω and $\|a\|$ then implies

$$\mathbb{E}\left[(1+\|a\|)^6 e^{\|a\|^2/4} 1_{A_{+,t}\cap A_{-,t}}\right] \le C_d t^2. \tag{E.37}$$

Since $(2a^\top x + ta^\top w)^2(a^\top\zeta)^2 \le C_d\|a\|^4$, (E.37) bounds the intersection term in (E.35) by $O_d(t^2)$. Consequently,

$$\mathbb{E}(\zeta^\top V)^2 = \sum_{\sigma=\pm1}\mathbb{E}\left[(2a^\top x + ta^\top w)^2(a^\top\zeta)^2 1_{A_{\sigma,t}}\right] + O_d(t^2). \tag{E.38}$$

By (E.38), it now suffices to compute the limit of the sum divided by t and show that it is uniformly positive. For each $\sigma \in \{-1,1\}$ in this sum, let $b = b_\sigma$, put $\widehat{b} = b/\|b\|$, and write $a = z + v\widehat{b}$ with $z \perp b$. For sufficiently small t and fixed z, the crossing interval in v has endpoints 0 and

$$v_t(z) = -\frac{tz^\top w}{\|b\| + t\widehat{b}^\top w}.$$

Thus, for a continuous function F bounded by a polynomial times a Gaussian,

$$\int_{\{(a^\top b)(a^\top(b+tw))<0\}} F(a)\mathrm{d}a = \int_{b^\perp}\int_0^1 F(z+\ell v_t(z)\widehat{b})\,|v_t(z)|\,\mathrm{d}\ell\,\mathrm{d}z. \tag{E.39}$$

Here and below, $\mathrm{d}z$ denotes Lebesgue measure in orthonormal coordinates on the hyperplane. Moreover, $|v_t(z)| \le C_d t\|z\|$ and $\left\|z+\ell v_t(z)\widehat{b}\right\| \ge \|z\|$. Dividing by t and applying dominated convergence gives the leading contribution of this crossing:

$$\frac{1}{t}\int_{\{(a^\top b)(a^\top(b+tw))<0\}} F(a)\mathrm{d}a \longrightarrow \frac{1}{\|b\|}\int_{b^\perp}\left|z^\top w\right| F(z)\mathrm{d}z. \tag{E.40}$$

We also need (E.40) to hold uniformly as x, w, ζ vary. The same dominated-convergence argument applies along convergent sequences of parameters b, w, F with a common Gaussian bound, using orthonormal coordinates on $b^\perp$ that vary continuously with b. Applying (E.40) to $(2a^\top x + ta^\top w)^2(a^\top\zeta)^2\varphi_{I_d}(a)$

and using (E.38), we therefore obtain

$$\frac{1}{t}\mathbb{E}(\zeta^\top V)^2 \longrightarrow 4\sum_{\sigma=\pm1}\frac{1}{\|b_\sigma\|}\int_{H_\sigma}\left|z^\top w\right|(z^\top x)^2(z^\top\zeta)^2\varphi_{I_d}(z)\mathrm{d}z. \tag{E.41}$$

To obtain the covariance lower bound, we show that the limit in (E.41) is uniformly positive. Put $\delta=\delta(x,y)$. Lemma E.3 and $\rho(x,y)\le\delta$ imply

$$t^2=\rho(x,y)^2+(\tau+\beta_x)(\tau+\beta_y)\delta^2\le C_d\delta^2.$$

For $\sigma=\pm1$, (D.8) and (D.9), together with $\|x-\sigma e\|\le C_d$, therefore imply

$$\|P_{H_\sigma}w\|=\frac{\left\|P_{(x-\sigma e)^\perp}(y-x)\right\|}{t}\ge\frac{c_d\delta}{t}\ge c_d. \tag{E.42}$$

On each H_σ, (E.42) and (E.36) imply that the restrictions of $z\mapsto z^\top w$ and $z\mapsto z^\top x$ are nonzero linear forms. Thus the corresponding integral vanishes exactly when ζ is parallel to b_σ. Since b_+ and b_- are nonparallel, the sum of the two integrals is positive for every unit ζ.

The triples (x,ζ,w), with $x\in T_{n_0}$ and unit vectors ζ,w satisfying the fixed positive lower bound in (E.42), form a compact set. The right side of (E.41) is continuous there, by Gaussian domination in local orthonormal coordinates, so it has a positive minimum depending only on d.

The representation (E.39) implies convergence along every convergent sequence of parameters, hence uniform convergence in (E.41). The positive lower bound therefore transfers to $t^{-1}\mathbb{E}(\zeta^\top V)^2$ for sufficiently small t. Since $\|\mathbb{E}V\|\le C_dt$, after choosing $t_1\le t_0$ sufficiently small, we obtain, uniformly over unit ζ,

$$\mathrm{Var}(\zeta^\top V)=\mathbb{E}(\zeta^\top V)^2-(\zeta^\top\mathbb{E}V)^2\ge c_dt-C_dt^2\ge\frac{c_d}{2}t.$$

Lemma E.5(I) with $j=2$ implies $\mathbb{E}\|V\|^2\le C_dt$, which proves the upper covariance bound. Thus

$$c_dtI_d\preceq\mathrm{Cov}(V)\preceq C_dtI_d. \tag{E.43}$$

This proves the first estimate in (E.34).

Mixed covariance. Since $\|Z_x(a)\|\le C_d\|a\|^2$ and $\|\mu(x)\|\le C_d$, (E.28) and (E.30) imply

$$\begin{aligned}\|\mathrm{Cov}(Z_x(a),V)\|_{\mathrm{op}}&\le\mathbb{E}[\|Z_x(a)\|\,\|V\|]+\|\mu(x)\|\,\|\mathbb{E}V\|\\&\le C_d\left(t\mathbb{E}\|a\|^4+\mathbb{E}\big[\|a\|^4\,1_{A_{+,t}\cup A_{-,t}}\big]\right)+C_dt\le C_dt.\end{aligned} \tag{E.44}$$

This proves the second estimate in (E.34). Substituting (E.43) and (E.44) into (E.33) and applying Young's inequality, we obtain

$$\begin{aligned}\mathrm{Var}(\xi^\top Z_x(a)+\eta^\top V)&\ge c_d\|\xi\|^2+c_dt\|\eta\|^2-C_dt\|\xi\|\,\|\eta\|\\&\ge\frac{c_d}{2}\|\xi\|^2+(c_dt-C_dt^2)\|\eta\|^2\\&\ge c_d(\|\xi\|^2+t\|\eta\|^2),\end{aligned}$$

where the last inequality holds after reducing t_1 once more. This proves (I).

Concentration estimate. For any fixed $T>0$, Lemma F.7 with $m=2d$ and $\varepsilon=RT$, together with independence, implies

$$\sup_{z\in\mathbb{R}^{2d}}\mathbb{P}(X\in z+B(0,R))\le C_{2d,RT}R^{2d}\int_{\|(\xi,\eta)\|\le T}|\chi(\xi,\eta)|^k\,\mathrm{d}\xi\,\mathrm{d}\eta, \tag{E.45}$$

where

$$X = \begin{pmatrix} S(x;C_k) \\ S(y;C_k) - S(x;C_k) \end{pmatrix}, \qquad \text{and} \qquad \chi(\xi,\eta) = \mathbb{E}e^{i[\xi^\top Z_x(a) + \eta^\top V]}.$$

It therefore suffices to prove that, for some fixed $T > 0$ depending only on d,

$$|\chi(\xi,\eta)| \le e^{-c_d(\|\xi\|^2 + t\|\eta\|^2)} \qquad \text{when } \|\xi\|, \|\eta\| \le T, \tag{E.46}$$

since the resulting Gaussian integral in (E.45) has order $k^{-d}t^{-d/2}$.

To prove (E.46), we separate the variance from the third-order remainder. Let $Y = \xi^\top[Z_x(a) - \mu(x)] + \eta^\top[V - \mathbb{E}V]$, and let Y' be an independent copy. We use the same symmetrization idea as in [5, proof of Proposition 8.2]. Independence and the inequality $\cos s \le 1 - s^2/2 + |s|^3/6$ imply

$$|\chi(\xi,\eta)|^2 = \mathbb{E}\cos(Y - Y') \le 1 - \mathrm{Var}(Y) + \frac{1}{6}\mathbb{E}\left|Y - Y'\right|^3.$$

For the remainder, Jensen's inequality, $|a+b|^3 \le 4(|a|^3 + |b|^3)$, and (E.28) imply

$$\begin{aligned}
\mathbb{E}\left|Y - Y'\right|^3 &\le 8\mathbb{E}\left|Y\right|^3 \\
&\le 32\left\|\xi\right\|^3 \mathbb{E}\left\|Z_x(a) - \mu(x)\right\|^3 + 32\left\|\eta\right\|^3 \mathbb{E}\left\|V - \mathbb{E}V\right\|^3 \\
&\le C_d\left(\left\|\xi\right\|^3 \mathbb{E}\left\|a\right\|^6 + \left\|\eta\right\|^3 \mathbb{E}\left\|V\right\|^3\right) \\
&\le C_d(\left\|\xi\right\|^3 + t\left\|\eta\right\|^3),
\end{aligned}$$

where the last inequality uses Lemma E.5(I) with $j = 3$ and the finiteness of Gaussian sixth moments. Since part (I) implies $\mathrm{Var}(Y) \ge c_d(\|\xi\|^2 + t\|\eta\|^2)$, we obtain

$$|\chi(\xi,\eta)|^2 \le 1 - c_d(\|\xi\|^2 + t\|\eta\|^2) + C_d(\|\xi\|^3 + t\|\eta\|^3).$$

For $\|\xi\|, \|\eta\| \le T$, we have $\|\xi\|^3 + t\|\eta\|^3 \le T(\|\xi\|^2 + t\|\eta\|^2)$. Choosing $T > 0$ sufficiently small depending only on d therefore allows us to absorb the cubic terms into the quadratic terms. Taking square roots and using $1 - s \le e^{-s}$ proves (E.46). Finally, substituting (E.46) into (E.45) and enlarging the integration domain, we obtain

$$\begin{aligned}
\sup_{z\in\mathbb{R}^{2d}} \mathbb{P}(X \in z + B(0,R)) &\le C_{d,R}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} e^{-c_d k(\|\xi\|^2 + t\|\eta\|^2)}\mathrm{d}\xi\,\mathrm{d}\eta \\
&\le C_{d,R}k^{-d/2}(kt)^{-d/2} = C_{d,R}k^{-d}t^{-d/2}.
\end{aligned}$$

This proves (II). □

E.3 Normal-dominated region

We continue to use the notation Ω^{nor} and $D_{r,n}^{\mathrm{nor}}$ from the proof of Theorem C.3.

Lemma E.7 (Normal-dominated region)**.** *Fix $d \ge 4$ and $0 < \delta_0 < 1$. As $n \to \infty$, we have*

$$D_{r,n}^{\mathrm{nor}} = O_d(n^{d/2-1}) = o_d(n^{d-2}) \qquad \textit{for } 0 \le r \le d-1.$$

Proof. By rotational invariance of the standard Gaussian measurements, we may assume without loss of generality that $x^\star = e$, where e is the first standard basis vector.

There are $\binom{n}{d}\binom{d}{r}\binom{n-d}{d-r} = O_d(n^{2d-r})$ ordered pairs (I,J) with overlap r. It therefore suffices to show that, for any fixed $I \ne J$ with $|I| = |J| = d$ and $|I \cap J| = r$, the integral

$$\begin{aligned}
I_{r,n}^{\mathrm{nor}} := \int_{\Omega^{\mathrm{nor}}} & p_{X_I(x),X_J(y)}(0,0) \\
&\times \mathbb{E}\left[\begin{array}{l} |\det A_I|\,|\det A_J|\,1_{\mathcal{K}_x}(A_I)1_{\mathcal{K}_y}(A_J) \\ \quad \times \pi_{n,I,J}(x,y;A_{I\cup J}) \end{array} \,\middle|\, X_I(x) = X_J(y) = 0\right]\mathrm{d}x\,\mathrm{d}y
\end{aligned}$$

satisfies

$$I_{r,n}^{\mathrm{nor}} \le C_d n^{-3d/2-1+r} \qquad \text{for } 0 \le r \le d-1. \tag{E.47}$$

We first estimate the integral over

$$\Omega_+^{\mathrm{nor}} = \Omega^{\mathrm{nor}} \cap \{(x,y) : u_x^\top u_y > 0,\ \delta(x,y) > 0\}.$$

Denote this contribution by I_+. By Lemma E.8, for pairs in Ω_+^{nor},

$$\sup_{\substack{B: B_I \in \mathcal{K}_x \\ B_J \in \mathcal{K}_y}} \pi_{n,I,J}(x,y;B) \le C_d n^{-d/2} e^{-c_d n \|y-x\|}. \tag{E.48}$$

When $A_I \in \mathcal{K}_x$ and $A_J \in \mathcal{K}_y$, $|\det A_I|\,|\det A_J| \le M_d^{2d}$. Therefore, (E.48) implies

$$I_+ \le C_d n^{-d/2} \int_{\Omega_+^{\mathrm{nor}}} p_{X_I(x),X_J(y)}(0,0) e^{-c_d n\|y-x\|} \mathrm{d}x\,\mathrm{d}y.$$

Write $h = y - x$. Lemma E.3 implies

$$\|h\|^2 = \rho(x,y)^2 + (\tau+\beta_x)(\tau+\beta_y)\delta(x,y)^2 > 0,$$

where positivity follows from $\rho(x,y) > \delta(x,y)$ and $\tau+\beta_x, \tau+\beta_y > 0$ for sufficiently large n. By (D.3),

$$p_{X_I(x),X_J(y)}(0,0) \le \frac{C_d}{\left\|P_{v_+^\perp} h\right\|^{r_+} \left\|P_{v_-^\perp} h\right\|^{r_-}},$$

where $v_\pm = x \mp e$ and $r_\pm$ count the shared rows with matching signs ± 1. Put $q = r_+ + r_- \le r$. The chosen sign pattern implies $r_+ \le d-2$ and $r_- \le 2 \le d-2$. The disjoint case is included by taking $r_+ = r_- = 0$.

After enlarging the displacement domain to $\mathbb{R}^d \setminus \{0\}$ and changing variables $h = t\omega$, we obtain

$$I_+ \le C_d n^{-d/2} \int_{T_n} \underbrace{\int_{\mathbb{S}^{d-1}} \frac{\mathrm{d}\sigma(\omega)}{\left\|P_{v_+^\perp}\omega\right\|^{r_+} \left\|P_{v_-^\perp}\omega\right\|^{r_-}}}_{I_1} \underbrace{\int_0^\infty e^{-c_d n t} t^{d-1-q} \mathrm{d}t}_{I_2} \mathrm{d}x.$$

By Lemma D.2(III), the first integral satisfies $I_1 \le C_d$ uniformly in x. For the second integral, the change of variables $s = nt$ implies

$$I_2 = n^{-(d-q)} \int_0^\infty e^{-c_d s} s^{d-1-q} \mathrm{d}s \le C_d n^{-(d-q)},$$

since $q \le r \le d-1$. Substituting the bounds on I_1 and I_2, we obtain

$$\begin{aligned} I_+ &\le C_d n^{-d/2} n^{-(d-q)} \int_{T_n} \mathrm{d}x \\ &= C_d n^{-3d/2+q} \int_{\mathbb{S}^{d-2}} \int_{[-n^{-1/2}, n^{-1/2}]^2} (\tau+\beta_x)^{d-2} \mathrm{d}\alpha_x\,\mathrm{d}\beta_x\,\mathrm{d}\sigma(u_x) \\ &\le C_d n^{-3d/2-1+q} \le C_d n^{-3d/2-1+r}, \end{aligned} \tag{E.49}$$

where we used the change of variables $x = \alpha_x e + (\tau+\beta_x)u_x$ and $q \le r$.

The antipodal contribution. For pairs with $u_x^\top u_y < 0$, replace y by $y' = -y$ and reverse all active signs assigned to J, as at the end of the proof of Lemma E.2. This change preserves $\delta(x,y)$, $\rho(x,y)$, and Lebesgue measure, and transforms the antipodal contribution into the corresponding integral over Ω_+^{nor} with reversed signs for J. The proof of Lemma E.8 applies under this sign reversal, so (E.48) holds with y' in place of y. The density bound (D.3) and the angular bound in

Lemma D.2(III) also remain valid: the signs on I are unchanged, so $r_+ \le d-2$ and $r_- \le 2 \le d-2$. Hence the antipodal contribution satisfies (E.49) as well. Since the sets $\{\delta(x,y)=0\}$ and $\{u_x^\top u_y = 0\}$ have Lebesgue measure zero, combining the two contributions proves (E.47), uniformly over I, J, and completes the proof. □

Lemma E.8 (Probability estimate in the normal-dominated region)**.** *Fix $d \ge 4$ and active sets $I \ne J$ with $|I| = |J| = d$. There are constants $C_d, c_d > 0$ such that, for all sufficiently large n and every $x, y \in T_n$ satisfying $u_x^\top u_y > 0$ and $0 < \delta(x,y) < \rho(x,y)$,*

$$\sup_{\substack{B: B_I \in \mathcal{K}_x \\ B_J \in \mathcal{K}_y}} \pi_{n,I,J}(x,y;B) \le C_d n^{-d/2} e^{-c_d n \|y-x\|}.$$

Proof. Fix active rows B satisfying $B_I \in \mathcal{K}_x$ and $B_J \in \mathcal{K}_y$, let $r = |I \cap J|$ and $\nu = n - 2d + r$, and relabel the remaining independent standard Gaussian rows as $g_1, \dots, g_\nu$. We split the remaining rows into two independent groups to obtain the two factors in the desired bound. The joint event forces at least one group's increment to have a projection bounded above by a fixed constant in the direction $\mu(y) - \mu(x)$, yielding the exponential factor $e^{-c_d n\|y-x\|}$. Conditional on that group, a uniform concentration estimate for the other yields the factor $n^{-d/2}$. Take the first $\nu_1 = \lfloor \nu/2 \rfloor$ rows for C_1 and the rest for C_2:

$$C_1 = (g_1, \dots, g_{\nu_1})^\top, \qquad \text{and} \qquad C_2 = (g_{\nu_1+1}, \dots, g_\nu)^\top.$$

With $\nu_2 = \nu - \nu_1$, both groups have $n/2 + O_d(1)$ rows. As in the proof of Lemma E.4, the target sets for the sums at x and y are

$$\mathcal{B}_x = -Q(x; B_I)(-1,1)^d - S(x; B_{J\setminus I}), \qquad \text{and} \qquad \mathcal{B}_y = -Q(y; B_J)(-1,1)^d - S(y; B_{I\setminus J}),$$

and both lie in $B(0, R_d)$ by the row norm bounds on B. Write

$$U_j = S(x; C_j), \qquad \text{and} \qquad V_j = S(y; C_j) - S(x; C_j) \qquad \text{for } j = 1, 2.$$

Then $\pi_{n,I,J}(x,y;B) = \mathbb{P}(F)$ for the pair event

$$F = \{U_1 + U_2 \in \mathcal{B}_x, \quad U_1 + U_2 + V_1 + V_2 \in \mathcal{B}_y\}.$$

Put $h = y - x$ and $w = [\mu(y) - \mu(x)] / \|\mu(y) - \mu(x)\|$. For a single row $a \sim N(0, I_d)$, write $V = Z_y(a) - Z_x(a)$. Let λ_0 be as in Lemma E.9. For a sufficiently small fixed $0 < \lambda \le \lambda_0$,

$$\begin{aligned} \mathbb{E}e^{-\lambda w^\top V} &\le 1 - \lambda \mathbb{E}[w^\top V] + \frac{\lambda^2}{2} \mathbb{E}\left[(w^\top V)^2 e^{\lambda |w^\top V|} \right] \\ &\le 1 - \lambda \|\mu(y) - \mu(x)\| + \frac{\lambda^2}{2} \mathbb{E}\left[\|V\|^2 e^{\lambda_0 \|V\|} \right] \\ &\le 1 - \lambda c_d \|h\| + C_d \lambda^2 \|h\| \\ &\le e^{-c_d \|h\|}. \end{aligned} \tag{E.50}$$

The first inequality in (E.50) uses $e^{-t} \le 1 - t + t^2 e^{|t|}/2$; the third uses (E.52). The last inequality in (E.50) follows by choosing λ sufficiently small. On F, the difference sum $V_1 + V_2$ belongs to $\mathcal{B}_y - \mathcal{B}_x \subset B(0, 2R_d)$. Thus F implies at least one of the events $E_j = \{w^\top V_j \le R_d\}$. For $j \in \{1, 2\}$, the exponential Markov inequality implies

$$\mathbb{P}(E_j) \le e^{\lambda R_d} \left(\mathbb{E}e^{-\lambda w^\top [Z_y(a) - Z_x(a)]} \right)^{\nu_j} \le C_d e^{-c_d n \|h\|}.$$

The local limit argument for $S(x; C_j)$ in the proof of Lemma D.7(I), applied to a fixed cube containing $B(0, R_d)$, implies

$$\sup_{z \in \mathbb{R}^d} \mathbb{P}\big(S(x; C_j) \in z + B(0, R_d)\big) \le C_d \nu_j^{-d/2} \tag{E.51}$$

uniformly for $x \in T_n$ and all sufficiently large n. For $j \in \{1,2\}$, conditional on C_j, the event $F \cap E_j$ requires $U_{3-j} \in \mathcal{B}_x - U_j$. Since C_{3-j} is independent of C_j and $\mathcal{B}_x \subset B(0, R_d)$, (E.51) implies

$$\begin{aligned}\mathbb{P}(F \cap E_j) &= \mathbb{E}\big[1_{E_j}\mathbb{P}(F \mid C_j)\big] \\ &\le \mathbb{E}\big[1_{E_j}\mathbb{P}(U_{3-j} \in \mathcal{B}_x - U_j \mid C_j)\big] \\ &\le C_d \nu_{3-j}^{-d/2}\mathbb{P}(E_j) \le C_d n^{-d/2}\mathbb{P}(E_j).\end{aligned}$$

Summing over $j = 1, 2$ proves (E.48). The argument uses only the row bounds on B, so it also applies after reversing the active signs at an antipodal endpoint. □

Lemma E.9 (Moments of increments in the normal-dominated region)**.** *Fix $d \ge 4$ and let $a \sim N(0, I_d)$. There exist n_0 and constants $c_d, C_d, \lambda_0 > 0$ depending only on d such that, for every $n \ge n_0$ and $x, y \in T_n$ with $u_x^\top u_y > 0$ and $\delta(x,y) < \rho(x,y)$,*

$$\|\mu(y) - \mu(x)\| \ge c_d \|y - x\|, \quad \text{and} \quad \mathbb{E}\left[\|Z_y(a) - Z_x(a)\|^2 e^{\lambda_0 \|Z_y(a) - Z_x(a)\|}\right] \le C_d \|y - x\|. \tag{E.52}$$

Proof. By rotational invariance of the standard Gaussian measurements, we may assume without loss of generality that $x^\star = e$, where e is the first standard basis vector. Put $h = y - x$. Since $u_x^\top u_y > 0$, we have

$$\delta = \|u_x - u_y\|, \qquad \text{and} \qquad \rho^2 = (\alpha_y - \alpha_x)^2 + (\beta_y - \beta_x)^2.$$

Lemma E.3 and the condition $\delta < \rho$ of the normal-dominated region imply

$$\|h\|^2 = \rho^2 + (\tau + \beta_x)(\tau + \beta_y)\delta^2 \le C_d \rho^2 \le C_d n^{-1}.$$

The segment from x to y therefore lies within $O_d(n^{-1/2})$ of τu_x. Since $\mu = \nabla F/2$, Proposition 3.1 and continuity of $D\mu$ near $\mathcal{S}$ give, uniformly over these pairs,

$$\mu(y) - \mu(x) = \kappa(u_x u_x^\top - ee^\top)h + o(\|h\|) \qquad \text{with } \kappa = \frac{8\tau}{\pi(1+\tau^2)^2} > 0. \tag{E.53}$$

Moreover,

$$e^\top h = \alpha_y - \alpha_x, \qquad \text{and} \qquad u_x^\top h = \beta_y - \beta_x - \frac{\tau + \beta_y}{2}\delta^2.$$

Hence,

$$\begin{aligned}\left\|(u_x u_x^\top - ee^\top)h\right\| &= \sqrt{(\alpha_y - \alpha_x)^2 + \left(\beta_y - \beta_x - \frac{\tau + \beta_y}{2}\delta^2\right)^2} \\ &\ge \rho - \frac{\tau + \beta_y}{2}\delta^2 \\ &\ge \rho - C_d\rho^2 \ge \frac{\rho}{2} \ge c_d \|h\|.\end{aligned} \tag{E.54}$$

The first equality in (E.54) uses that e and u_x are orthogonal unit vectors. The bound $\rho - C_d\rho^2 \ge \rho/2$ in (E.54) holds for sufficiently large n, since $\rho \le C_d n^{-1/2}$. The expansion (E.53) and (E.54) therefore imply, for sufficiently large n,

$$\|\mu(y) - \mu(x)\| \ge \kappa \left\|(u_x u_x^\top - ee^\top)h\right\| - o(\|h\|) \ge c_d \|h\|.$$

To prove the second bound in (E.52), write $s_x(a) = \operatorname{sign}((a^\top x)^2 - (a^\top e)^2)$ and similarly for s_y. Then

$$Z_y(a) - Z_x(a) = s_y(a)(a^\top h)a + [s_y(a) - s_x(a)](a^\top x)a,$$

so

$$\|Z_y(a) - Z_x(a)\| \le C_d \|a\|^2 \left(\|h\| + 1_{\{s_x(a) \ne s_y(a)\}}\right). \tag{E.55}$$

If $s_x(a) \neq s_y(a)$, continuity of $(a^\top z)^2 - (a^\top e)^2 = a^\top(z-e)a^\top(z+e)$ along the segment from x to y implies that $a^\top(x + th - \sigma e) = 0$ for some $t \in [0,1]$ and $\sigma \in \{-1,1\}$. Writing $a = \|a\|\,\omega$, we obtain

$$\left|\omega^\top(x - \sigma e)\right| = t\left|\omega^\top h\right| \leq \|h\|\,.$$

For each fixed σ, a direct computation of the surface measure of the spherical band, together with $\|x - \sigma e\| \geq \tau/2$, shows that

$$\mathbb{P}\big(\omega \in \mathbb{S}^{d-1} : \left|\omega^\top(x - \sigma e)\right| \leq \|h\|\big) \leq C_d\,\|h\|\,.$$

Taking the union over $\sigma \in \{-1,1\}$ yields $\mathbb{P}(s_x(a) \neq s_y(a)) \leq C_d\,\|h\|$. This event depends only on ω, which is independent of $\|a\|$. Since $\|a\|^2$ has a χ^2_d distribution, for $0 < \lambda_0 < 1/(2C_d)$ we have

$$\mathbb{E}\big[\|a\|^4\, e^{C_d\lambda_0\|a\|^2}\big] = d(d+2)(1 - 2C_d\lambda_0)^{-d/2-2} < \infty.$$

Thus, for sufficiently large n, (E.55) implies

$$\mathbb{E}\left[\|Z_y(a) - Z_x(a)\|^2\, e^{\lambda_0\|Z_y(a) - Z_x(a)\|}\right] \leq C_d(\|h\|^2 + \|h\|) \leq C_d\,\|h\|\,.$$

□

F Auxiliary lemmas

We collect auxiliary lemmas used throughout the paper. The proof of the first lemma is simple, so we omit it.

Lemma F.1. *Let $G \in \mathbb{R}^d$ be a nondegenerate Gaussian vector, and let $p : \mathbb{R}^d \to \mathbb{R}$ be a polynomial that is not identically zero. Then, $\mathbb{P}(p(G) = 0) = 0$.*

Lemma F.2 (Almost sure nonsingularity of projected matrices)**.** *Suppose $n > d$, and let $(a_i)_{i=1}^n$ be i.i.d. standard Gaussian vectors in $\mathbb{R}^d$. Then, almost surely, for every index set $I \subset [n]$ with $|I| < d$, every sign pattern $\sigma \in \{-1,1\}^{I^c}$, and every matrix V whose columns form an orthonormal basis of the orthogonal complement of* $\operatorname{span}\{a_i\}_{i\in I}$*,*

$$\textit{the matrix} \quad V^\top\left(\sum_{i\notin I}\sigma_i a_i a_i^\top\right)V \quad \textit{is nonsingular.}$$

Proof. Fix an index set $I \subset [n]$ with $|I| < d$ and a sign pattern σ. Almost surely, the orthogonal complement of $\operatorname{span}\{a_i\}_{i\in I}$ equals that of $\operatorname{span}\{u_i\}_{i\in I}$, where $u_i = a_i/\|a_i\|$. We can therefore choose an orthonormal basis matrix $V = (v_1, \ldots, v_{d-|I|})$ as a measurable function of the directions $(u_i)_{i\in I}$ alone. Define $H = \sum_{i\notin I}\sigma_i a_i a_i^\top$. Since any orthonormal basis can be written as $\tilde{V} = VQ$ with $Q^\top Q = I$, the singularity of $V^\top HV$ does not depend on the choice of V.

The vectors $(a_i)_{i\notin I}$ are independent of the directions $(u_i)_{i\in I}$. Conditional on these directions, $\det(V^\top HV)$ is a polynomial in the components of $(a_i)_{i\notin I}$. This polynomial is not identically zero. Indeed, choose distinct inactive indices $j_1, \ldots, j_{d-|I|}$ from I^c and set

$$a_{j_l} = v_l, \qquad \text{and} \qquad a_j = 0 \quad \text{for all remaining indices in } I^c.$$

For this choice, $V^\top HV = \operatorname{diag}(\sigma_{j_1}, \ldots, \sigma_{j_{d-|I|}})$, so its determinant is nonzero. Lemma F.1 gives

$$\mathbb{P}(\det(V^\top HV) = 0 \mid (u_i)_{i\in I}) = 0 \quad \text{almost surely.}$$

Taking expectations and then a finite union over all choices of I and σ completes the proof. □

Lemma F.3 ([24, Proposition 7.2]). *Let f be closed and weakly convex, and let $\mathcal{M}$ be an active manifold at a critical point $\bar{x}$. Then $\bar{x}$ is a (strict) local minimizer of $f|_{\mathcal{M}}$ if, and only if, it is a (strict) local minimizer of f.*

Lemma F.4. *Let $\tau > 0$ be the unique solution of $\arctan\tau + \tau/(1+\tau^2) = \pi/4$. Let $P \sim N(0,1)$ be independent of jointly standard Gaussian variables X, Y with $\mathbb{E}[XY] = c \in [-1,1]$, and define*

$$q(c) = \mathbb{E}\big[XY\mathrm{sign}(\tau^2X^2 - P^2)\mathrm{sign}(\tau^2Y^2 - P^2)\big].$$

Then there exists a constant $\gamma > 0$ such that

$$1 - q(c)^2 \geq (1-c^2)(1+\gamma c^2) \qquad \text{for } -1 \leq c \leq 1.$$

Proof. For $p \in \mathbb{R}$, put $f_p(t) = t\mathrm{sign}(\tau^2t^2 - p^2)$. This function is odd and has unit norm in $L^2(N(0,1))$. Expanding f_p in the orthonormal Hermite basis, applying the Gaussian correlation identity, and averaging over P yields

$$q(c) = \sum_{j\geq 0} w_j c^{2j+1} \qquad \text{with } w_j \geq 0, \qquad \text{and} \qquad \sum_{j\geq 0} w_j = 1.$$

Here w_j is the expectation of the squared Hermite coefficient of degree $2j+1$. In particular,

$$w_0 = \mathbb{E}_P\Big[\big(\mathbb{E}_X[X^2\mathrm{sign}(\tau^2X^2 - P^2)]\big)^2\Big] < 1,$$

because the inner expectation lies strictly between -1 and 1 whenever $0 < |P| < \infty$. Hence

$$|q(c)| \leq w_0|c| + (1-w_0)|c|^3.$$

Since $[w_0 + (1-w_0)c^2]^2 \leq w_0^2 + (1-w_0^2)c^2$, the claim follows with $\gamma = 1 - w_0^2 > 0$. □

For the next lemma, let $\tau > 0$ and let X, Z, P be independent standard Gaussian variables. For $-1 < c < 1$, set

$$s = \sqrt{1-c^2}, \qquad Y_c = cX + sZ, \qquad \text{and} \qquad T_c = -sX + cZ,$$

and define

$$r(c) = \mathbb{E}\big[XZY_cT_c\mathrm{sign}(\tau^2X^2 - P^2)\mathrm{sign}(\tau^2Y_c^2 - P^2)\big].$$

As in the proof of Lemma F.4, write

$$w_0 = \mathbb{E}_P\Big[\big(\mathbb{E}_X[X^2\mathrm{sign}(\tau^2X^2 - P^2)]\big)^2\Big], \qquad \text{and} \qquad \gamma = 1 - w_0^2.$$

Lemma F.5. *Suppose that $2/5 \leq \tau \leq 9/20$ and $w_0 \geq 3/5$. Then $\gamma > 0$ and*

$$1 - r(c)^2 \geq \gamma\sqrt{1-c^2} \qquad \text{for } -1 < c < 1,$$

with equality if, and only if, $c = 0$.

Proof. We first obtain global bounds for $r(c)$ from the Hermite expansion, then prove a quantitative endpoint estimate, and finally combine the two bounds.

Global bounds. Use the functions $f_p(x) = x\mathrm{sign}(\tau^2x^2 - p^2)$ and Hermite weights w_j from the proof of Lemma F.4, whose construction applies for every $\tau > 0$. In particular, $w_j \geq 0$, $\sum_{j\geq 0} w_j = 1$, and $w_0 < 1$. Write $h_i = \mathrm{He}_i/\sqrt{i!}$ for the orthonormal probabilists' Hermite polynomials. For $i, j \geq 1$, the Gaussian moment generating function gives

$$\mathbb{E}[h_i(X)Z\,h_j(Y_c)T_c] = \mathbf{1}_{\{i=j\}}\big(c^{i+1} - i(1-c^2)c^{i-1}\big). \tag{F.1}$$

Indeed, for real a, b,

$$\mathbb{E}\Big[ZT_ce^{aX - a^2/2}e^{bY_c - b^2/2}\Big] = \big(c - (1-c^2)ab\big)e^{cab},$$

and comparing coefficients proves (F.1). Applying this identity to the Hermite expansion of f_p and averaging over P, we obtain, with $u = c^2$,

$$r(c) = \sum_{j\ge 0} w_j\big((2j+2)u - (2j+1)\big)u^j. \tag{F.2}$$

To justify the passage to the infinite expansion, multiplication by the independent Z preserves the L^2 norm, as does the orthogonal rotation giving (Y_c, T_c). Thus finite Hermite sums converge in the required expectations. Cauchy–Schwarz bounds each diagonal expectation in (F.1) by one in absolute value, so $\sum_j w_j = 1$ also justifies averaging over P.

For $0 \le u < 1$, each summand polynomial satisfies

$$\big((2j+2)u - (2j+1)\big)u^j = u^j\big(2u - 1 - 2j(1-u)\big) \le \max\{0, 2u-1\}.$$

For $j \ge 1$, its negative is at most $(2j+1)(1-u)u^j \le 3u$, since

$$j(1-u)u^{j-1} \le (1-u)\sum_{i=0}^{j-1} u^i = 1 - u^j \le 1, \qquad \text{and} \qquad \frac{2j+1}{j} \le 3.$$

Separating the $j = 0$ term in (F.2) gives

$$-w_0 + (5w_0 - 3)c^2 \le r(c) \le \max\{0, 2c^2 - 1\}. \tag{F.3}$$

In particular, $r(c) \ge -w_0$ because $w_0 \ge 3/5$.

Endpoint estimate. We next show that

$$1 - r(c) \ge \frac{2}{5}\sqrt{1-c^2} \qquad \text{for } 31/32 \le c < 1. \tag{F.4}$$

Write $c = \cos(2\delta)$ with $0 < \delta < \pi/2$, and put $h = \sin\delta$ and $\ell = \cos\delta$. Then $0 < h \le 1/8$, $\ell^2 = 1 - h^2 \ge 63/64$, and $s = 2h\ell$. Conditioning on X, Z gives

$$r(c) = \mathbb{E}[XZY_cT_c\,(1 - 4\,|\Phi(\tau|X|) - \Phi(\tau|Y_c|)|)]\,,$$

where Φ is the standard Gaussian distribution function. Define the radial integral

$$\Psi(a) = \int_0^\infty \rho^5 e^{-\rho^2/2}\left(\Phi(\tau\rho a) - \frac{1}{2}\right)\mathrm{d}\rho \qquad \text{for } 0 \le a \le 1.$$

The integrable bound $\tau\rho^6 e^{-\rho^2/2}/\sqrt{2\pi}$ justifies differentiation under the integral. The Gaussian sixth moment, together with $1 + \tau^2 \le (11/10)^2$ and $(11/10)^7 < 2$, yields

$$\frac{15\tau}{4} \le \Psi'(a) = \frac{15\tau}{2(1+\tau^2a^2)^{7/2}} \le \frac{15\tau}{2} \qquad \text{for } 0 \le a \le 1. \tag{F.5}$$

Using polar coordinates $X = \rho\cos\phi$ and $Z = \rho\sin\phi$, followed by $\theta = \phi - \delta$, we obtain

$$r(c) = 2c^2 - 1 - \frac{4}{\pi}I, \qquad \text{with } I = 2\int_0^{\pi/2} K(\theta)D(\theta)\,\mathrm{d}\theta, \tag{F.6}$$

where

$$K(\theta) = \sin^2\theta\cos^2\theta - h^2\ell^2, \quad \text{and} \quad D(\theta) = \Psi(\cos(\theta-\delta)) - \Psi(|\cos(\theta+\delta)|).$$

Here $\mathbb{E}[XZY_cT_c] = 2c^2 - 1$, the Gaussian density and polar Jacobian leave the radial weight $\rho^5 e^{-\rho^2/2}/(2\pi)$, and the original angular integrand is π-periodic and invariant under $\theta \mapsto \pi - \theta$. These identities give the factor $4/\pi$ in (F.6).

For $0 \le \theta \le \pi/2$, the two arguments of Ψ differ by $2\min\{h\sin\theta, \ell\cos\theta\}$. Consequently, (F.5) gives

$$\frac{15\tau}{2}\min\{h\sin\theta, \ell\cos\theta\} \le D(\theta) \le 15\tau\min\{h\sin\theta, \ell\cos\theta\}.$$

Moreover, since $\ell \geq 1/2$,

$$\sin^2\theta\cos^2\theta\min\{h\sin\theta, \ell\cos\theta\} \geq h\sin^3\theta\cos^2\theta - \frac{h^2}{2}\sin^4\theta\cos\theta.$$

When the minimum is $h\sin\theta$, this is immediate. Otherwise it follows from

$$\ell\cos^2\theta - h\sin\theta\cos\theta + \frac{h^2}{2}\sin^2\theta \geq \frac{1}{2}(\cos\theta - h\sin\theta)^2 \geq 0.$$

Apply the lower bound for $D(\theta)$ to the nonnegative term $\sin^2\theta\cos^2\theta$ in $K(\theta)$ and its upper bound to $-h^2\ell^2$. Using

$$\int_0^{\pi/2}\sin^3\theta\cos^2\theta\,\mathrm{d}\theta = \frac{2}{15}, \quad \int_0^{\pi/2}\sin^4\theta\cos\theta\,\mathrm{d}\theta = \frac{1}{5}, \quad \text{and} \quad \int_0^{\pi/2}\sin\theta\,\mathrm{d}\theta = 1,$$

we obtain

$$I \geq 15\tau h\left(\frac{2}{15} - \frac{h}{10} - 2h^2\ell^2\right) \geq h\left(\frac{4}{5} - \frac{3}{5}h - 12h^2\ell^2\right) \geq 0.$$

The first bracket is positive because $h \leq 1/8$ and $\ell^2 \leq 1$, and the second inequality uses $\tau \geq 2/5$. Since $\pi < 4$, (F.6) implies

$$1 - r(c) = 8h^2\ell^2 + \frac{4}{\pi}I \geq \frac{4}{5}h + h^2\left(8\ell^2 - \frac{3}{5} - 12h\ell^2\right) \geq \frac{4}{5}h \geq \frac{2}{5}s.$$

The last two inequalities use $\ell^2 \geq 63/64$, $h\ell^2 \leq 1/8$, and $s = 2h\ell \leq 2h$. This proves (F.4).

The minimum. Equation (F.2) shows that $r(-c) = r(c)$ and $r(0) = -w_0$, so it suffices to consider $0 \leq c < 1$. We have $0 < \gamma = 1 - w_0^2 \leq 16/25$. If $r(c) \leq 0$, (F.3) gives $|r(c)| \leq w_0$, and hence

$$\frac{1 - r(c)^2}{s} \geq \frac{\gamma}{s} \geq \gamma,$$

with strict inequality when $c > 0$.

Suppose that $r(c) > 0$ and $c \leq 31/32$. By (F.3), $u = c^2 > 1/2$ and $r(c) \leq 2u - 1$, so

$$\frac{1 - r(c)^2}{s} \geq 4u\sqrt{1-u}.$$

For $u \leq 3/4$, the right-hand side is at least 1. For $u \geq 3/4$, we have $s \geq \sqrt{63}/32 \geq 7/32$, so it is at least $21/32 > 16/25 \geq \gamma$.

Finally, suppose that $r(c) > 0$ and $31/32 \leq c < 1$. The endpoint estimate (F.4) gives $r(c) \leq 1 - (2/5)s$, whence

$$\frac{1 - r(c)^2}{s} \geq \frac{1 - (1 - (2/5)s)^2}{s} = \frac{4}{5} - \frac{4}{25}s > \frac{16}{25} \geq \gamma.$$

At $c = 0$, the identity $r(0) = -w_0$ gives equality, completing the proof. □

Lemma F.6 (Angular integration)**.** *Fix $d \geq 4$. Define $g(u, v) = \min\{\|u - v\|, \|u + v\|\}$ for $u, v \in \mathbb{S}^{d-2}$. There exists a constant $0 < C_d < \infty$, depending only on d, such that for every nonnegative measurable function f on $[0, \infty)$ and every $0 \leq a < b$,*

$$\int_{\substack{u,v\in\mathbb{S}^{d-2}\\ a\leq g(u,v)\leq b}} f(g(u,v))\mathrm{d}\sigma(u)\,\mathrm{d}\sigma(v) \leq C_d\int_a^b f(t)t^{d-3}\mathrm{d}t.$$

Here $\mathrm{d}\sigma$ denotes the surface measure on $\mathbb{S}^{d-2}$.

Proof. Fix u. Reflecting v through the origin produces a factor of two and reduces the angular integral to $0 \leq \theta \leq \pi/2$, where $u^\top v = \cos\theta$. Write $v = \cos\theta\, u + \sin\theta\, w$, where $w \in \mathbb{S}^{d-2} \cap u^\perp$, a

sphere identified with $\mathbb{S}^{d-3}$. On the hemisphere $\{v : u^\top v \geq 0\}$,

$$g(u,v) = \|u - v\| = \sqrt{2 - 2\cos\theta} = 2\sin(\theta/2), \qquad \text{and} \qquad \mathrm{d}\sigma(v) = (\sin\theta)^{d-3}\mathrm{d}\theta\,\mathrm{d}\sigma(w).$$

Integrating over v implies

$$\begin{aligned}
\int_{\substack{v\in\mathbb{S}^{d-2}\\ a\leq g(u,v)\leq b}} f(g(u,v))\mathrm{d}\sigma(v) &= 2\int_{\substack{0\leq\theta\leq\pi/2\\ a\leq 2\sin(\theta/2)\leq b}} \int_{\mathbb{S}^{d-3}} f(2\sin(\theta/2))(\sin\theta)^{d-3}\mathrm{d}\sigma(w)\,\mathrm{d}\theta \\
&= 2\left|\mathbb{S}^{d-3}\right| \int_{\substack{0\leq\theta\leq\pi/2\\ a\leq 2\sin(\theta/2)\leq b}} f(2\sin(\theta/2))(\sin\theta)^{d-3}\mathrm{d}\theta \\
&= 2\left|\mathbb{S}^{d-3}\right| \int_{[a,b]\cap[0,\sqrt{2}]} f(t)t^{d-3}(1-t^2/4)^{(d-4)/2}\mathrm{d}t,
\end{aligned}$$

where the last equality uses the change of variables $t = 2\sin(\theta/2)$. Integrating over u as well yields

$$\int_{\substack{u,v\in\mathbb{S}^{d-2}\\ a\leq g(u,v)\leq b}} f(g(u,v))\mathrm{d}\sigma(u)\,\mathrm{d}\sigma(v) = 2\left|\mathbb{S}^{d-2}\right|\left|\mathbb{S}^{d-3}\right| \int_{[a,b]\cap[0,\sqrt{2}]} f(t)t^{d-3}(1-t^2/4)^{(d-4)/2}\mathrm{d}t.$$

Since $d \geq 4$, the last factor is at most one. The bound therefore holds with $C_d = 2\left|\mathbb{S}^{d-2}\right|\left|\mathbb{S}^{d-3}\right|$. □

Lemma F.7 (Esséen concentration inequality [62, Exercise 2.2.11, p. 85])**.** *For every integer $m \geq 1$ and every $\varepsilon > 0$, there exists a constant $C_{m,\varepsilon} > 0$, depending only on m and ε, such that for every random vector $X \in \mathbb{R}^m$ and every $R > 0$,*

$$\sup_{z\in\mathbb{R}^m} \mathbb{P}\big(X \in z + B(0,R)\big) \leq C_{m,\varepsilon} R^m \int_{\|v\|\leq\varepsilon/R} \left|\mathbb{E}e^{iv^\top X}\right| \mathrm{d}v.$$

G Numerical experiment details

This appendix details how we test the criticality condition (4.2) numerically for the candidates in the first experiment of Section 5. The signal is fixed at $x^\star = e_1$, and the dimension, measurement ratios, and trial counts are specified in the main text.

Candidate construction. For each enumerated index set $I = \{i_1 < \cdots < i_d\}$ and nonconstant sign vector $s \in \{\pm 1\}^d$, we impose the active equations $a_{i_j}^\top \widehat{x} = s_j a_{i_j}^\top e_1$ for $j = 1, \ldots, d$ and form the candidate $\widehat{x} = x_{I,s}$ from (4.3). Let $S_I(\widehat{x})$ and $Q_I(\widehat{x})$ be as defined in Section 4, with the columns of Q_I ordered by $i_1, \ldots, i_d$.

Numerical verification. We recompute each candidate selected for verification and its multiplier $\lambda \in \mathbb{R}^d$ using 60-digit decimal arithmetic and then apply the following checks.

1. *Consistency of the active set.* We solve the selected active equations and require every unselected measurement to have a nonzero residual:

$$\min_{i\notin I} |(a_i^\top \widehat{x})^2 - (a_i^\top e_1)^2| > 10^{-30}.$$

   This check numerically separates the prescribed active measurements from the remaining ones.

2. *Criticality with an interior multiplier.* We solve $Q_I(\widehat{x})\lambda = -S_I(\widehat{x})$ and require

$$1 - \|\lambda\|_\infty > 10^{-30}.$$

   This condition keeps the multiplier strictly inside $(-1,1)^d$, as required by (4.4).

We count a problem instance as positive if at least one candidate passes both checks.