

Saddlepoint Approximation and Central Limit Theorem for Densities in high dimensions

Alexander Katsevich

Alexander.Katsevich@ucf.edu

University of Central Florida

Research Article

Keywords:

Posted Date: October 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11242148/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

SADDLEPOINT APPROXIMATION AND CENTRAL LIMIT THEOREM FOR DENSITIES IN HIGH DIMENSIONS

ALEXANDER KATSEVICH¹

ABSTRACT. We study the saddlepoint approximation (SPA) for densities of sums of high-dimensional random vectors, allowing the dimension d to grow with the sample size n . For averages of i.i.d. random vectors in $\mathbb{R}^d$, we obtain an explicit non-asymptotic bound for the multiplicative SPA error whose leading term is of order d^2/n . This improves on the generic d^4/n rate in recent work of Tang and Reid.

In the natural central region $\|x\| \lesssim (d/n)^{1/2}$, we obtain a further refinement. We prove that the leading SPA error is essentially constant throughout this region, up to $O(d^{5/2}/n^{3/2})$. This allows us, at the cost of computing a single Gaussian Monte Carlo integral, to estimate the SPA error in the central region and use the resulting value as a multiplicative correction to obtain a more accurate SPA. As further corollaries to the main SPA bound, we derive a local central limit theorem with explicit multiplicative error and, on the natural high-dimensional central region, a cubic-corrected Gaussian approximation with multiplicative error $O(d^2/n)$. We also obtain a more general SPA theorem that applies, in particular, to averages of independent but not necessarily identically distributed random vectors.

We further give readily verifiable sufficient conditions for the SPA assumptions, reducing the main high-dimensional Fourier condition to one-dimensional estimates for tilted marginals. These conditions apply, in particular, to arbitrary isotropic log-concave densities, uniformly well-conditioned affine transformations thereof and their convex mixtures, as well as to distributions obtained by smoothing with Gaussian or log-concave noise. The results are illustrated by multivariate generalized Laplace distributions and by a likelihood problem for averaged images with Gaussian-mixture noise.

1. INTRODUCTION

The saddlepoint approximation method (SPA), first proposed by Daniels in [16], is a powerful method with numerous applications in probability and statistics [12, 13, 21]. SPA provides accurate pointwise approximations to probability density functions (pdfs) and cumulative distribution functions (cdfs), based on the moment generating function (mgf). Initially proposed for scalar random variables, SPA has since been extended to vector-valued cases [28, 29], primarily in fixed dimension.

The behavior of saddlepoint approximations in growing dimension has received much less attention. Jensen [22] gave examples showing that the accuracy familiar from fixed-dimensional theory need not persist in high-dimensional inference. A significant recent advance is the work of Tang and Reid [35], who extended Laplace and saddlepoint approximations to settings in which the dimension d may grow with the sample size n . Under their most general assumptions, the relative SPA error is shown to be asymptotically of order $O(d^4/n)$.

In this paper we further extend SPA analysis and obtain an explicit non-asymptotic multiplicative error bound whose leading term is of order $O(d^2/n)$, in contrast to the asymptotic error estimate in [35]. Thus our result applies under the substantially weaker dimensional restriction $d^2/n \rightarrow 0$. The analysis also yields a refined description of the SPA error in the natural high-dimensional central region, including identification of its dominant contribution. In Section 5 we consider a multivariate generalized Laplace family with uniformly bounded means and uniformly well-conditioned covariance matrices. For this family, the estimates in [35] give the $O(d^4/n)$ rate, while our Theorem 3.1 gives $O(d^2/n)$.

¹School of Data, Mathematical, and Statistical Sciences, University of Central Florida, Orlando, FL 32816 (Alexander.Katsevich@ucf.edu).

A key step in [35] and in our paper is an asymptotic calculation of a high-dimensional Laplace-type integral arising after Fourier inversion through the multivariate saddlepoint. In our notation, the relevant local integral has the form

$$(1.1) \quad \int_{\mathcal{U}} \exp\{-nG(\tau + it)\} dt,$$

where $\mathcal{U}$ is a neighborhood of zero in $\mathbb{R}^d$ whose size goes to zero as $d, n \rightarrow \infty$. The difference between the value of this integral, suitably normalized, and one gives the multiplicative SPA error.

Very recently, similar Laplace-type integrals have been studied by Anya Katsevich [26]. She showed that complete asymptotic expansions can be obtained under a variety of conditions. In the case closest to ours, the leading error is also of order $O(d^2/n)$. Closely related work on the Laplace approximation of probability measures of the form $d\pi \propto e^{-nV(x)}dx$ in growing dimensions is contained in [23, 24]. High-dimensional posterior normality and Bernstein–von Mises-type results are also closely related; see, e.g., [3, 25] and the references therein. Such results give global Gaussian approximations of the posterior in probability metrics such as total variation, whereas the SPA studied here gives a pointwise approximation with a multiplicative error bound for the density on a prescribed region.

The approach in our paper is a blend of those in [35] and [26]. Similarly to [35], we use the inverse Laplace/Fourier transform and integrate across the multivariate saddlepoint; see also [29]. Similarly to [26], we use a concentration inequality. This is the step that allows us to reduce the dimension dependence to d^2 . There are two main differences from [26]. First, the exponent in our problem is complex-valued, whereas in [26] it is real-valued, and the argument there uses this fact in an essential way. We instead bound the oscillatory part using a Poincaré/spectral-gap type inequality [2, eq. (4.9.4)]. Second, to obtain a sharper estimate, we employ thermodynamic integration, which expresses one of the key integrals in terms of a variance that can be estimated using the same inequality. The result is an explicit non-asymptotic pointwise bound for the multiplicative SPA error.

Our *first main result*, Theorem 3.1, gives this bound for the density of the average of n i.i.d. random vectors $X_i \in \mathbb{R}^d$. Our *second main result* is an additional refinement obtained in Subsection 4.1. In centered coordinates, let $I(a)$ denote the multiplicative correction to the ordinary SPA and set $\Sigma = \text{Var}(X)$. We show that, for every fixed $L > 0$,

$$(1.2) \quad I(a) - 1 = J(0) - 1 + O_L(d^{5/2}/n^{3/2}) + \text{e.s.t.}, \quad \|\Sigma^{-1/2}a\| \leq L(d/n)^{1/2},$$

where e.s.t. stands for exponentially small terms, and $J(0)$ is a single Gaussian integral evaluated at the central saddlepoint. Thus the leading SPA error is essentially constant throughout the natural high-dimensional central region. In particular, multiplying the SPA by the correction factor $J(0)$ improves the relative error from order d^2/n to order $d^{5/2}/n^{3/2}$ there, a gain of a factor $(d/n)^{1/2}$; see Theorem 4.1. Moreover, $J(0)$ is an expectation with respect to the standard Gaussian measure and can be estimated by Monte Carlo only once for each fixed model, d , and n . Subsection 4.1 provides an explicit probabilistic bound for the resulting approximation; see (4.24).

We next obtain a local central limit theorem (CLT). After affine normalization, Corollary 4.2 gives an explicit multiplicative error bound. On the above natural high-dimensional central region, our *third main result*, Corollary 4.3, further gives a cubic-corrected Gaussian approximation with multiplicative error $O(d^2/n)$.

The correction appearing in the exponent is precisely the first multivariate Edgeworth correction; see, e.g., [7, Chapters 2 and 4]. In the standard first-order Edgeworth approximation, however, this term appears additively in the polynomial prefactor multiplying the Gaussian density. Our result instead places the same correction in the exponent and can therefore be viewed as an exponentiated first-order Edgeworth approximation.

Moving this correction to the exponent is important in high dimension. A closely related phenomenon occurs for high-dimensional Laplace expansions. For the usual additive expansion, the condition $d^2/n \rightarrow 0$ is in general sharp and cannot simply be relaxed by obtaining better

remainder estimates; see [26, Example 2.19 and Proposition 2.20]. By contrast, after passing to an expansion of the logarithm, and hence to an exponential rather than additive representation of the correction, asymptotic expansions can remain valid in regimes arbitrarily close to the concentration threshold $d/n \rightarrow 0$ [27]. The reason is that exponentiation retains all powers generated by the leading correction, whereas an additive first-order expansion truncates after its linear contribution.

The same mechanism is relevant here. On the natural high-dimensional central region, the first Edgeworth correction is generally of order $(d^3/n)^{1/2}$ and therefore need not be small when $d^2/n \rightarrow 0$. Retaining this term in the exponent nevertheless leaves a multiplicative remainder of order $O(d^2/n)$ in Corollary 4.3. Thus our result remains accurate in regimes in which the corresponding additive first-order Edgeworth correction is no longer perturbative. As a result we show that multiplicative convergence to the *uncorrected* Gaussian density generically requires the stronger condition $d^3/n \rightarrow 0$; the condition $d^2/n \rightarrow 0$ is not sufficient; see (4.33).

CLTs for vectors of fixed dimension are classical [8, Sections 27 and 29], but the case in which $d \rightarrow \infty$ together with n is much less understood; see [14] for a recent review. An important early distributional result of Portnoy [34] identifies $d^2/n \rightarrow 0$ as a fundamental scaling for multivariate central limit behavior in growing dimension.

Recent work of Bobkov and Götze [11] gives multidimensional Berry–Esseen bounds for the uniform local error $\sup_x |p_n(x) - \gamma_d(x)|$ under substantially weaker moment assumptions than those imposed here. Their results provide global absolute bounds, while ours give pointwise multiplicative bounds. The latter can therefore capture improved accuracy in a prescribed region, including regions where the density itself is smaller than a global absolute error bound.

Pointwise Edgeworth expansions for densities in growing dimension have also been considered recently. Withers [37] derives classical additive Edgeworth expansions for a broad class of statistics of dimension d , organized in powers of $d^3/\sqrt{n}$.

There is an extensive related literature on high-dimensional Gaussian approximation in metrics of the form $\sup_{D \in \mathcal{D}} |\pi(D) - \gamma(D)|$, where π is the target distribution, γ is its Gaussian approximation, and $\mathcal{D}$ is a family of sets. Classical Berry–Esseen bounds for convex sets include [4, 5, 18], while sharper results are available for special geometric classes; see, for example, [6]. More recently, Gaussian approximation for hyperrectangles and related classes has been developed in [15] and subsequent work. Higher-order approximations for probabilities of sets in growing dimension have also been studied; see, for example, Anderson, Hall and Titterton [1] for asymptotic Edgeworth expansions and Zhilova [40] for non-asymptotic Edgeworth-type expansions for Euclidean balls and half-spaces.

Many such results are obtained under substantially weaker moment assumptions and can allow the dimension to grow much faster than is possible for pointwise density approximation. This broader range of applicability reflects the fact that approximation of probabilities of sets and pointwise density approximation are fundamentally different problems, both in the form of the approximation and in the conditions required for its validity. In particular, approximation of probabilities of sets does not determine the density at an observed point. One important consequence is that it does not provide the likelihood needed for likelihood-based inference.

We also give a more general SPA theorem, Theorem 4.5, which does not require the random vector under consideration to be an average of identically distributed summands. This includes, in particular, averages of independent but non-identically distributed random vectors.

Classical SPA theory applies to broad classes of distributions in fixed dimension. In particular, uniform saddlepoint approximations for log-concave (LC) and related densities were developed by Jensen [20, 21]; the classical theory also includes multivariate, conditional, and non-identically distributed settings. In this paper we establish conditions under which some of these results generalize to increasing dimension $d = d_n \rightarrow \infty$.

A further main contribution is that the assumptions of the SPA theorem can be verified from relatively simple properties of the underlying distribution. Our *fourth main result*, Lemma 6.1, gives a sufficient criterion in terms of uniform exponential-moment bounds, covariance control, and integral Hölder regularity of one-dimensional tilted marginals. In this way the d -dimensional

Fourier condition is reduced to 1D estimates. Thus, once the 1D tilted marginals are established to possess the required property, these multidimensional estimates need not be checked separately. Using this approach, we verify the assumptions for important and broad classes of distributions; see item (2) below.

Taken together, these results go beyond an improvement in the dimension dependence of the conventional SPA. In addition to obtaining the explicit non-asymptotic $O(d^2/n)$ error bound, we provide readily verifiable sufficient conditions for its applicability to broad classes of high-dimensional distributions, a refined analysis of the SPA error in the natural central region together with a computable multiplicative correction, and a local CLT with a multiplicative cubic correction. Thus the results address both the accuracy of high-dimensional SPA and the practical verification of the assumptions under which that accuracy holds.

We illustrate the theory in three complementary settings.

- (1) Section 5 treats the multivariate generalized Laplace distribution, a non-Gaussian and generally nonseparable family for which all assumptions can be checked explicitly. It also provides the direct comparison with the dimension dependence in [35] discussed above.
- (2) Section 7 applies our sufficient criterion to several broad classes of distributions. We first consider isotropic LC densities, then uniformly well-conditioned affine transformations and convex mixtures of LC densities, and then Gaussian-smoothed distributions $X = Y + Z$. Here the law of Y may be supported on a lower-dimensional set, be singular, or discrete, while a uniformly nondegenerate Gaussian component Z ensures the required regularity. More generally, Z can be any random vector satisfying the conditions of Remark 7.5.
- (3) Finally, Section 8 considers likelihood-based inference from averaged images under a non-Gaussian, nonseparable two-component Gaussian-mixture noise model. SPAs have previously been used to construct approximate likelihoods in fixed-dimensional and applied multivariate problems; see, e.g., [17, 38]. Here we study the same use in the growing-dimensional setting. The exact likelihood of the averaged image is available as a finite Gaussian mixture, allowing direct comparison of the conventional SPA with its Gaussian and cubic central expansions. This illustrates a use of pointwise density approximation that is not accessible from approximations to probabilities of sets alone. The experiment also tests the non-asymptotic SPA bound and the Monte Carlo refinement of Subsection 4.1: a single Monte Carlo estimate of the SPA error at the central saddlepoint remains representative of the true error throughout the displayed central region.

The paper is organized as follows. Section 2 introduces the notation and problem setting, and Section 3 proves the main SPA theorem. Section 4 contains the central Monte Carlo refinement, the local CLT, the cubic central density expansion, and the more general SPA theorem. Section 5 treats the multivariate generalized Laplace family and compares the resulting dimension dependence with [35]. Section 6 develops a sufficient criterion for the assumptions, which is easier to verify. Section 7 applies the criterion to several broad classes of distributions. Section 8 presents the averaged-image likelihood experiment described above. Auxiliary calculations are collected in the appendices.

2. NOTATION. PROBLEM SETTING

Let $\|u\|$ denote the standard Euclidean norm in $\mathbb{R}^d$. The operator norm of a k -th order tensor is defined as

$$(2.1) \quad \|T\|_{\text{op}} := \sup_{\|u_1\|=\dots=\|u_k\|=1} |T[u_1, \dots, u_k]|,$$

where

$$(2.2) \quad T[u_1, \dots, u_k] = \sum_{i_1, \dots, i_k=1}^d T_{i_1 \dots i_k} (u_1)_{i_1} (u_2)_{i_2} \dots (u_k)_{i_k}.$$

If T is symmetric, we can set $u_1 = \dots = u_k$ in (2.1) [39]. For a function $f \in C^k(\mathbb{R}^d)$, the k -th derivative tensor is

$$(2.3) \quad (\nabla^k f(x))_{i_1 \dots i_k} = \partial_{x_{i_1}} \dots \partial_{x_{i_k}} f(x).$$

For matrices, we will write $\|\cdot\|$, but the operator/spectral norm is understood.

Let X and X_i , $i \geq 1$, be independent, identically distributed (i.i.d.) random vectors in $\mathbb{R}^d$. Let the mgf and cgf of X be denoted

$$(2.4) \quad \phi(\tau) = \mathbb{E}e^{\langle \tau, X \rangle}, \quad \varphi(\tau) = \log \phi(\tau),$$

respectively. We analytically continue $\varphi(\tau)$ to $\mathbb{C}^d$ in a neighborhood of the origin. Consider a sequence of dimensions $d = d_n$, $d \rightarrow \infty$ as $n \rightarrow \infty$. Let S^{d-1} be the unit sphere in $\mathbb{R}^d$, I_d be the $d \times d$ identity matrix, and denote $\epsilon = d/n$.

Throughout the paper we use the convention that the inequality sign $\lesssim$ absorbs various absolute multiplicative constants which are independent of X , d , and n . Similarly, we write $A \lesssim B$ for positive semidefinite matrices if $\langle Ax, x \rangle \leq c \langle Bx, x \rangle$ for some $c > 0$ and all $x \in \mathbb{R}^d$. If $A \lesssim B$ and $B \lesssim A$, then we write $A \asymp B$.

Suppose $\phi(t)$ satisfies the following assumption.

Assumption 2.1. For all n sufficiently large,

(1) One has

$$(2.5) \quad \sup_{\Theta \in S^{d-1}} \sup_{\tau \in V_d} \mathbb{E}[\langle \Theta, X \rangle^4 e^{\langle \tau, X \rangle}] < \infty$$

for a convex neighborhood of the origin $0 \in V_d \subset \mathbb{R}^d$.

(2) $H(\tau) := \nabla^2 \varphi(\tau) > 0$ for all $\tau \in V_d$ and $\sup_{\tau \in V_d} \|H(\tau)^{-1}\| < \infty$.

(3) One has

$$(2.6) \quad \left| \frac{\phi(\tau + iH^{-1/2}t)}{\phi(\tau)} \right| \leq \max \left(\exp \{ -\epsilon^{1/2} \|t\| \}, (1 + \|t\|)^{-\frac{\kappa}{d^{1/2}}} \right),$$

$$\tau \in V_d, \quad \|t\| \geq 2.5\epsilon^{1/2},$$

for some $\kappa > 0$. Here $H = H(\tau)$.

(4) There exists $\delta > 0$ such that

$$(2.7) \quad |\text{Arg } \phi(\tau + iH^{-1/2}t)| \leq \pi - \delta, \quad |\phi(\tau + iH^{-1/2}t)| \geq \delta,$$

$$\tau \in V_d, \quad \|t\| \leq 2.5\epsilon^{1/2}.$$

Here, $\text{Arg } z$ denotes the principal branch of the argument of $z \in \mathbb{C}$. For notational simplicity, we drop the subscript n from d_n in what follows. To clarify, the constants κ and δ are independent of d and n . In what follows we continue using the notation $H := H(\tau)$.

Introduce the constants

$$(2.8) \quad c_3 := \sup_{\tau \in V_d, \|t\| < 2.5\epsilon^{1/2}} \|\nabla_t^3 \text{Im } \varphi(\tau + iH^{-1/2}t)\|_{\text{op}},$$

$$c_4 := \sup_{\tau \in V_d, \|t\| < 2.5\epsilon^{1/2}} \|\nabla_t^4 \text{Re } \varphi(\tau + iH^{-1/2}t)\|_{\text{op}}.$$

Assumptions 2.1(1, 2, 4) imply that $c_3, c_4 < \infty$ for each d , but they may grow with d . Also, Assumption 2.1(1) implies that $\phi(z)$ is analytic in the polystrip $z = \tau + it$, $\tau \in V_d$, $t \in \mathbb{R}^d$.

A useful stronger version of Assumption 2.1(1) is the uniform bound

$$(2.9) \quad \sup_{\|\tau\| < r} \mathbb{E}e^{\langle \tau, X \rangle} \lesssim 1$$

for some $r > 0$ and all d . We have dropped here the factor $\langle \Theta, X \rangle^4$, because,

$$(2.10) \quad |\langle \Theta, x \rangle|^m \leq C_{h,m} (e^{h\langle \Theta, x \rangle} + e^{-h\langle \Theta, x \rangle}), \quad x \in \mathbb{R}^d,$$

for any $h, m > 0$ and some $C_{h,m}$. Thus, reducing r to r' and taking any $0 < h < r - r'$, we obtain (2.5) with $V_d = \{\|\tau\| \leq r'\}$.

Suppose in addition that

$$(2.11) \quad \text{Var}(X) \asymp I_d.$$

Then there exists $0 < r' \leq r$, independent of d , such that, with $V_d = \{\|\tau\| < r'\}$,

$$(2.12) \quad H(\tau) \asymp I_d, \quad \tau \in V_d.$$

Indeed, (2.9) gives uniform bounds on the derivatives of ϕ of every fixed order in a fixed complex neighborhood of the origin. After decreasing r' , if necessary, ϕ is uniformly bounded away from zero there, and hence the corresponding derivatives of $\varphi = \log \phi$ are uniformly bounded as well. In particular, $H(0) = \text{Var}(X)$ and $\|H(\tau) - H(0)\| \lesssim \|\tau\|$. After decreasing r' once more, if necessary, Assumption 2.1(2) therefore follows from (2.11) and (2.9). The same argument shows that Assumption 2.1(4) holds for all ϵ sufficiently small and that the constants c_3, c_4 in (2.8) are bounded independently of d .

We summarize the preceding discussion in a lemma.

Lemma 2.2. *Let the random vector X satisfy (2.9) and (2.11). There exist $0 < r' \leq r$ and $\epsilon_0 > 0$ such that, provided $\epsilon = d/n \leq \epsilon_0$,*

- (1) *parts (1), (2) and (4) of Assumption 2.1 are satisfied with $V_d := \{\|\tau\| < r'\}$, and*
- (2) *constants c_3, c_4 , defined in (2.8) using the same V_d , are bounded independently of d, n .*

Introduce the key function

$$(2.13) \quad \begin{aligned} G_a(t) &= -\varphi(\tau + iH^{-1/2}t) + \varphi(\tau) + i\langle H^{-1/2}t, a \rangle, \\ \tau &= (\nabla\varphi)^{-1}(a), \quad \|t\| < 2.5\epsilon^{1/2}, \quad a \in U_d := \nabla\varphi(V_d). \end{aligned}$$

This relation between a and τ is assumed in what follows. Note that $(\nabla\varphi)^{-1}(a)$ is well-defined (i.e., single-valued) due to the assumption $\nabla^2\varphi(\tau) > 0$, $\tau \in V_d$. By assumption and construction,

$$(2.14) \quad G_a(t) = \frac{1}{2}\|t\|^2 + O(\|t\|^3), \quad t \rightarrow 0.$$

The absence of the linear term in the Taylor expansion follows from the relation $\nabla\varphi(\tau) = a$ and the analyticity of $\varphi(z)$ in a neighborhood of $z = \tau$. Indeed, since $\varphi(z)$ is analytic, the derivative can be taken along any direction in $\mathbb{C}^d$. Therefore,

$$(2.15) \quad \varphi(\tau + iH^{-1/2}t) = \varphi(\tau) + \nabla\varphi(\tau)iH^{-1/2}t - \frac{1}{2}\|t\|^2 + O(\|t\|^3), \quad t \rightarrow 0,$$

where $\nabla\varphi(\tau) = \nabla_\tau\varphi(\tau)$, and the assertion follows.

From (2.4) it follows immediately that $\text{Re } G_a(t)$ and $\text{Im } G_a(t)$ are even and odd in t , respectively.

Let us now discuss Assumption 2.1(3). The bound on the right is the maximum of two functions. The first of these is similar to the one in [26, Assumption 2.3]. This function dominates on the disk $\|t\| \leq \rho_n$, where ρ_n solves the equation

$$(2.16) \quad \frac{\log(1 + \rho_n)}{\rho_n} = \frac{d}{\kappa\sqrt{n}}.$$

Thus, ρ_n grows slightly faster than linearly with $\sqrt{n}/d$. Taking logarithm on both sides of (2.6) we get

$$(2.17) \quad \text{Re } \varphi(\tau + iH^{-1/2}t) - \varphi(\tau) \leq -\epsilon^{1/2}\|t\|, \quad 2.5\epsilon^{1/2} \leq \|t\| \leq \rho_n.$$

Using (2.15), this means that for any fixed $c > 2.5$, one should have

$$(2.18) \quad \epsilon^{1/2}\|t\| \leq \frac{1}{2}\|t\|^2 + O(\|t\|^3), \quad 2.5\epsilon^{1/2} \leq \|t\| \leq c\epsilon^{1/2},$$

which clearly holds provided ϵ is sufficiently small. This means that the first term in (2.6) is consistent with the local quadratic behavior in (2.15). See [26, Remark 2.4] for a more detailed discussion of this condition.

For $\|t\| \geq \rho_n$, the second function inside the maximum in (2.6) dominates. It is used to avoid exponential decay at infinity. In fact, since $d \rightarrow \infty$, we see that a fairly slow decay as $t \rightarrow \infty$ is allowed when $d \gg 1$.

Observe that the mgf of $\bar{X}_n := (1/n) \sum_{i=1}^n X_i$ is $\phi^n(\tau/n)$. By the Laplace transform inversion formula, the pdf of $\bar{X}_n$ is given by

$$(2.19) \quad \rho_n(y) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \phi^n((p+iq)/n) e^{-\langle p+iq, y \rangle} dq, \quad y \in \mathbb{R}^d,$$

for any p such that $p/n \in V_d$. Pick any $a \in U_d$, set $y = a$, select p so that $\tau = p/n \in V_d$, and change variables $t = H^{1/2}q/n$. This gives

$$(2.20) \quad \begin{aligned} \rho_n(a) &= \frac{n^d}{(2\pi)^d \sqrt{\det H}} \int_{\mathbb{R}^d} \phi^n(\tau + iH^{-1/2}t) e^{-n\langle \tau + iH^{-1/2}t, a \rangle} dt \\ &= \frac{n^d}{(2\pi)^d \sqrt{\det H}} e^{-n\varphi^*(a)} \int_{\mathbb{R}^d} \left[\frac{\phi(\tau + iH^{-1/2}t)}{\phi(\tau)} \right]^n e^{-in\langle H^{-1/2}t, a \rangle} dt, \\ \varphi^*(a) &= \langle \tau, a \rangle - \varphi(\tau), \quad a \in U_d. \end{aligned}$$

Denote

$$(2.21) \quad I(a) = C_{d,n} \int_{\mathbb{R}^d} \left[\frac{\phi(\tau + iH^{-1/2}t)}{\phi(\tau)} \right]^n e^{-in\langle H^{-1/2}t, a \rangle} dt, \quad C_{d,n} := \frac{n^{d/2}}{(2\pi)^{d/2}}.$$

Then

$$(2.22) \quad \rho_n(a) = \frac{C_{d,n}}{\sqrt{\det H}} e^{-n\varphi^*(a)} I(a), \quad a \in U_d.$$

The conventional SPA step consists of replacing $I(a)$ by 1 in (2.22).

3. FIRST SPA RESULT AND ITS PROOF

Set

$$(3.1) \quad \begin{aligned} R &= 2.5, \quad \mathcal{U} := \{t \in \mathbb{R}^d : \|t\| < R\epsilon^{1/2}\}, \quad \tilde{\mathcal{U}} := \{s \in \mathbb{R}^d : \|s\| < Rd^{1/2}\}, \\ \lambda_* &:= 1 - \frac{c_4 R^2}{2} \frac{d}{n}, \quad A_0 := \frac{c_4 d(d+2)}{24n}, \quad B_0 := \frac{c_4^2 d(d+4)(d+9)}{72\lambda_*^4 n^2}, \\ D_0 &:= A_0 + B_0. \end{aligned}$$

Theorem 3.1. *Suppose Assumption 2.1 is satisfied and $\lambda_* > 0$. For all d^2/n sufficiently small, one has*

$$(3.2) \quad |I(a) - 1| \leq e^{D_0} \left(D_0 + \frac{c_3^2}{8\lambda_*^3} \frac{d(d+4)}{n} \right) + c \left[\exp(-d) + \left(\frac{e}{\kappa^2} \frac{d^2}{n} \right)^{d/2} \right], \quad a \in U_d,$$

for an absolute constant $c > 0$.

Importantly, the unspecified absolute constant c appears only in the exponentially small remainder. It can be computed, but we do not do it to avoid irrelevant technical details. The leading, algebraically decaying part of the bound is completely explicit and has a simple, transparent form in terms of c_3 , c_4 , and λ_* . In particular, if $c_3, c_4 = O(1)$ and $d^2/n \rightarrow 0$, then

$$(3.3) \quad \lambda_*^{-3} = 1 + O(d/n), \quad D_0 = \left(\frac{c_4}{24} + O(1/d) \right) \frac{d^2}{n},$$

and we get an estimate

$$(3.4) \quad |I(a) - 1| \leq \exp \left\{ \left(\frac{c_4}{24} + O(1/d) \right) \frac{d^2}{n} \right\} \left(\frac{3c_3^2 + c_4}{24} + O(1/d) \right) \frac{d^2}{n} + \text{e.s.t.}$$

Here and throughout the paper, e.s.t. denotes a term bounded in absolute value, up to a constant prefactor independent of d and n , by

$$(3.5) \quad \mathcal{E}_{d,n} := e^{-d} + \left(\frac{e}{\kappa^2} \frac{d^2}{n} \right)^{d/2}.$$

Our proof broadly follows the approach in [26]. The tail estimate is treated similarly, but there are two important differences in the analysis of the local integral. First, the exponent

$-nG_a(t)$ in our problem (see (3.17)) is complex-valued, whereas the exponent in [26] is real-valued. The logarithmic Sobolev argument used there therefore does not apply directly to the oscillatory factor. We instead separate the real and imaginary parts of the exponent and bound the contribution of the phase using a Poincaré/Brascamp–Lieb inequality for the probability measure generated by $\exp\{-n \operatorname{Re} G_a(t)\}$.

Second, to carefully estimate one of the integrals involving $\exp\{-n \operatorname{Re} G_a(t)\}$, we use thermodynamic integration. This expresses the remainder in terms of a variance, which can again be bounded by the same Poincaré/Brascamp–Lieb inequality.

3.1. Tail integrals. Denote (cf. (2.21))

$$(3.6) \quad I_{\mathcal{U}^c} := C_{d,n} \int_{\mathcal{U}^c} \left[\frac{\phi(\tau + iH^{-1/2}t)}{\phi(\tau)} \right]^n e^{-in\langle H^{-1/2}t, a \rangle} dt.$$

By (2.6),

$$(3.7) \quad \begin{aligned} |I_{\mathcal{U}^c}| &\leq I_1 + I_2, \\ I_1 &:= C_{d,n} \int_{\|t\| \geq R\epsilon^{1/2}} \exp\{- (dn)^{1/2} \|t\|\} dt, \\ I_2 &:= C_{d,n} \int_{\mathbb{R}^d} (1 + \|t\|)^{-\mu} dt, \quad \mu := \kappa \frac{n}{d^{1/2}}. \end{aligned}$$

Write the integral in I_1 in spherical coordinates and change variables

$$(3.8) \quad \begin{aligned} I_1 &= \frac{|S^{d-1}|}{(2\pi)^{d/2}} \int_{\mathbb{R}^{d^{1/2}}} \exp\{-d^{1/2}\rho\} \rho^{d-1} d\rho \\ &= \frac{|S^{d-1}|}{(2\pi)^{d/2}} \int_{\mathbb{R}^{d^{1/2}}} \exp\{-\psi(\rho)\} d\rho, \\ \psi(\rho) &:= d^{1/2}\rho - (d-1)\log \rho. \end{aligned}$$

Here $|S^{d-1}|$ is the area of the unit sphere in $\mathbb{R}^d$. Since $R > 1$, the critical point $\rho_* = (d-1)/d^{1/2}$ satisfies $\rho_* < Rd^{1/2}$. Hence

$$(3.9) \quad \psi'(\rho) \geq \psi'(Rd^{1/2}) \geq d^{1/2}[1 - (1/R)], \quad \rho \geq Rd^{1/2},$$

and therefore

$$(3.10) \quad \psi(\rho) \geq \psi(Rd^{1/2}) + d^{1/2}[1 - (1/R)](\rho - Rd^{1/2}), \quad \rho \geq Rd^{1/2}.$$

Using this inequality in (3.8), integrating, and using the Stirling formula, we obtain

$$(3.11) \quad I_1 \leq \frac{|S^{d-1}|}{(2\pi)^{d/2}} \frac{R^{d-1} d^{(d-1)/2} e^{-Rd}}{d^{1/2}[1 - (1/R)]} \lesssim \frac{R^d e^{d/2} e^{-Rd}}{d^{1/2}(R-1)}.$$

Since $R = 2.5$ satisfies $R \geq [(1/2) + \log R] + 1$,

$$(3.12) \quad I_1 \lesssim \frac{\exp(-d)}{d^{1/2}}.$$

Finally, consider I_2 . Our assumptions imply

$$(3.13) \quad d/\mu = \frac{d^{3/2}}{\kappa n} = \frac{1}{\kappa d^{1/2}} \frac{d^2}{n} \rightarrow 0$$

as $d, n \rightarrow \infty$. Evaluate the integral in the definition of I_2 and use the Stirling formula to get

$$(3.14) \quad \begin{aligned} I_2 &= C_{d,n} |S^{d-1}| \int_0^\infty (1 + \rho)^{-\mu} \rho^{d-1} d\rho \\ &= \frac{n^{d/2}}{(2\pi)^{d/2}} \frac{2\pi^{d/2}}{\Gamma(d/2)} \frac{\Gamma(d)\Gamma(\mu-d)}{\Gamma(\mu)} = \frac{(2n)^{d/2} \Gamma((d+1)/2)}{\pi^{1/2}} \frac{\Gamma(\mu-d)}{\Gamma(\mu)}. \end{aligned}$$

Using Stirling's formula we compute

$$(3.15) \quad I_2 \lesssim \frac{(edn)^{d/2}(\mu - d)^{\mu-d}}{\mu^\mu} = \left(\frac{edn}{\mu^2}\right)^{d/2} (1 - (d/\mu))^{\mu-d}.$$

Since $edn/\mu^2 = (e/\kappa^2)d^2/n$ and $(1 - (d/\mu))^{\mu-d} < 1$,

$$(3.16) \quad I_2 \lesssim \left(\frac{e}{\kappa^2} \frac{d^2}{n}\right)^{d/2}.$$

3.2. Local integrals. Consider the part of the integral in (2.21) over $\mathcal{U}$. Assumption 2.1(4) implies that the cgf $\varphi(\tau + iH^{-1/2}t)$ is well defined for $\tau \in V_d$, $t \in \mathcal{U}$. Using (2.13),

$$(3.17) \quad I_{\mathcal{U}} = C_{d,n} \int_{\mathcal{U}} e^{-nG_a(t)} dt.$$

For simplicity, drop the subscript a from G_a and write $G(t) = G_r(t) + iG_i(t)$. By construction,

$$(3.18) \quad G_r(t) = \frac{1}{2}\|t\|^2 + O(\|t\|^4), \quad G_i(t) = O(\|t\|^3), \quad t \rightarrow 0.$$

Set $r(t) := G_r(t) - \frac{1}{2}\|t\|^2$. Changing variables $s = n^{1/2}t$ gives

$$(3.19) \quad \begin{aligned} I_{\mathcal{U}} &= I_0 - I_r - iI_i, \\ I_0 &:= \int_{\tilde{\mathcal{U}}} e^{-q(s)} \gamma(ds), \quad q(s) := nr(s/n^{1/2}), \\ I_r &:= \int_{\tilde{\mathcal{U}}} (1 - \cos(nG_i(s/n^{1/2}))) e^{-q(s)} \gamma(ds), \\ I_i &:= C_{d,n} \int_{\mathcal{U}} \sin(nG_i(t)) e^{-nG_r(t)} dt, \end{aligned}$$

where $\gamma = \mathcal{N}(0, I_d)$ denotes the standard Gaussian measure on $\mathbb{R}^d$. As was mentioned before, G_r is even and G_i is odd. Using that $\mathcal{U}$ is symmetric gives $I_i = 0$.

To bound the remaining integrals, we first obtain a convexity estimate which will be used repeatedly. Since G_r is even, $\nabla^2 G_r(0) = I_d$, and $\nabla^3 G_r(0) = 0$, we get

$$(3.20) \quad \nabla^2 G_r(t) - I_d = \int_0^1 (1-u) \nabla^4 G_r(ut) [t, t, \cdot, \cdot] du.$$

By (2.8), for $t \in \mathcal{U}$,

$$(3.21) \quad \|\nabla^2 G_r(t) - I_d\| \leq \frac{c_4}{2} \|t\|^2 \leq \frac{c_4 R^2}{2} \frac{d}{n},$$

and therefore

$$(3.22) \quad \nabla^2 G_r(t) \succeq \lambda_* I_d, \quad t \in \mathcal{U}.$$

We use the following result. If ν is a probability measure on $\tilde{\mathcal{U}}$ with density proportional to $e^{-V(s)}$ and $\nabla^2 V \succeq \lambda_* I_d$, then by [10, Proposition 2.1]

$$(3.23) \quad \text{Var}_\nu(f) \leq \frac{1}{\lambda_*} \int_{\tilde{\mathcal{U}}} \|\nabla f(s)\|^2 \nu(ds).$$

For the measures ν considered below, $\mathbb{E}_\nu s = 0$. Hence, applying (3.23) first to the coordinate functions $f = s_j$ and then to $f = \|s\|^2$ gives

$$(3.24) \quad \int_{\tilde{\mathcal{U}}} \|s\|^2 \nu(ds) \leq \frac{d}{\lambda_*}, \quad \int_{\tilde{\mathcal{U}}} \|s\|^4 \nu(ds) \leq \frac{d(d+4)}{\lambda_*^2}.$$

Similarly, applying (3.23) to $\|s\|^3$ and using Cauchy–Schwarz,

$$(3.25) \quad \int_{\tilde{\mathcal{U}}} \|s\|^6 \nu(ds) \leq \frac{d(d+4)(d+9)}{\lambda_*^3}.$$

We next estimate I_r . Define the probability measure

$$(3.26) \quad \nu_1(ds) := I_0^{-1} e^{-q(s)} 1_{\tilde{\mathcal{U}}}(s) \gamma(ds)$$

and let $v(s) := \sin(nG_i(s/n^{1/2})/2)$. Here we use that I_0 , defined in (3.19), is the normalizing constant of the measure proportional to $e^{-q(s)}1_{\tilde{\mathcal{U}}}(s)\gamma(ds)$. Since v is odd and q in (3.19) is even, we see that ν_1 is even and $\mathbb{E}_{\nu_1}v = 0$. Thus

$$(3.27) \quad I_r = 2I_0\mathbb{E}_{\nu_1}v^2 = 2I_0\text{Var}_{\nu_1}(v).$$

The potential of ν_1 , apart from an additive constant, is

$$(3.28) \quad V_1(s) = \frac{1}{2}\|s\|^2 + q(s) = nG_r(s/n^{1/2}).$$

By (3.22), $\nabla^2 V_1(s) \succeq \lambda_* I_d$ on $\tilde{\mathcal{U}}$. Therefore, by (3.23), with $\nu = \nu_1$:

$$(3.29) \quad I_r \leq \frac{2I_0}{\lambda_*} \int_{\tilde{\mathcal{U}}} \|\nabla v(s)\|^2 \nu_1(ds).$$

We have

$$(3.30) \quad \nabla v(s) = \frac{n^{1/2}}{2} \cos(nG_i(s/n^{1/2})/2) \nabla G_i(s/n^{1/2}).$$

Using (A.3) and (2.8),

$$(3.31) \quad \|\nabla v(s)\| \leq \frac{c_3}{4n^{1/2}} \|s\|^2.$$

Combining (3.29), (3.31), and (3.24),

$$(3.32) \quad I_r \leq \frac{c_3^2 I_0}{8\lambda_*^3 n} d(d+4).$$

It remains to estimate I_0 . For $0 \leq \beta \leq 1$, set

$$(3.33) \quad Z(\beta) := \int_{\tilde{\mathcal{U}}} e^{-\beta q(s)} \gamma(ds), \quad \nu_\beta(ds) := Z(\beta)^{-1} e^{-\beta q(s)} 1_{\tilde{\mathcal{U}}}(s) \gamma(ds).$$

Thus $Z(0) = \gamma(\tilde{\mathcal{U}})$ and $Z(1) = I_0$. Differentiating under the integral,

$$(3.34) \quad (\log Z)'(\beta) = -\mathbb{E}_{\nu_\beta} q, \quad (\log Z)''(\beta) = \text{Var}_{\nu_\beta}(q).$$

Taylor's formula with integral remainder gives the identity

$$(3.35) \quad \log(I_0/\gamma(\tilde{\mathcal{U}})) = -\mathbb{E}_{\nu_0} q + \int_0^1 (1-\beta) \text{Var}_{\nu_\beta}(q) d\beta.$$

The first term is estimated directly. Since $\nabla^l r(0) = 0$, $0 \leq l \leq 3$, Taylor's formula gives

$$(3.36) \quad r(t) = \frac{1}{6} \int_0^1 (1-u)^3 \nabla^4 G_r(ut)[t, t, t, t] du.$$

Since ν_0 is the standard Gaussian measure conditioned on the centered ball $\tilde{\mathcal{U}}$, conditioning can only decrease the fourth radial moment. Hence from (3.19) and (3.36)

$$(3.37) \quad |\mathbb{E}_{\nu_0} q| \leq \frac{c_4}{24n} \mathbb{E}_{\nu_0} \|s\|^4 \leq \frac{c_4}{24n} \mathbb{E}_\gamma \|s\|^4 \leq \frac{c_4}{24n} d(d+2) = A_0.$$

For the variance term in (3.35), the potential of ν_β is

$$(3.38) \quad V_\beta(s) = \frac{1}{2}\|s\|^2 + \beta q(s).$$

Since $q(s) = nG_r(s/n^{1/2}) - \frac{1}{2}\|s\|^2$,

$$(3.39) \quad \nabla^2 V_\beta(s) = (1-\beta)I_d + \beta \nabla^2 G_r(s/n^{1/2}) \succeq \lambda_* I_d;$$

see (3.22). Also, Taylor's formula applied to ∇r gives

$$(3.40) \quad \|\nabla r(t)\| \leq \frac{c_4}{6} \|t\|^3,$$

and hence

$$(3.41) \quad \|\nabla q(s)\| \leq \frac{c_4}{6n} \|s\|^3.$$

Applying (3.23) to q under ν_β and using (3.25),

$$(3.42) \quad \text{Var}_{\nu_\beta}(q) \leq \frac{c_4^2}{36\lambda_* n^2} \mathbb{E}_{\nu_\beta} \|s\|^6 \leq \frac{c_4^2}{36\lambda_*^4 n^2} d(d+4)(d+9).$$

Substituting (3.37) and (3.42) into (3.35),

$$(3.43) \quad |\log(I_0/\gamma(\tilde{\mathcal{U}}))| \leq A_0 + B_0 = D_0.$$

Consequently, $I_0 \leq e^{D_0}$ and

$$(3.44) \quad |I_0 - 1| \leq (1 - \gamma(\tilde{\mathcal{U}})) + D_0 e^{D_0}.$$

Combining (3.32) and (3.44), and using $I_0 \leq e^{D_0}$,

$$(3.45) \quad |I_{\mathcal{U}} - 1| \leq (1 - \gamma(\tilde{\mathcal{U}})) + e^{D_0} \left(D_0 + \frac{c_3^2}{8\lambda_*^3} \frac{d(d+4)}{n} \right).$$

The standard Gaussian tail bound used above gives $1 - \gamma(\tilde{\mathcal{U}}) \leq \exp\{-(R-1)^2 d/2\}$. Since $R = 2.5$, this term is absorbed by $\exp(-d)$. Combining (3.45) with the estimates for I_1 and I_2 proves Theorem 3.1.

4. CONSEQUENCES AND REFINEMENTS

4.1. A Monte Carlo refinement in the central region. The bound in Theorem 3.1 is uniform in $a \in U_d$. For numerical purposes, a sharper bound can be obtained if a is restricted to a central region. We describe this observation in centered coordinates, so that $\mathbb{E}X = 0$ and hence $\tau(0) = (\nabla\varphi)^{-1}(0) = 0$. The noncentered case follows by translation, with the central region centered at $\mathbb{E}X$.

In this subsection we suppose that (2.9) and (2.11) hold. Let $V_d = \{\|\tau\| < r'\}$, with $r' > 0$ chosen as in Lemma 2.2, and suppose also that (2.6) holds on this V_d .

Set $\Sigma := H(0) = \text{Var}(X)$. Under the Gaussian approximation, the covariance of the sample mean is Σ/n , and the typical standardized radius is of order $\sqrt{d}$. Accordingly, for fixed $L > 0$ we take the central region to be

$$(4.1) \quad \|\Sigma^{-1/2}a\| \leq L(d/n)^{1/2}.$$

Under the uniform spectral bounds (2.12), this is equivalent, up to changing L , to $\|a\| \leq L(d/n)^{1/2}$. For every fixed $L > 0$, these same bounds imply that, for d/n sufficiently small, the central region (4.1) is contained in U_d . Moreover, $\|\tau(a)\| \lesssim_L (d/n)^{1/2}$ uniformly on this region.

Restore temporarily the dependence on a in the notation of Subsection 3.2. For $\tau = \tau(a)$, set

$$(4.2) \quad J(a) := \int_{\tilde{\mathcal{U}}} e^{-q_\tau(s)} \cos \theta_\tau(s) \gamma(ds), \quad q_\tau(s) := nG_{r,a}(s/\sqrt{n}) - \frac{1}{2}\|s\|^2, \quad \theta_\tau(s) := nG_{i,a}(s/\sqrt{n}).$$

Thus $J(a) = I_0(a) - I_r(a)$. Since $I_i = 0$, the tail estimates in Subsection 3.1 give $I(a) = J(a) + \text{e.s.t.}$.

Fix any $L > 0$. Since $G_{r,a}(t) - \|t\|^2/2$ is even and begins at fourth order in t , while $G_{i,a}(t)$ is odd and begins at third order, Taylor expansion in τ about $\tau = 0$ gives

$$(4.3) \quad q_\tau(s) = q_0(s) + R_q(s), \quad \theta_\tau(s) = \theta_0(s) + R_\theta(s),$$

where, uniformly for $s \in \tilde{\mathcal{U}}$,

$$(4.4) \quad |R_q(s)| \lesssim_L \frac{\sqrt{d}\|s\|^4}{n^{3/2}}, \quad |R_\theta(s)| \lesssim_L \frac{\sqrt{d}\|s\|^3}{n}.$$

Both bounds tend to zero uniformly on $\tilde{\mathcal{U}}$ provided that $d^2/n \rightarrow 0$. The first estimate uses derivatives through order five, whereas the second requires only derivatives through order four (cf. the paragraph following (2.12)).

Substituting (4.3) into (4.2) gives

$$(4.5) \quad J(a) = \int_{\tilde{\mathcal{U}}} e^{-q_0(s)} e^{-R_q(s)} \cos(\theta_0(s) + R_\theta(s)) \gamma(ds).$$

Since d^2/n is sufficiently small, q_0 and R_q are uniformly bounded on $\tilde{\mathcal{U}}$. Hence

$$(4.6) \quad e^{-R_q} = 1 - R_q + O(R_q^2), \quad \cos(\theta_0 + R_\theta) = \cos \theta_0 - R_\theta \sin \theta_0 + O(R_\theta^2),$$

and therefore

$$(4.7) \quad J(a) - J(0) = - \int_{\tilde{\mathcal{U}}} e^{-q_0} R_q \cos \theta_0 d\gamma - \int_{\tilde{\mathcal{U}}} e^{-q_0} R_\theta \sin \theta_0 d\gamma + O \left(\int_{\tilde{\mathcal{U}}} (R_q^2 + R_\theta^2) d\gamma \right).$$

The first integral is bounded directly using (4.4):

$$(4.8) \quad \int_{\tilde{\mathcal{U}}} |R_q| d\gamma \lesssim_L \frac{\sqrt{d} \mathbb{E} \|S\|^4}{n^{3/2}} \lesssim_L \frac{d^{5/2}}{n^{3/2}}, \quad S \sim \mathcal{N}(0, I_d).$$

For the second integral we use the fact that both θ_0 and R_θ are odd. Moreover,

$$(4.9) \quad \|\nabla_s \theta_0(s)\| \lesssim \frac{\|s\|^2}{\sqrt{n}}, \quad \|\nabla_s R_\theta(s)\| \lesssim_L \frac{\sqrt{d} \|s\|^2}{n}.$$

Applying (3.23) to the Gaussian measure conditioned on the symmetric ball $\tilde{\mathcal{U}}$ therefore gives

$$(4.10) \quad \|\theta_0\|_{L^2(\gamma_{\tilde{\mathcal{U}}})} \lesssim \frac{d}{\sqrt{n}}, \quad \|R_\theta\|_{L^2(\gamma_{\tilde{\mathcal{U}}})} \lesssim_L \frac{d^{3/2}}{n}.$$

Consequently, since $|\sin x| \leq |x|$,

$$(4.11) \quad \int_{\tilde{\mathcal{U}}} |R_\theta \sin \theta_0| d\gamma \lesssim \left[\int_{\tilde{\mathcal{U}}} R_\theta^2 d\gamma \int_{\tilde{\mathcal{U}}} \theta_0^2 d\gamma \right]^{1/2} \lesssim_L \frac{d^{5/2}}{n^{3/2}}.$$

Finally,

$$(4.12) \quad \int_{\tilde{\mathcal{U}}} R_\theta^2 d\gamma \lesssim_L \frac{d^3}{n^2}, \quad \int_{\tilde{\mathcal{U}}} R_q^2 d\gamma \lesssim_L \frac{d^5}{n^3},$$

which are smaller than $d^{5/2}/n^{3/2}$ when d^2/n is sufficiently small. Thus (4.7) yields

$$(4.13) \quad J(a) = J(0) + O_L(d^{5/2}/n^{3/2}), \quad \|\Sigma^{-1/2}a\| \leq L(d/n)^{1/2}.$$

Hence

$$(4.14) \quad I(a) = J(0) + O_L(d^{5/2}/n^{3/2}) + \text{e.s.t.}, \quad \|\Sigma^{-1/2}a\| \leq L(d/n)^{1/2}.$$

This observation has two useful consequences. First, the ordinary saddlepoint approximation amounts to replacing $I(a)$ by 1. Therefore,

$$(4.15) \quad I(a) - 1 = J(0) - 1 + O_L(d^{5/2}/n^{3/2}) + \text{e.s.t.}, \quad \|\Sigma^{-1/2}a\| \leq L(d/n)^{1/2}.$$

Thus the quantity $J(0) - 1$ describes the leading SPA error throughout the central region with accuracy

$$(4.16) \quad O_L \left(\frac{d^{5/2}}{n^{3/2}} \right) = O_L \left(\frac{d^2}{n} \sqrt{\frac{d}{n}} \right),$$

which is smaller by a factor of $(d/n)^{1/2}$ than the generic $O(d^2/n)$ SPA error scale.

Second, either $I(0)$ or $J(0)$ can be used as a multiplicative correction factor. Since $I(0) = J(0) + \text{e.s.t.}$, (4.14) gives

$$(4.17) \quad I(a) = I(0) + O_L(d^{5/2}/n^{3/2}) + \text{e.s.t.}$$

uniformly for $\|\Sigma^{-1/2}a\| \leq L(d/n)^{1/2}$.

Since $I(0)$, and hence also $J(0)$, is bounded away from zero for sufficiently small d^2/n by Theorem 3.1, the preceding argument proves the following result.

Theorem 4.1 (Improved saddlepoint approximation). *Suppose that (2.9) and (2.11) hold. Let $V_d = \{\|\tau\| < r'\}$, with $r' > 0$ chosen as in Lemma 2.2, and suppose that (2.6) holds on this V_d . Suppose also that $\mathbb{E}X = 0$. Let K_0 be either $I(0)$ or $J(0)$. Then, for every fixed $L > 0$, provided d^2/n is sufficiently small,*

$$(4.18) \quad \sup_{\|a\| \leq L(d/n)^{1/2}} \left| \frac{I(a)}{K_0} - 1 \right| \lesssim_L \frac{d^{5/2}}{n^{3/2}} + \text{e.s.t.}.$$

Consequently, if

$$(4.19) \quad \rho_n^{\text{imp}}(a) := \frac{C_{d,n} K_0}{\sqrt{\det H(\tau)}} e^{-n\varphi^*(a)}, \quad \tau = (\nabla\varphi)^{-1}(a),$$

then

$$(4.20) \quad \sup_{\|a\| \leq L(d/n)^{1/2}} \left| \frac{\rho_n(a)}{\rho_n^{\text{imp}}(a)} - 1 \right| \lesssim_L \frac{d^{5/2}}{n^{3/2}} + \text{e.s.t.}.$$

The representation (4.2) is especially convenient for Monte Carlo computation, since the reference measure is standard Gaussian. Let $S_1, \dots, S_M$ be independent $\mathcal{N}(0, I_d)$ random vectors and define

$$(4.21) \quad \hat{J}_M(0) := \frac{1}{M} \sum_{k=1}^M \chi_{\tilde{\mathcal{U}}}(S_k) e^{-q_0(S_k)} \cos \theta_0(S_k).$$

Here $\chi_{\tilde{\mathcal{U}}}$ is the characteristic function of the set $\tilde{\mathcal{U}}$. Then $\mathbb{E}\hat{J}_M(0) = J(0)$. Moreover, set

$$(4.22) \quad B_{d,n} := \exp \left\{ \frac{c_4 R^4 d^2}{24 n} \right\}.$$

By the bound on q_0 , every summand in (4.21) is bounded in absolute value by $B_{d,n}$. Hoeffding's inequality [33, Theorem 1], [36, Theorem 2.2.6] therefore gives, with probability at least $1 - \eta$ over the Monte Carlo sample,

$$(4.23) \quad |\hat{J}_M(0) - J(0)| \leq B_{d,n} \sqrt{\frac{2 \log(2/\eta)}{M}}.$$

Combining this with (4.14), we obtain with the same probability

$$(4.24) \quad \sup_{\|\Sigma^{-1/2}a\| \leq L(d/n)^{1/2}} |I(a) - \hat{J}_M(0)| \lesssim_L \frac{d^{5/2}}{n^{3/2}} + B_{d,n} \sqrt{\frac{2 \log(2/\eta)}{M}} + \text{e.s.t.}.$$

When $c_4 \lesssim 1$ and d^2/n is sufficiently small, $B_{d,n} \lesssim 1$. Thus the Monte Carlo error has the usual $M^{-1/2}$ rate with a dimension-independent prefactor. The quantity $\hat{J}_M(0)$ is computed only once for each fixed model, d , and n , and can then be used throughout the entire central region $\|\Sigma^{-1/2}a\| \leq L(d/n)^{1/2}$.

4.2. Local limit theorem and central expansion. For convenience, in this subsection we suppose that X is isotropic, i.e., $\mathbb{E}X = 0$ and $\mathbb{E}(XX^\top) = I_d$. A standard result on the Legendre transform (which we derive for convenience in Appendix A.2) implies

$$(4.25) \quad \left| \varphi^*(a) - \frac{1}{2}\|a\|^2 \right| \lesssim C_3(a)\|a\|^3,$$

where φ^* is defined in (2.22) and $a \in U_d$ satisfies

$$(4.26) \quad 4\|a\|C_3(a) \leq 1, \quad C_3(a) := \sup_{\|\tau\| \leq 2\|a\|} \|\nabla^3 \varphi(\tau)\|_{\text{op}}.$$

We will only consider those a for which all the τ appearing in the supremum above belong to V_d . By (2.22), the following local CLT is a corollary to Theorem 3.1.

Corollary 4.2. *Suppose X is isotropic and the assumptions of Theorem 3.1 hold. For any $x \in \mathbb{R}^d$ and $n \in \mathbb{N}$ such that*

- (1) $\{\tau \in \mathbb{R}^d : \|\tau\| \leq 2\|x\|/n^{1/2}\} \subset V_d$, and

(2) (4.26) is satisfied with $a = x/n^{1/2} \in U_d$,
one has

$$(4.27) \quad \begin{aligned} \rho_n(a) da &= I(a) e^{-\mu(a)} \gamma(dx), \quad a = x/n^{1/2}, \\ \mu(a) &= n(\varphi^*(a) - \tfrac{1}{2}\|a\|^2) + \tfrac{1}{2} \log \det H(\tau), \end{aligned}$$

where $|I(a) - 1|$ is bounded by (3.2), and

$$(4.28) \quad |\mu(a)| \lesssim C_3(a) \frac{\|x\|^3 + d\|x\|}{n^{1/2}}.$$

Proof. In view of (4.25), it only remains to show that the determinant in (2.22) gives the second term in the fraction in (4.28). Since $H(0) = I_d$, we have

$$(4.29) \quad H(\tau) - I_d = \int_0^1 \nabla^3 \varphi(s\tau) [\tau, \cdot, \cdot] ds,$$

and hence, using $\|\tau\| \leq 2\|a\|$,

$$(4.30) \quad \|H(\tau) - I_d\| \leq C_3(a) \|\tau\| \leq 2C_3(a) \|a\| \leq \tfrac{1}{2}.$$

Thus, all the eigenvalues of $H(\tau)$ are bounded away from zero and

$$(4.31) \quad |\log \det H(\tau)| \lesssim d \|H(\tau) - I_d\| \lesssim d C_3(a) \|a\|.$$

Setting $a = x/\sqrt{n}$ completes the proof. $\square$

The leading central approximation to the density is, of course, the Gaussian approximation furnished by the CLT. A natural next step is to retain the first non-Gaussian terms in the central expansion of the saddlepoint representation (4.27). These terms are determined by the third cumulant tensor and give a cubic correction to the Gaussian approximation, as shown in the following corollary.

Corollary 4.3. *Suppose X is isotropic and (2.9) holds for some $r > 0$ independent of d . Suppose also that Assumption 2.1(3) holds with $V_d = \{\|\tau\| < r\}$. Let $T_3 := \nabla^3 \varphi(0)$, f_n denote the density of*

$$(4.32) \quad S_n := \frac{1}{\sqrt{n}} \sum_{j=1}^n X_j,$$

and let γ_d denote the standard Gaussian density on $\mathbb{R}^d$. Then, for every fixed $L > 0$, uniformly for $\|x\| \leq L\sqrt{d}$, one has

$$(4.33) \quad f_n(x) = \gamma_d(x) \exp \left\{ \frac{1}{6\sqrt{n}} T_3[x, x, x] - \frac{1}{2\sqrt{n}} \operatorname{tr} T_3[x, \cdot, \cdot] \right\} (1 + O_L(d^2/n) + \text{e.s.t.}),$$

provided d^2/n is sufficiently small. Here

$$(4.34) \quad \operatorname{tr} T_3[x, \cdot, \cdot] := \sum_{j=1}^d T_3[x, e_j, e_j].$$

The proof of the corollary is straightforward and is given in Appendix A.3.

4.3. More general SPA theorem. Let Y_n , $n \geq 1$, be a sequence of random vectors in $\mathbb{R}^d$, where d may grow with n . Denote by

$$(4.35) \quad \phi_n(\tau) = (\mathbb{E} e^{n\langle \tau, Y_n \rangle})^{1/n}, \quad \varphi_n(\tau) = \log \phi_n(\tau).$$

For complex arguments, the powers in (4.35) are understood via the analytic branch obtained by continuation from the positive real axis. The random vector Y_n is an analog of the average $n^{-1} \sum_{i=1}^n X_i$ in Section 2. We suppose that $\phi_n(t)$, $n \geq 1$, satisfy the following assumption.

Assumption 4.4. For all n sufficiently large,

- (1) $\sup_{\tau \in V_d} \|\nabla^k \phi_n(\tau)\|_{\text{op}} < \infty$, $0 \leq k \leq 4$, for some convex neighborhood of the origin $0 \in V_d \subset \mathbb{R}^d$.
- (2) $H_n(\tau) := \nabla^2 \varphi_n(\tau) > 0$ for all $\tau \in V_d$ and $\sup_{\tau \in V_d} \|H_n^{-1}(\tau)\| < \infty$.

(3) One has

$$(4.36) \quad \left| \frac{\phi_n(\tau + iH_n^{-1/2}t)}{\phi_n(\tau)} \right| \leq \max \left(\exp \{ -\epsilon^{1/2} \|t\| \}, (1 + \|t\|)^{-\frac{\kappa}{d^{1/2}}} \right),$$

$$\tau \in V_d, \quad \|t\| \geq 2.5\epsilon^{1/2},$$

for some $\kappa > 0$, where $H_n = H_n(\tau)$.

(4) There exists $\delta > 0$ such that

$$(4.37) \quad |\text{Arg } \phi_n(\tau + iH_n^{-1/2}t)| \leq \pi - \delta, \quad |\phi_n(\tau + iH_n^{-1/2}t)| \geq \delta,$$

for all $\tau \in V_d, \|t\| \leq 2.5\epsilon^{1/2}$.

The constants κ and δ are independent of d and n , and the size of V_d may depend on d . Similarly to (2.8), we define

$$(4.38) \quad c_3 := \sup_{\tau \in V_d, \|t\| < 2.5\epsilon^{1/2}} \|\nabla_t^3 \text{Im } \varphi_n(\tau + iH_n^{-1/2}t)\|_{\text{op}},$$

$$c_4 := \sup_{\tau \in V_d, \|t\| < 2.5\epsilon^{1/2}} \|\nabla_t^4 \text{Re } \varphi_n(\tau + iH_n^{-1/2}t)\|_{\text{op}}.$$

Assumption 4.4 implies that the constants c_k are finite, but they may grow with d and n .

The proof of the following result is identical to that of Theorem 3.1, upon replacing ϕ , φ , and H with ϕ_n , φ_n , and H_n , respectively. The quantities λ_* , A_0 , B_0 , D_0 are defined as in (3.1), where c_4 is given in (4.38).

Theorem 4.5. *Suppose Assumption 4.4 is satisfied and $\lambda_* > 0$. For all d^2/n sufficiently small, one has*

$$(4.39) \quad |I(a) - 1| \leq e^{D_0} \left(D_0 + \frac{c_3^2}{8\lambda_*^3} \frac{d(d+4)}{n} \right) + c \left[\exp(-d) + \left(\frac{e}{\kappa^2} \frac{d^2}{n} \right)^{d/2} \right], \quad a \in U_d,$$

for an absolute constant $c > 0$.

Theorem 4.5 generalizes Theorem 3.1 in that it imposes no specific form on Y_n ; in particular, Y_n need not be the average of n i.i.d. random vectors. A common case is when Y_n is the average of n independent, but not necessarily identically distributed random vectors.

5. MULTIVARIATE GENERALIZED LAPLACE DISTRIBUTIONS

We consider the multivariate generalized Laplace distribution [30, eq. (4)]. Its pdf is

$$(5.1) \quad \rho(x) = \frac{2 \exp\{\langle \Sigma^{-1}b, x \rangle\}}{(2\pi)^{d/2} \Gamma(s) (\det \Sigma)^{1/2}} \left(\frac{\langle x, \Sigma^{-1}x \rangle}{2 + \langle b, \Sigma^{-1}b \rangle} \right)^{\frac{1}{2}(s-d/2)} K_{s-d/2} \left(\sqrt{(2 + \langle b, \Sigma^{-1}b \rangle) \langle x, \Sigma^{-1}x \rangle} \right),$$

where $b \in \mathbb{R}^d$, $\Sigma > 0$, and K_ν is the modified Bessel function of the second kind. Its mgf and cgf are [30, eq. (3)]

$$(5.2) \quad \phi(z) = \left(1 - \langle b, z \rangle - \frac{1}{2} \langle z, \Sigma z \rangle \right)^{-s},$$

$$(5.3) \quad \varphi(z) = -s \log \left(1 - \langle b, z \rangle - \frac{1}{2} \langle z, \Sigma z \rangle \right).$$

We suppose that, uniformly in d ,

$$(5.4) \quad \|b\| \lesssim 1, \quad 0 < \sigma_- \leq \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) \leq \sigma_+ < \infty.$$

For simplicity, in what follows we assume that $s > 0$ is fixed.

The density has a singularity at the origin once $d > 2s$. Indeed, setting $q(x) := \langle x, \Sigma^{-1}x \rangle$, and using $K_\nu(t) = K_{-\nu}(t)$ together with

$$(5.5) \quad K_\nu(t) \sim 2^{\nu-1} \Gamma(\nu) t^{-\nu}, \quad t \rightarrow 0^+, \quad \nu > 0,$$

we obtain, for $d > 2s$,

$$(5.6) \quad \rho(x) \sim q(x)^{s-d/2} \asymp \|x\|^{2s-d}, \quad x \rightarrow 0.$$

Thus $\rho(x) \rightarrow \infty$ as $x \rightarrow 0$. In particular, in anticipation of Section 7, the generalized Laplace densities considered here are not LC whenever $d > 2s$, since a full-dimensional LC probability density is bounded.

Choose $r > 0$ sufficiently small and set

$$(5.7) \quad V_d = \{\tau \in \mathbb{R}^d : \|\tau\| < r\}.$$

Shrinking r if needed, we have

$$(5.8) \quad A(\tau) := 1 - \langle b, \tau \rangle - \frac{1}{2} \langle \tau, \Sigma \tau \rangle \asymp 1, \quad \tau \in V_d,$$

with constants independent of d .

Differentiating (5.3) gives

$$(5.9) \quad \nabla \varphi(\tau) = s \frac{b + \Sigma \tau}{A(\tau)}, \quad H(\tau) = s \left\{ \frac{\Sigma}{A(\tau)} + \frac{(b + \Sigma \tau)(b + \Sigma \tau)^\top}{A(\tau)^2} \right\}.$$

It follows from (5.4)–(5.9) that

$$(5.10) \quad 0 < \lambda_- \leq \lambda_{\min}(H(\tau)) \leq \lambda_{\max}(H(\tau)) \leq \lambda_+ < \infty, \quad \tau \in V_d,$$

uniformly in d . Thus Assumption 2.1(2) is satisfied. Assumption 2.1(1) follows immediately from (5.2) after reducing r , if necessary.

We next verify Assumption 2.1(3). For $u \in \mathbb{R}^d$,

$$(5.11) \quad A(\tau + iu) = A(\tau) + \frac{1}{2} \langle u, \Sigma u \rangle - i \langle b + \Sigma \tau, u \rangle.$$

Hence

$$(5.12) \quad \left| \frac{\phi(\tau + iu)}{\phi(\tau)} \right| = \left[\frac{A(\tau)}{\left[(A(\tau) + \frac{1}{2} \langle u, \Sigma u \rangle)^2 + \langle b + \Sigma \tau, u \rangle^2 \right]^{1/2}} \right]^s \leq \left(1 + \frac{\langle u, \Sigma u \rangle}{2A(\tau)} \right)^{-s}.$$

Take $u = H(\tau)^{-1/2}t$. By (5.4), (5.8), and (5.10),

$$(5.13) \quad \langle H(\tau)^{-1/2}t, \Sigma H(\tau)^{-1/2}t \rangle \geq c\|t\|^2.$$

Therefore

$$(5.14) \quad \left| \frac{\phi(\tau + iH(\tau)^{-1/2}t)}{\phi(\tau)} \right| \leq (1 + c\|t\|^2)^{-s}.$$

For $\|t\|$ bounded away from zero, (5.14) is much stronger than the second bound in (2.6). For small t , the Hessian normalization gives as in (2.14), uniformly in τ ,

$$(5.15) \quad \log \left| \frac{\phi(\tau + iH(\tau)^{-1/2}t)}{\phi(\tau)} \right| = -\frac{1}{2}\|t\|^2 + O(\|t\|^4).$$

Consequently, for some $r_1 > 0$,

$$(5.16) \quad \left| \frac{\phi(\tau + iH(\tau)^{-1/2}t)}{\phi(\tau)} \right| \leq e^{-0.4\|t\|^2}, \quad \|t\| \leq r_1.$$

If $2.5\epsilon^{1/2} \leq \|t\| \leq r_1$, then

$$(5.17) \quad 0.4\|t\|^2 \geq \epsilon^{1/2}\|t\|.$$

Combining this observation with (5.14) verifies Assumption 2.1(3) for all ϵ sufficiently small.

The quantities c_3 and c_4 are uniformly bounded as well. Indeed, all derivatives of order three and four of (5.3) are finite sums of terms involving Σ , $b + \Sigma z$, and negative powers of $A(z)$. In the region

$$(5.18) \quad z = \tau + iH(\tau)^{-1/2}t, \quad \tau \in V_d, \quad \|t\| \leq 2.5\epsilon^{1/2},$$

the quantities $\|b + \Sigma z\|$, $\|\Sigma\|$, and $|A(z)|^{-1}$ are uniformly bounded. Thus, using (5.10),

$$(5.19) \quad c_3 + c_4 \lesssim 1.$$

Finally, since $A(\tau + iu)$ remains in a fixed neighborhood of the positive real axis when $\|u\| \lesssim \epsilon^{1/2}$, the nonvanishing and phase requirements in Assumption 2.1(4) follow for all ϵ sufficiently small.

We have therefore proved the following.

Proposition 5.1. *Suppose (5.4) holds and let X have the mgf (5.2). Then, setting $V_d = \{\|\tau\| < r\}$ for a sufficiently small fixed $r > 0$, Assumption 2.1 holds. Moreover, c_3, c_4 are bounded independently of d and n and, provided d^2/n is sufficiently small,*

$$(5.20) \quad |I(a) - 1| \lesssim \frac{d^2}{n} + \text{e.s.t.}, \quad a \in \nabla\varphi(V_d).$$

Let us now compare (5.20) with the high-dimensional saddlepoint theorem of [35]. To map our notation to theirs, let

$$(5.21) \quad \mathcal{X}_n := \sum_{j=1}^n X_j, \quad p = d, \quad s_n = na.$$

Then the cgf of $\mathcal{X}_n$ is $K_{\mathcal{X}_n}(z) = n\varphi(z)$, and the saddlepoint $\hat{t}_n$ in [35] is precisely our τ , since

$$(5.22) \quad \nabla K_{\mathcal{X}_n}(\tau) = n\nabla\varphi(\tau) = na = s_n.$$

Writing, as in [35],

$$(5.23) \quad K_{\mathcal{X}_n}(x + iy) = U(x, y) + iV(x, y),$$

and setting

$$(5.24) \quad H_r(z) := \text{Re } \nabla^2\varphi(z),$$

we therefore have

$$(5.25) \quad U^{(x,2)}(\tau, y) = n \text{Re } \nabla^2\varphi(\tau + iy) = nH_r(\tau + iy).$$

Thus the Hessian appearing in [35, Assumption 8] is the same as $nH_r(z)$ in our notation.

The bounds (5.10) and (5.19), together with the same argument on a sufficiently small fixed complex neighborhood, show that the second, third, and fourth derivatives appearing in [35, Assumptions 8–10] have the required spectral bounds. Moreover, (5.14), after raising its right-hand side to the n th power, gives the required estimate of the outer Fourier integral in their Assumption 7.

For a matrix $A = (A_{ij})$, we denote by $\|A\|_\infty$ the induced ℓ_∞ operator norm,

$$(5.26) \quad \|A\|_\infty := \sup_{\|x\|_\infty=1} \|Ax\|_\infty = \max_{1 \leq i \leq d} \sum_{j=1}^d |A_{ij}|.$$

After reducing the fixed complex neighborhood if necessary, the argument above (5.10) gives uniformly $H_r(z) \asymp I_d$. Consequently,

$$(5.27) \quad \|(nH_r(z))^{-1/2}\|_\infty \leq d^{1/2} \|(nH_r(z))^{-1/2}\| \lesssim d^{1/2} n^{-1/2}.$$

The factor $d^{1/2}$ in (5.27) cannot in general be removed under (5.4). Indeed, for admissible choices of b and Σ one has

$$(5.28) \quad \|(nH_r(0))^{-1/2}\|_\infty \gtrsim (d/n)^{1/2};$$

see Appendix A.4. Thus, in the notation of [35], one should in general take $c_\infty = 1/2$.

Next, consider Assumptions 9–10 in [35]. To avoid confusion with the constants c_3, c_4 in (2.8), denote the exponents appearing in these assumptions by c_3^{TR} and c_4^{TR} . On the fixed complex neighborhood considered above, direct differentiation of (5.3) gives

$$(5.29) \quad \|\nabla^3\varphi(z)\|_{\text{op}} + \|\nabla^4\varphi(z)\|_{\text{op}} \lesssim 1.$$

Since $K_{\mathcal{X}_n}(z) = n\varphi(z)$,

$$(5.30) \quad \|\nabla^k K_{\mathcal{X}_n}(z)\|_{\text{op}} \lesssim n, \quad k = 3, 4.$$

Each matrix slice of either derivative tensor has spectral norm bounded by the operator norm of the corresponding tensor. Hence the matrix slices of $U^{(x,3)}$, $U^{(x,4)}$, and $V^{(x,4)}$ appearing in [35, Assumptions 9–10] all have spectral norm $\lesssim n$. Thus one may take $c_3^{\text{TR}} = c_4^{\text{TR}} = 1$. These are the smallest exponents that can be guaranteed uniformly under (5.4): admissible

choices of b and Σ give nonzero third- and fourth-order derivatives of φ of order one, and hence corresponding derivatives of $n \operatorname{Re} \varphi$ of order n .

Thus, with $c_\infty = 1/2$ and $c_3^{\text{TR}} = c_4^{\text{TR}} = 1$, the error in [35, Theorem 5.1] becomes

$$(5.31) \quad |I(a) - 1| \lesssim \max \left\{ \frac{d^4}{n}, \frac{d^3}{n} \right\} \lesssim \frac{d^4}{n}.$$

By contrast, (5.20) gives the error $O(d^2/n)$ under the same parameter bounds (5.4), and hence convergence under the substantially weaker condition $d^2/n \rightarrow 0$. If additional structure is imposed so that the ℓ_∞ Hessian bound in [35] holds with $c_\infty = 0$, their error improves to $O(d^3/n)$, which is still one power of d larger than (5.20).

6. A SUFFICIENT CRITERION

Let $\rho(x)$ be the density of X . For $\Theta \in S^{d-1}$ and $\tau \in V_d$, define the 1D tilted marginal

$$(6.1) \quad \rho_\Theta^\tau(s) := \frac{1}{\phi(\tau)} \int_{\mathbb{R}^d} e^{\langle \tau, x \rangle} \rho(x) \delta(\langle \Theta, x \rangle - s) dx.$$

Here $\phi(\tau)$ is the normalizing constant. Let X_Θ^τ be the random variable with the pdf ρ_Θ^τ . The quantile $m_\Theta^\tau(a)$ of the random variable $\rho_\Theta^\tau(X_\Theta^\tau)$ is any number that satisfies

$$(6.2) \quad \mathbb{P}\{\rho_\Theta^\tau(X_\Theta^\tau) < m_\Theta^\tau(a)\} \leq a, \quad \mathbb{P}\{\rho_\Theta^\tau(X_\Theta^\tau) > m_\Theta^\tau(a)\} \leq 1 - a, \quad 0 < a < 1.$$

Lemma 6.1. *Suppose (2.9) and (2.11) hold for some $r > 0$. Suppose further that*

$$(6.3) \quad \sup_{\|\tau\| \leq r} \sup_{\Theta \in S^{d-1}} \int_{\mathbb{R}} |\rho_\Theta^\tau(s+h) - \rho_\Theta^\tau(s)| ds \lesssim |h|^\alpha, \quad |h| \leq h_0,$$

for some $0 < \alpha \leq 1$ and $h_0 > 0$, with all constants independent of d and n . Then there exists $0 < r' \leq r$ such that, with $V_d = \{\|\tau\| < r'\}$, Assumption 2.1(3) holds on V_d for all $\epsilon = d/n$ sufficiently small.

Proof of Lemma 6.1. Choose $r' \in (0, r]$ sufficiently small so that (2.12) holds. From the assumptions of the lemma and (2.10) it follows that

$$(6.4) \quad \mathbb{E}|X_\Theta^\tau|^j \lesssim 1, \quad 1 \leq j \leq 4, \quad \|\tau\| < r', \quad \Theta \in S^{d-1}.$$

Next, consider quantiles of ρ_Θ^τ . By an elementary calculation, (6.3) yields

$$(6.5) \quad |\widetilde{\rho_\Theta^\tau}(\lambda)| \lesssim (1 + |\lambda|)^{-\alpha}, \quad \|\tau\| < r,$$

uniformly in τ . Suppose first that $0 < \alpha < 1$. The Besov embedding $B_{1,\infty}^\alpha(\mathbb{R}) \hookrightarrow L^{q,\infty}(\mathbb{R})$, where $q = 1/(1 - \alpha) > 1$ [19, Part (iii) of Theorem 1.15], implies that

$$(6.6) \quad L(m) := |\{s : \rho_\Theta^\tau(s) > m\}| \lesssim m^{-q}, \quad m \rightarrow \infty.$$

Applying this in (6.2) we get (with ρ_Θ^τ denoted by ρ for simplicity)

$$(6.7) \quad \mathbb{P}\{\rho(X) > m\} = \int_{\rho(s) > m} \rho(s) ds = mL(m) + \int_m^\infty L(u) du.$$

Using (6.6) gives an upper bound

$$(6.8) \quad \mathbb{P}\{\rho(X) > m\} \lesssim m^{1-q} + \int_m^\infty u^{-q} du \rightarrow 0, \quad m \rightarrow \infty.$$

Thus, $m_\Theta^\tau(1/2) \lesssim 1$ uniformly in $\Theta \in S^{d-1}$ and $\|\tau\| \leq r$.

If $\alpha = 1$, then (6.3) implies a uniform BV bound on ρ_Θ^τ , and hence $\|\rho_\Theta^\tau\|_\infty \lesssim 1$; in particular, $m_\Theta^\tau(1/2) \lesssim 1$ directly.

Comparing (6.1) with (2.6), we see that the ratio in (2.6) is precisely $|\widetilde{\rho_\Theta^\tau}(\lambda)|$, where Θ and λ satisfy $H(\tau)^{-1/2}t = \lambda\Theta$ and

$$(6.9) \quad \operatorname{Var}(X_\Theta^\tau)\lambda^2 = \|t\|^2.$$

This relation between λ and t is assumed in what follows.

To prove Assumption 2.1(3), begin by considering the interval $2.5\epsilon^{1/2} \leq \|t\| \leq t_0$ for a sufficiently small $t_0 > 0$. The function $|\widetilde{\rho_\Theta^\tau}|$ is even, because ρ_Θ^τ is real-valued. Using (6.4) we get

$$(6.10) \quad |\widetilde{\rho_\Theta^\tau}(\lambda)| \leq 1 - \frac{1}{2} \text{Var}(X_\Theta^\tau) \lambda^2 + O(\lambda^4), \quad \lambda \rightarrow 0.$$

By (2.12), $\text{Var}(X_\Theta^\tau) = \langle H(\tau)\Theta, \Theta \rangle \asymp 1$ and hence $\lambda \asymp \|t\|$. This gives the inequality

$$(6.11) \quad |\widetilde{\rho_\Theta^\tau}(\lambda)| \leq \exp\{-\epsilon^{1/2}\|t\|\}, \quad 2.5\epsilon^{1/2} \leq \|t\| \leq t_0, \quad \|\tau\| \leq r', \quad \Theta \in S^{d-1},$$

for sufficiently small $t_0 > 0$ and $r' > 0$.

Next, consider the set $\|t\| \geq t_0$. In view of (6.4) and the uniform bound on $m_\Theta^\tau(1/2)$, [9, Theorem 2] implies

$$(6.12) \quad |\widetilde{\rho_\Theta^\tau}(\lambda)| \leq \begin{cases} 1 - b_1 \lambda^2, & |\lambda| \leq b_3, \\ 1 - b_2, & |\lambda| \geq b_3, \end{cases}$$

for some $b_k > 0$, $k = 1, 2, 3$. Using the bound (6.5), we can find $t_1 \geq t_0$, independent of d and n , such that

$$(6.13) \quad |\widetilde{\rho_\Theta^\tau}(\lambda)| \lesssim (1 + |\lambda|)^{-\alpha} \leq (1 + \|t\|)^{-\kappa/d^{1/2}}, \quad \|t\| \geq t_1, \quad \|\tau\| \leq r', \quad \Theta \in S^{d-1},$$

for all d sufficiently large.

Finally, by (6.12), (2.6) holds for $t_0 \leq \|t\| \leq t_1$ for all ϵ sufficiently small and d sufficiently large. $\square$

7. STRUCTURED CLASSES OF DISTRIBUTIONS

We now verify the dimension-uniform conditions of Lemma 6.1 for several structured classes of distributions.

7.1. LC isotropic case. Suppose that X is isotropic and the pdf of X is $e^{-V(x)}$, where $V : \mathbb{R}^d \rightarrow (-\infty, \infty]$ is extended value convex. Allowing V to take the value $+\infty$ permits, in particular, compactly supported LC densities. Since X is isotropic, (2.11) trivially holds with absolute constants.

Proof of (2.9). Using that marginals of a LC isotropic random vector are LC isotropic, it suffices to prove (2.9) only in $\mathbb{R}$, i.e., for its one-dimensional marginal $e^{-V_1(x)}$, $x \in \mathbb{R}$, on the subspace spanned by τ .

In $\mathbb{R}$, [32, Theorem 5.14] implies that $V_1(0) \leq 7 \log 2$. Consider the sublevel set $D := \{x : V_1(x) \leq V_1(0) + 1\}$, which is an interval. The bound on $V_1(0)$ implies that $|D| \leq R$, where $|D|$ is the length of the interval and $R > 0$ is an absolute constant. Since V_1 is a convex function, this means that the rate of linear growth of V_1 outside D is bounded below by some absolute positive constant, say, μ . Therefore,

$$(7.1) \quad V_1(x) \geq V_1(0) + 1 + \mu \text{dist}(x, D), \quad x \notin D.$$

Selecting $\delta_2 < 0.9\mu$ gives

$$(7.2) \quad \int_{\mathbb{R}} e^{\tau x - V_1(x)} dx \lesssim 1, \quad |\tau| \leq \delta_2,$$

where the absorbed bound is absolute. This proves (2.9) with an absolute bound.

In (7.2) and similarly in the proof below, we tacitly identify $\tau \in \mathbb{R}^d$ with its projection to the appropriate one- or two-dimensional subspace.

Proof of (6.3). To prove (6.3) it suffices to consider the two-dimensional marginal $e^{-V_2(x)}$, $x \in \mathbb{R}^2$, on the subspace spanned by Θ and τ . If the two are linearly dependent, we can take any subspace containing them.

In $\mathbb{R}^2$, [32, Theorem 5.14 and Lemma 5.18] imply that $\|x_0\| \leq 3$ and $V_2(0) \leq 14 \log 2$, where x_0 is the location of the minimum of $V_2(x)$. Consider the sublevel set $D := \{x : V_2(x) \leq V_2(x_0) + 1\}$.

Let $s \in (0, 1]$ be its probability. Given that $V_2(x_0) \leq V_2(0)$, the bound on $V_2(0)$ implies that $s \geq c_1|D|$, where $|D|$ is the area of D and $c_1 > 0$ is an absolute constant.

By [32, Lemma 5.13], $e^{-V_2(x)}$ is $1/e$ -rounded, so D contains a ball of radius s/e . Pick any $x' \in D$, and let x_c be the center of the ball. Denote $R' = \|x_c - x'\|$. Clearly, $|D| \geq c_2 s R'$ for an absolute constant $c_2 > 0$. To see this, we consider the triangle formed by x' and the diameter of the ball orthogonal to the line through x_c and x' . The triangle is inside D , because D is convex. Thus, $s \geq c_1 c_2 s R'$, so R' is bounded by an absolute constant, i.e., the diameter of D is bounded by an absolute constant.

In turn, this implies that there exists an absolute constant R such that $D \subset \{\|x - x_0\| \leq R\}$. Since V_2 is a convex function, this means that the rate of linear growth of V_2 along any ray $x_0 + t\Theta$, $t \geq R$, is bounded below by some absolute positive constant, say, μ . Selecting $\delta_2 < 0.9\mu$ we get similarly to the 1D case:

$$(7.3) \quad e^{\langle \tau, x \rangle - V_2(x)} \leq c_1 e^{-c_2 \|x - x_0\|}, \quad x \in \mathbb{R}^2, \quad \|\tau\| \leq \delta_2,$$

for some absolute constants $c_1, c_2 > 0$. Furthermore, by the Jensen inequality applied to the convex function $\psi(x) = e^{\langle \tau, x \rangle}$, the normalization constant satisfies

$$(7.4) \quad \int_{\mathbb{R}^2} e^{\langle \tau, x \rangle - V_2(x)} dx = \mathbb{E} e^{\langle \tau, X \rangle} \geq e^{\langle \tau, \mathbb{E} X \rangle} = 1,$$

where we have used that $\mathbb{E} X = 0$. Integrating the bound in (7.3) along any line and using the last inequality proves that $\rho_{\Theta}^{\tau}(s)$ is bounded by an absolute constant.

Since the original d -dimensional density is LC, its pdf tilted by $e^{\langle \tau, X \rangle}$ is still LC, but it is no longer isotropic. Thus, ρ_{Θ}^{τ} is also LC (and not necessarily isotropic). Let s_0 be a location of its maximum (this point is not necessarily unique). The fact that ρ_{Θ}^{τ} is LC implies that it is nondecreasing on $(-\infty, s_0]$ and nonincreasing on $[s_0, \infty)$. Therefore,

$$(7.5) \quad \int_{\mathbb{R}} |\rho_{\Theta}^{\tau}(s+h) - \rho_{\Theta}^{\tau}(s)| ds \leq 2 \int_{|s-s_0| \leq |h|} \rho_{\Theta}^{\tau}(s) ds.$$

Using the uniform bound on ρ_{Θ}^{τ} obtained above, the right-hand side is $\lesssim |h|$. Hence (6.3) holds with $\alpha = 1$ (taking, for example, $h_0 = 1$).

Thus, any isotropic LC random variable X satisfies the assumptions of Lemma 6.1. We summarize the result.

Proposition 7.1. *Let $X = X_d$ be an isotropic random vector with a full-dimensional LC density on $\mathbb{R}^d$. There exists a universal $r_* > 0$ such that (2.9) and (6.3) hold with $r = r_*$ and $\alpha = 1$, and with $V_d = \{\|\tau\| < r_*\}$ one has*

$$(7.6) \quad |I(a) - 1| \lesssim \frac{d^2}{n} + \text{e.s.t.}, \quad a \in \nabla \varphi(V_d),$$

provided d^2/n is sufficiently small. Finally, c_3, c_4 are bounded by a universal constant.

Proof. Since X is isotropic, (2.11) clearly holds with absolute constants. The preceding argument verifies the remaining assumptions of Lemmas 2.2 and 6.1 with absolute constants (for example, taking $h_0 = 1$ in (6.3)). After decreasing the resulting universal radius if necessary, denote it by r_* . Lemmas 2.2 and 6.1, together with Theorem 3.1, then prove the proposition. $\square$

7.2. Affine images and convex mixtures of LC random vectors. The preceding result is stable under uniformly bounded affine transformations. This observation also gives a broad class of non-LC distributions to which Lemma 6.1 applies.

Lemma 7.2. *Let $X = X_d$ have a full-dimensional LC density on $\mathbb{R}^d$, and suppose that*

$$(7.7) \quad \|\mathbb{E} X\| \leq c_0, \quad c_1 I_d \preceq \text{Var}(X) \preceq c_2 I_d,$$

with $c_k > 0$ independent of d . Then, for $\alpha = 1$, there exist $r, h_0 > 0$ such that (2.9) and (6.3) hold. Consequently, the conclusions of Proposition 7.1 hold for X , with constants depending only on c_0, c_1, c_2 .

Proof. Set

$$(7.8) \quad b = \mathbb{E}X, \quad A = \text{Var}(X)^{1/2}, \quad Y = A^{-1}(X - b).$$

Clearly, Y is isotropic and LC. The argument preceding Proposition 7.1 shows that there are absolute constants $r_0, h_0^{(Y)} > 0$ such that

$$(7.9) \quad \sup_{\|\sigma\| \leq r_0} \mathbb{E}e^{\langle \sigma, Y \rangle} \lesssim 1,$$

and (6.3) holds for Y , uniformly in d .

Fix $\Theta \in S^{d-1}$ and set

$$(7.10) \quad r_\Theta := \|A^\top \Theta\|, \quad \tilde{\Theta} := \frac{A^\top \Theta}{r_\Theta}, \quad \sigma := A^\top \tau.$$

By (7.7),

$$(7.11) \quad c_1^{1/2} \leq r_\Theta \leq c_2^{1/2}, \quad \|\sigma\| \leq c_2^{1/2} \|\tau\|.$$

Moreover,

$$(7.12) \quad \langle \tau, X \rangle = \langle \sigma, Y \rangle + \langle \tau, b \rangle.$$

Therefore, after taking $r \leq r_0/c_2^{1/2}$ and decreasing it if necessary,

$$(7.13) \quad \mathbb{E}e^{\langle \tau, X \rangle} \lesssim e^{c_0 \|\tau\|} \sup_{\|\sigma\| \leq r_0} \mathbb{E}e^{\langle \sigma, Y \rangle} \lesssim 1, \quad \|\tau\| < r.$$

Thus (2.9) holds uniformly.

The assumption (2.11) is the same as the second half of (7.7). It remains to verify (6.3).

Write $\rho_{\Theta, X}^\tau$ and $\rho_{\Theta, Y}^\sigma$ for the tilted one-dimensional marginals defined in (6.1). Exponential tilting of X by τ corresponds to exponential tilting of Y by $\sigma = A^\top \tau$. Hence

$$(7.14) \quad \rho_{\Theta, X}^\tau(s) = \frac{1}{r_\Theta} \rho_{\Theta, Y}^\sigma\left(\frac{s - \langle \Theta, b \rangle}{r_\Theta}\right).$$

Consequently,

$$(7.15) \quad \int_{\mathbb{R}} |\rho_{\Theta, X}^\tau(s+h) - \rho_{\Theta, X}^\tau(s)| ds = \int_{\mathbb{R}} |\rho_{\Theta, Y}^\sigma(u + \frac{h}{r_\Theta}) - \rho_{\Theta, Y}^\sigma(u)| du \lesssim \left| \frac{h}{r_\Theta} \right| \lesssim |h|,$$

uniformly for $\|\tau\| \leq r$ and $|h|$ sufficiently small. This proves (6.3), where we can take $h_0 = c_1^{1/2} h_0^{(Y)}$. Here we use that $r_\Theta \geq c_1^{1/2}$; see (7.11).

Lemmas 2.2 and 6.1, together with Theorem 3.1, then prove the lemma. $\square$

We can now pass from individual LC densities to their convex combinations. The components need not have the same mean or covariance matrix.

Proposition 7.3. *For $1 \leq j \leq J$, let $X_j = X_{j,d}$ have LC density ρ_j and satisfy*

$$(7.16) \quad \|\mathbb{E}X_j\| \leq c_0, \quad c_1 I_d \preceq \text{Var}(X_j) \preceq c_2 I_d,$$

for some constants $c_k > 0$ independent of j and d . Let X have density

$$(7.17) \quad \rho(x) = \sum_{j=1}^J p_j \rho_j(x), \quad p_j \geq 0, \quad \sum_{j=1}^J p_j = 1.$$

Then (2.11) holds. Moreover, for $\alpha = 1$, there exist $r, h_0 > 0$ such that (2.9) and (6.3) hold. Consequently, the conclusions of Proposition 7.1 hold for X , with constants depending only on c_0, c_1, c_2 , and, in particular, independent of d, J , and the weights p_j .

Proof. Clearly, X need not be LC, so we verify that it satisfies the assumptions of Lemma 6.1. Denote $b_j = \mathbb{E}X_j$. Let

$$(7.18) \quad b := \mathbb{E}X = \sum_{j=1}^J p_j b_j, \quad \Sigma := \text{Var}(X) = \sum_{j=1}^J p_j [\text{Var}(X_j) + (b_j - b)(b_j - b)^\top].$$

Thus $\|b\| \leq c_0$ and

$$(7.19) \quad c_1 I_d \preceq \Sigma \preceq (c_2 + c_0^2) I_d.$$

Indeed,

$$(7.20) \quad \sum_{j=1}^J p_j (b_j - b)(b_j - b)^\top = \sum_{j=1}^J p_j b_j b_j^\top - b b^\top \preceq \sum_{j=1}^J p_j b_j b_j^\top \preceq c_0^2 I_d,$$

which implies the second inequality in (7.19). Therefore, X satisfies (2.11).

By Lemma 7.2, X_j satisfy (2.9) and (6.3) uniformly in j, d . By linearity of expectation, X satisfies (2.9). It remains to check (6.3). We have

$$(7.21) \quad \begin{aligned} \rho_\Theta^\tau(s) &= \frac{1}{\phi(\tau)} \sum_{j=1}^J p_j \int_{\mathbb{R}^d} e^{\langle \tau, x \rangle} \rho_j(x) \delta(\langle \Theta, x \rangle - s) dx = \frac{1}{\phi(\tau)} \sum_{j=1}^J p_j \phi_j(\tau) (\rho_j)_\Theta^\tau(s), \\ \phi(\tau) &= \sum_{j=1}^J p_j \phi_j(\tau), \quad \phi_j(\tau) := \mathbb{E} e^{\langle \tau, X_j \rangle}. \end{aligned}$$

Therefore, the triangle inequality and individual bounds imply

$$(7.22) \quad \int_{\mathbb{R}} |\rho_\Theta^\tau(s+h) - \rho_\Theta^\tau(s)| ds \lesssim |h| \frac{1}{\phi(\tau)} \sum_{j=1}^J p_j \phi_j(\tau) = |h|.$$

Lemmas 2.2 and 6.1, together with Theorem 3.1, then prove the proposition. $\square$

The same proof also applies to general convex mixtures provided the bounds in (7.16) hold uniformly over the mixing parameter.

7.3. Gaussian-smoothed distributions. We next consider a broad class of distributions obtained by adding nondegenerate Gaussian noise to an otherwise arbitrary random vector. Let

$$(7.23) \quad X = Y + Z, \quad Z \sim \mathcal{N}(0, \Sigma), \quad Y \perp Z.$$

No density or full-dimensionality assumption is imposed on Y . The Gaussian component ensures that X has a smooth full-dimensional density.

Proposition 7.4. *Suppose that, for some $r > 0$ independent of d ,*

$$(7.24) \quad \sup_{\|\tau\| \leq r} \mathbb{E} e^{\langle \tau, Y \rangle} \lesssim 1, \quad \Sigma \asymp I_d.$$

Then (2.11) holds. Moreover, there exist $r', h_0 > 0$ such that (2.9) and (6.3) hold with $\alpha = 1$. Consequently, the conclusions of Proposition 7.1 hold for X , with constants depending only on r and the absorbed constants in (7.24).

Proof. Denote the mgf of Y by ϕ_Y . Independence in (7.23) gives

$$(7.25) \quad \phi(\tau) = \phi_Y(\tau) \exp \left\{ \frac{1}{2} \langle \Sigma \tau, \tau \rangle \right\}.$$

Therefore, (7.24) implies (2.9) for X .

Moreover, (7.24) and (2.10) give $\mathbb{E} \langle \Theta, Y \rangle^2 \lesssim 1$ uniformly in $\Theta \in S^{d-1}$. Thus $\text{Var}(Y) \lesssim I_d$, and hence

$$(7.26) \quad \text{Var}(X) = \text{Var}(Y) + \Sigma \asymp I_d.$$

This proves (2.11).

It remains to verify (6.3). For $\|\tau\| \leq r$, let Y^τ and Z^τ denote the exponentially tilted versions of Y and Z , respectively. Since Y and Z are independent, exponential tilting preserves their independence and

$$(7.27) \quad X^\tau \stackrel{d}{=} Y^\tau + Z^\tau, \quad Z^\tau \sim \mathcal{N}(\Sigma \tau, \Sigma).$$

Fix $\Theta \in S^{d-1}$ and denote by μ_Θ^τ the law of $\langle \Theta, Y^\tau \rangle$. Also, let g_Θ^τ be the pdf of $\langle \Theta, Z^\tau \rangle$. Then

$$(7.28) \quad \rho_\Theta^\tau(s) = \int_{\mathbb{R}} g_\Theta^\tau(s-u) \mu_\Theta^\tau(du), \quad s \in \mathbb{R}.$$

The Gaussian density g_Θ^τ has variance

$$(7.29) \quad \sigma_\Theta^2 = \langle \Sigma \Theta, \Theta \rangle \asymp 1.$$

Consequently, elementary transformations give

$$(7.30) \quad \begin{aligned} \int_{\mathbb{R}} |\rho_\Theta^\tau(s+h) - \rho_\Theta^\tau(s)| ds &\leq \int_{\mathbb{R}} |g_\Theta^\tau(s+h) - g_\Theta^\tau(s)| ds \\ &\leq |h| \int_{\mathbb{R}} |(g_\Theta^\tau)'(s)| ds \lesssim |h| \end{aligned}$$

uniformly in $\|\tau\| \leq r$ and $\Theta \in S^{d-1}$. Indeed,

$$(7.31) \quad \int_{\mathbb{R}} |(g_\Theta^\tau)'(s)| ds = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma_\Theta} \lesssim 1.$$

This proves (6.3).

Lemmas 2.2 and 6.1, together with Theorem 3.1, then prove the proposition. $\square$

A useful special case of Proposition 7.4 is the latent-factor model

$$(7.32) \quad X = A_d U + Z,$$

where $A_d : \mathbb{R}^{k_d} \rightarrow \mathbb{R}^d$ may have rank strictly smaller than d . Suppose $\|A_d\| \lesssim 1$ and, for some $r_0 > 0$,

$$(7.33) \quad \sup_{\|u\| \leq r_0} \mathbb{E} e^{\langle u, U \rangle} \lesssim 1$$

uniformly in d . Then, after taking $r > 0$ sufficiently small,

$$(7.34) \quad \sup_{\|\tau\| \leq r} \mathbb{E} e^{\langle \tau, A_d U \rangle} = \sup_{\|\tau\| \leq r} \mathbb{E} e^{\langle A_d^\top \tau, U \rangle} \lesssim 1,$$

so Proposition 7.4 applies. Thus the latent variable may have dimension $k_d < d$, and may even have a singular or discrete distribution; the uniformly nondegenerate Gaussian noise is sufficient to provide the regularity required by Lemma 6.1.

Remark 7.5. The Gaussian assumption on Z in Proposition 7.4 is not essential. More generally, Z may be any random vector satisfying (2.9), (2.11), and (6.3) uniformly in d . Indeed, the first two properties together with the exponential-moment bound on Y in (7.24) imply (2.9) and (2.11) for $X = Y + Z$ exactly as in the proof above, while the convolution representation (7.28) and (6.3) for Z imply (6.3) for X . In particular, Z may be any of the random vectors covered by Proposition 7.3.

8. APPLICATION: NON-GAUSSIAN NOISE IN AVERAGED IMAGES

We conclude with a numerical application of our results to likelihood-based inference from averaged images. The setup is loosely motivated by astronomical image stacking, where averaging many registered images can lead to an approximately Gaussian distribution by the central limit theorem; see, for example, [31]. We use this astronomical application as a natural high-dimensional test problem for the approximations developed above. In particular, the two-component Gaussian-mixture noise model below is chosen because it is non-Gaussian and nonseparable while still allowing the exact likelihood of the averaged image to be computed as a finite Gaussian mixture.

Suppose that the same image is observed repeatedly and that the images have been registered on a common pixel grid.

8.1. Noise model and exact likelihood. Let $\Lambda_d = \{1, \dots, m\}^2$, so that $d = m^2$, and let $m_\theta \in \mathbb{R}^d$ denote the underlying image on the patch. The indices are understood modulo m in both coordinate directions. We observe

$$(8.1) \quad Y_j = m_\theta + X_j, \quad j = 1, \dots, n,$$

where the X_j are i.i.d. realizations of a noise vector $X \in \mathbb{R}^d$.

To simulate spatially correlated noise we use the periodic five-point stencil

$$(8.2) \quad \begin{aligned} (G_\alpha z)(x) &= c_\alpha \left[z(x) + \alpha \sum_{e \in \{(1,0), (-1,0), (0,1), (0,-1)\}} z(x - e) \right], \\ c_\alpha &= (1 + 4\alpha^2)^{-1/2}, \quad x \in \Lambda_d, \end{aligned}$$

and set $C_\alpha := G_\alpha G_\alpha^\top$. The Fourier symbol of G_α is

$$(8.3) \quad c_\alpha \{1 + 2\alpha(\cos \omega_1 + \cos \omega_2)\}.$$

Thus, for $|\alpha| < 1/4$, $C_\alpha \asymp I_d$ uniformly in d .

Let $Z_j \sim \mathcal{N}(0, I_d)$, and let S_j be independent random variables taking the two values

$$(8.4) \quad s_- = -\frac{1}{\sqrt{3}}, \quad s_+ = \sqrt{3}, \quad \mathbb{P}\{S_j = s_-\} = \frac{3}{4}, \quad \mathbb{P}\{S_j = s_+\} = \frac{1}{4}.$$

Thus $\mathbb{E}S_j = 0$ and $\mathbb{E}S_j^2 = 1$. Fix a smooth spatial pattern $\mu \in \mathbb{R}^d$ and define

$$(8.5) \quad X_j = \sigma (S_j \mu + \sqrt{c_{S_j}} G_\alpha Z_j), \quad c_{s_-} = c_-, \quad c_{s_+} = c_+.$$

Here $\sigma > 0$ is fixed and controls the overall noise magnitude, $c_-, c_+ > 0$ are fixed, and the S_j and Z_j are mutually independent i.i.d. sequences.

The two values of S_j may be viewed as representing two different observation conditions. The first term in (8.5) represents a random coherent perturbation along the fixed spatial pattern μ , while the second represents spatially correlated Gaussian noise whose strength is modulated by the same latent state S_j .

Conditional on $S_j = s_\pm$,

$$(8.6) \quad X_j \mid S_j = s_\pm \sim \mathcal{N}(\sigma s_\pm \mu, \sigma^2 c_\pm C_\alpha).$$

We choose c_-, c_+ so that

$$(8.7) \quad \bar{c} := \frac{3}{4}c_- + \frac{1}{4}c_+ = 1.$$

Consequently,

$$(8.8) \quad \mathbb{E}X_j = 0, \quad \Sigma := \text{Cov}(X_j) = \sigma^2(C_\alpha + \mu\mu^\top).$$

The cgf of X_j is

$$(8.9) \quad \begin{aligned} \varphi(\tau) &= \log \left(\frac{3}{4} \exp \left\{ \sigma s_- \langle \mu, \tau \rangle + \frac{\sigma^2 c_-}{2} \langle \tau, C_\alpha \tau \rangle \right\} \right. \\ &\quad \left. + \frac{1}{4} \exp \left\{ \sigma s_+ \langle \mu, \tau \rangle + \frac{\sigma^2 c_+}{2} \langle \tau, C_\alpha \tau \rangle \right\} \right). \end{aligned}$$

The density of X_j is a convex combination of two Gaussian, hence LC, densities, with means $\sigma s_\pm \mu$ and covariance matrices $\sigma^2 c_\pm C_\alpha$. If $\|\mu\|$ is bounded uniformly in d and $|\alpha| < 1/4$, their means and covariance matrices satisfy the uniform bounds in (7.16). Hence Proposition 7.3 implies that the conclusions of Proposition 7.1 hold for the present model.

For the numerical experiment we can also evaluate the true likelihood exactly. Let

$$(8.10) \quad K := \#\{j : S_j = s_+\}, \quad K \sim \text{Bin}(n, 1/4).$$

Conditional on $K = k$, the average $\bar{X}_n = n^{-1} \sum_{j=1}^n X_j$ is Gaussian:

$$(8.11) \quad \bar{X}_n \mid K = k \sim \mathcal{N}(m_k, v_k C_\alpha),$$

where

$$(8.12) \quad m_k = \frac{\sigma}{n} [(n-k)s_- + ks_+] \mu, \quad v_k = \frac{\sigma^2}{n^2} [(n-k)c_- + kc_+].$$

By (8.1), $\bar{Y} = m_\theta + \bar{X}_n$, where the bars denote the average values, so the likelihood of θ based on the averaged observation $\bar{y}$ is the pdf of $\bar{X}_n$ evaluated at $\bar{y} - m_\theta$. Writing

$$(8.13) \quad \pi_k = \binom{n}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{n-k},$$

and setting $a_\theta := \bar{y} - m_\theta$, we obtain the exact likelihood, with the full density normalization retained,

$$(8.14) \quad \begin{aligned} L_n^{\text{true}}(\theta; \bar{y}) &= \rho_n(\bar{y} - m_\theta) = \rho_n(a_\theta) \\ &= \sum_{k=0}^n \frac{\pi_k}{(2\pi v_k)^{d/2} (\det C_\alpha)^{1/2}} \exp \left\{ -\frac{1}{2v_k} \langle a_\theta - m_k, C_\alpha^{-1}(a_\theta - m_k) \rangle \right\}. \end{aligned}$$

8.2. Conventional SPA and its central expansions. For the two-point variable in (8.4),

$$(8.15) \quad \kappa_3 := \mathbb{E}S_j^3 = 2/\sqrt{3}, \quad \beta := \mathbb{E}(S_j c_{S_j}).$$

For the numerical values used below, $c_- = 0.7$ and $c_+ = 1.9$, so $\beta = 3\sqrt{3}/10$.

In addition to the conventional SPA, we consider the first two central approximations discussed in Subsection 4.2: the leading Gaussian approximation and its first non-Gaussian, cubic correction.

Let $T_3 := \nabla^3 \varphi(0)$ and put $u_\theta := \Sigma^{-1} a_\theta$. The logarithm of the conventional SPA obtained from (2.22) by replacing $I(a_\theta)$ by 1 is

$$(8.16) \quad \ell_n^S(\theta; \bar{y}) = \frac{d}{2} \log n - \frac{d}{2} \log(2\pi) - \frac{1}{2} \log \det H(\tau_\theta) - n\varphi^*(a_\theta), \quad \nabla \varphi(\tau_\theta) = a_\theta.$$

Expanding (8.16) about the central saddlepoint $a_\theta = 0$, $\tau_\theta = 0$, the constant term is

$$(8.17) \quad C_{n,\Sigma} := \frac{d}{2} \log n - \frac{d}{2} \log(2\pi) - \frac{1}{2} \log \det \Sigma.$$

Using the expansions derived in Appendix B, retaining the zeroth-order term in $\log \det H$ and the quadratic term in φ^* (see (B.13) and (B.11), respectively) gives the Gaussian-order approximation

$$(8.18) \quad \ell_n^G(\theta; \bar{y}) = C_{n,\Sigma} - \frac{n}{2} \langle a_\theta, \Sigma^{-1} a_\theta \rangle.$$

Retaining also the first non-Gaussian terms in the Legendre transform and in the determinant of the Hessian gives the cubic SPA

$$(8.19) \quad \ell_n^C(\theta; \bar{y}) = \ell_n^G(\theta; \bar{y}) + \frac{n}{6} T_3[u_\theta, u_\theta, u_\theta] - \frac{1}{2} \langle q, u_\theta \rangle,$$

where

$$(8.20) \quad \begin{aligned} T_3[u, u, u] &:= \sigma^3 [\kappa_3 \langle \mu, u \rangle^3 + 3\beta \langle \mu, u \rangle \langle u, C_\alpha u \rangle], \quad u \in \mathbb{R}^d, \\ q &:= \sigma^3 [\kappa_3 \langle \mu, \Sigma^{-1} \mu \rangle \mu + \beta \{ \text{tr}(C_\alpha \Sigma^{-1}) \mu + 2C_\alpha \Sigma^{-1} \mu \}]. \end{aligned}$$

The first correction in (8.19) comes from the cubic term in the Legendre transform, and the second from the expansion of the determinant of the Hessian; see Appendix B for more details. The normalization constant $C_{n,\Sigma}$ in (8.17) is retained because below we compare absolute density values. It could be dropped for maximization over θ .

Fix $R > 0$ and define the central parameter region

$$(8.21) \quad \Theta_{R,n}(\bar{y}) := \{ \theta : \|\Sigma^{-1/2} a_\theta\| \leq R(d/n)^{1/2} \}.$$

The same central expansion used in the proof of Corollary 4.3, namely (A.11)–(A.16), gives

$$(8.22) \quad \ell_n^{\text{true}}(\theta; \bar{y}) := \log L_n^{\text{true}}(\theta; \bar{y}) = \ell_n^C(\theta; \bar{y}) + O_R(d^2/n)$$

uniformly for $\theta \in \Theta_{R,n}(\bar{y})$. Thus the cubic SPA retains the leading contribution of the third cumulant.

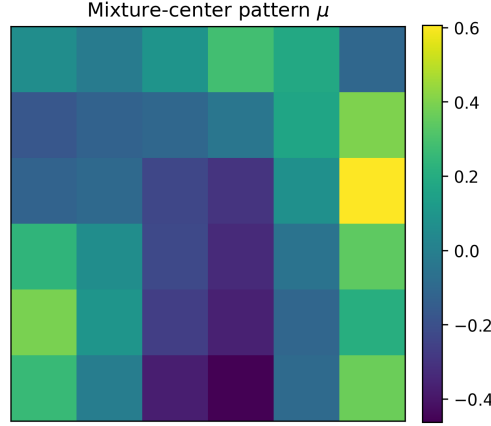


FIGURE 1. The smooth mixture-center pattern μ used in the numerical experiment. Its Euclidean norm is 1.5.

8.3. Numerical experiment. We take

$$(8.23) \quad \begin{aligned} m = 6, \quad d = 36, \quad n = 1000, \\ \alpha = 0.1, \quad \sigma = 0.25, \quad c_- = 0.7, \quad c_+ = 1.9. \end{aligned}$$

Notice that (8.7) is satisfied, and $d^2/n = 1.296$.

We carry out the numerical experiment in a regime where $d^2/n = 1.296$ is of order one rather than asymptotically small. There are two reasons for this choice. First, it provides a fairly stringent finite-dimensional stress test of the asymptotic theory. Second, when d^2/n is made small, the errors of the SPA and of its central expansions become so small that their differences are difficult to illustrate numerically. The present regime makes these differences visible.

For later use in the reconstruction experiment, and also to specify the fixed spatial pattern μ , we use the first 4×4 tensor-product orthonormal discrete cosine transform (DCT) modes on the 6×6 patch:

$$(8.24) \quad m_\theta = B\theta, \quad B : \mathbb{R}^{16} \longrightarrow \mathbb{R}^{36}.$$

The vector μ is a fixed smooth pattern in the same DCT subspace. The 16 DCT coefficients of μ were generated once with weights proportional to $(1 + k + \ell)^{-1/2}$, and the resulting vector was normalized so that $\|\mu\| = 1.5$. The resulting pattern is shown in Figure 1.

The exact likelihood (8.14) contains only $n + 1 = 1001$ Gaussian components, and the entire sum is evaluated deterministically. The conventional SPA is evaluated directly from (2.22) with $I(a)$ replaced by 1, while the Gaussian and cubic approximations are evaluated from (8.18) and (8.19).

We first compare the density approximations themselves, independently of any particular ground-truth image or data realization. Write

$$(8.25) \quad \rho_n^S(a) := \frac{C_{d,n}}{\sqrt{\det H(\tau)}} e^{-n\varphi^*(a)}, \quad \nabla\varphi(\tau) = a,$$

for the conventional SPA obtained from (2.22) by replacing $I(a)$ by 1. We similarly write $\rho_n^G(a)$ and $\rho_n^C(a)$ for the Gaussian and cubic approximations obtained from (8.18) and (8.19), respectively, with a_θ replaced by a .

We begin with the locations of the modes. Let a_*^{true} , a_*^S , and a_*^G denote the global maximizers of the exact density, the conventional SPA, and the Gaussian approximation, respectively, and let a_*^C denote the central local maximizer of the cubic approximation. The latter qualification is necessary because ρ_n^C is only a central approximation.

Set

$$(8.26) \quad e := \frac{\Sigma^{-1/2}\mu}{\|\Sigma^{-1/2}\mu\|}.$$

By (8.8) and (8.9), in the whitened coordinates ρ_n is invariant under rotations that fix e (see also (C.4)). The same symmetry is inherited by ρ_n^M , where $M \in \{S, C, G\}$. Writing

$$(8.27) \quad \Sigma^{-1/2}a = ze + y, \quad y \perp e, \quad r := \|y\|,$$

this symmetry implies that the exact density and the three approximations depend only on (z, r) . Numerically, the maximizing transverse radius is $r = 0$ in all four cases, so all four relevant modes lie on the axis spanned by e . Define

$$(8.28) \quad \zeta_M := \sqrt{n} \langle e, \Sigma^{-1/2} a_*^M \rangle, \quad \delta_M := \sqrt{n} \|\Sigma^{-1/2} (a_*^M - a_*^{\text{true}})\|,$$

where $M \in \{\text{true}, S, C, G\}$. Thus ζ_M measures the signed mode location in units of the one-dimensional standard deviation of the averaged noise, while δ_M measures the error in the mode location. The results are

	ζ_M	δ_M
Exact density	-0.24737	0
Conventional SPA	-0.24702	3.48×10^{-4}
Cubic SPA	-0.24630	1.06×10^{-3}
Gaussian	0	2.47×10^{-1}

TABLE 1. Locations of the exact and approximate modes. For the cubic approximation, the central local mode is used.

Thus, although $\mathbb{E}\bar{X}_n = 0$, the exact density has a nonzero mode as a result of the skewness of the noise distribution. Both the conventional SPA and the cubic approximation reproduce this finite- n displacement very accurately. The Gaussian approximation, whose mode is necessarily at zero, misses it completely.

We next assess the sharpness of the bound in Theorem 3.1. For simplicity, we do this only at the central point $a = 0$, for which the corresponding saddlepoint is $\tau = 0$. At this fixed saddlepoint, the apparently high-dimensional maximizations in (2.8) reduce, by the rotational symmetry of the whitened model, to four-dimensional compact optimization problems; see Appendix C. Numerically, this gives

$$(8.29) \quad c_{3,0} = 1.306, \quad c_{4,0} = 2.336,$$

where the subscript 0 indicates that the supremum over $\tau \in V_d$ in (2.8) is replaced by $\tau = 0$. Using these values in the proof of Theorem 3.1 gives

$$(8.30) \quad |I(0) - 1| \leq 1.064 + \text{e.s.t.}$$

Direct numerical evaluation gives instead

$$(8.31) \quad |I(0) - 1| = 1.595 \times 10^{-2}.$$

Thus, disregarding the exponentially small tail terms, the theoretical bound exceeds the actual error at the central point by a factor of approximately 67. We regard this as a reasonably informative non-asymptotic bound, particularly since it is being evaluated in a regime where $d^2/n = 1.296$ is not small.

We next illustrate the Monte Carlo refinement of Subsection 4.1. Using $M = 10^7$ samples in (4.21), we obtain

$$(8.32) \quad \hat{J}_M(0) - 1 = 1.5935 \times 10^{-2}.$$

As an a posteriori validation, comparing with the exact value gives

$$(8.33) \quad |I(0) - \hat{J}_M(0)| = 1.60 \times 10^{-5}.$$

Thus the Monte Carlo calculation gives an accurate estimate of the SPA error at the central point.

We next use the exact density to compare the conventional SPA and its two central expansions throughout the natural high-dimensional central region. Introduce the standardized shells

$$(8.34) \quad \mathcal{S}_r := \left\{ a : \|\Sigma^{-1/2}a\| = r(d/n)^{1/2} \right\}.$$

Thus r measures radial distance in units of the natural high-dimensional central scale $(d/n)^{1/2}$. Under the Gaussian approximation, the typical radius corresponds to $r \approx 1$.

For $M \in \{S, C, G\}$ define the signed log-density error

$$(8.35) \quad \Delta_M(a) := \log \frac{\rho_n^M(a)}{\rho_n(a)},$$

and set $\Delta_{\text{true}}(a) \equiv 0$. In particular,

$$(8.36) \quad \Delta_S(a) = -\log I(a).$$

By the rotational symmetry described above, each $\Delta_M(a)$ depends on a only through $\|\Sigma^{-1/2}a\|$ and $\langle e, \Sigma^{-1/2}a \rangle$. Hence, on each $\mathcal{S}_r$, the complete directional dependence is described by the single angular coordinate

$$(8.37) \quad \xi := \frac{\langle \Sigma^{-1/2}a, e \rangle}{\|\Sigma^{-1/2}a\|} \in [-1, 1].$$

Here $\xi = 1$ corresponds to the direction of e , $\xi = -1$ to the opposite direction, and $\xi = 0$ to directions orthogonal to e .

Figures 2 and 3 show the $\Delta_M(a)$ profiles as functions of ξ for $r = 1/6, 1/3, 1$, and $4/3$. The first two are inner central shells, $r = 1$ is the natural Gaussian radial scale, and $r = 4/3$ probes somewhat farther from the center.

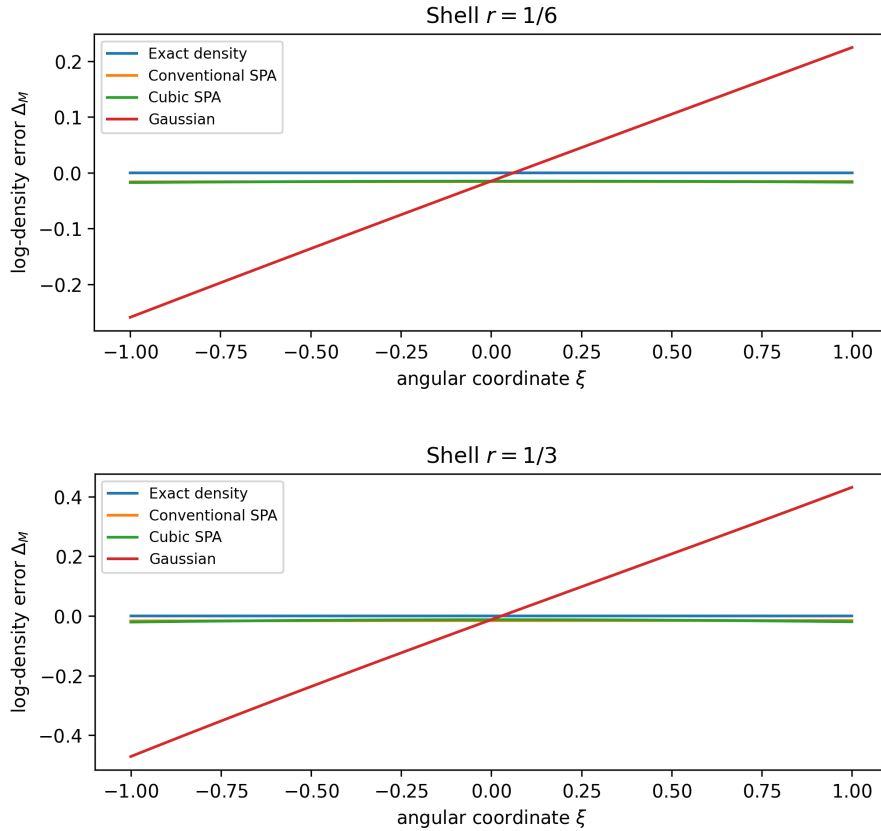


FIGURE 2. Signed log-density errors (8.35) of the conventional SPA, cubic SPA, and Gaussian approximation as functions of the angular coordinate (8.37). Top: the shell $\mathcal{S}_{1/6}$. Bottom: the shell $\mathcal{S}_{1/3}$. The exact density corresponds to the horizontal zero line.

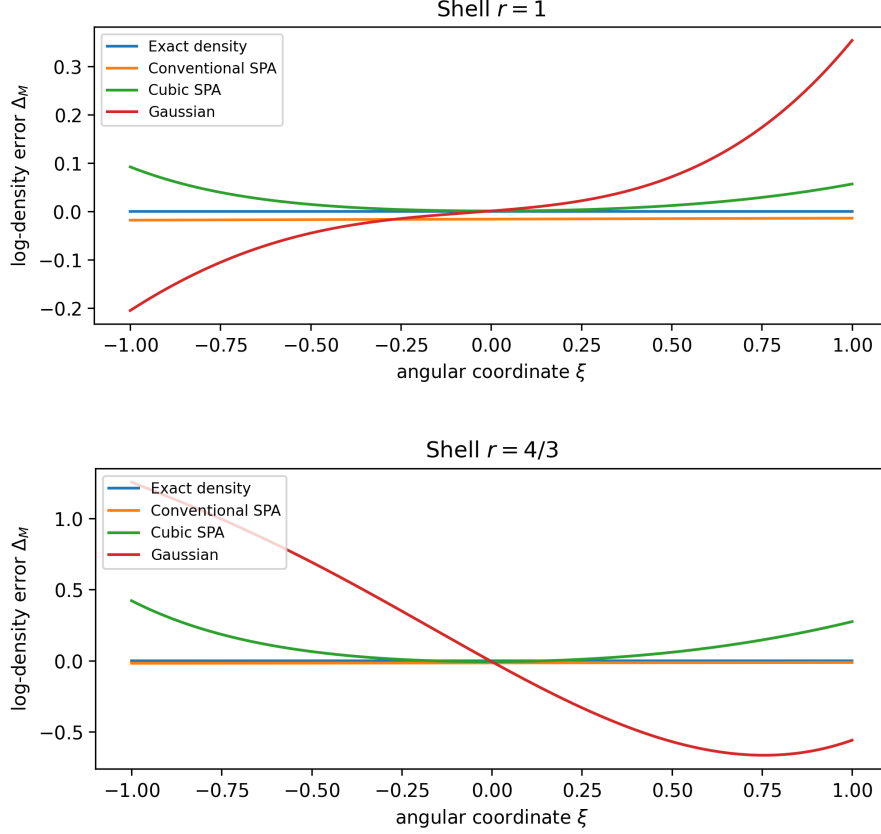


FIGURE 3. Signed log-density errors (8.35) of the conventional SPA, cubic SPA, and Gaussian approximation as functions of the angular coordinate (8.37). Top: the shell $\mathcal{S}_1$. Bottom: the shell $\mathcal{S}_{4/3}$. The exact density corresponds to the horizontal zero line.

The conventional-SPA error is nearly independent of direction on all four shells and remains close to its value at the central saddlepoint. Indeed, since $\Delta_S(a) = -\log I(a)$ and $I(a)$ is close to one, the near constancy of Δ_S is equivalent to the near constancy of $I(a) - 1$. Thus the single Monte Carlo estimate $\hat{J}_M(0) - 1 = 1.5935 \times 10^{-2}$ remains representative of the conventional-SPA error throughout the displayed region, which is qualitatively consistent with the refinement of Theorem 4.1.

The central expansions exhibit substantially stronger directional dependence. On the two inner shells, the cubic approximation remains close to the conventional SPA, whereas the Gaussian error varies almost linearly with ξ . This linear behavior has a simple local explanation. On a fixed shell S_r , the Gaussian log density $\log \rho_n^G(a)$ is independent of direction, since $\|\Sigma^{-1/2}a\|$ is fixed. Hence the directional dependence of $\Delta_G(a)$ comes entirely from the exact log density. By rotational symmetry in the whitened coordinates, near $a = 0$,

$$(8.38) \quad \log \rho_n(a) = \log \rho_n(0) + c\langle e, \Sigma^{-1/2}a \rangle + O(\|\Sigma^{-1/2}a\|^2)$$

for some constant c . Consequently, on S_r the leading directional term is linear in ξ , with a coefficient proportional to $r(d/n)^{1/2}$; see (8.34) and (8.37). Thus the nearly straight Gaussian profiles on the inner shells and the mode displacement in Table 1 have the same local origin. At $r = 1$, the cubic correction still removes most of the directional error of the Gaussian approximation. On the outer shell $r = 4/3$, both central expansions deteriorate in directions nearly parallel or antiparallel to e , with the Gaussian approximation deteriorating much more strongly.

We finally illustrate the use of SPA in one particular inverse problem. For the true image, we start from a fixed cross-shaped pattern on the 6×6 pixel grid and project it onto the 4×4

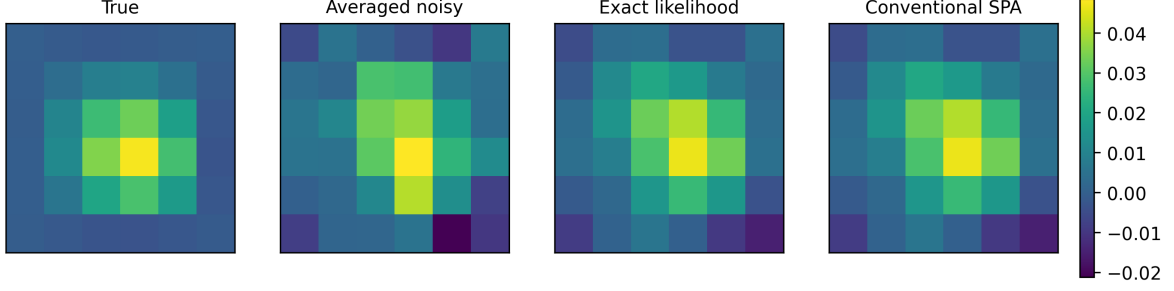


FIGURE 4. True image, averaged noisy image, and reconstructions obtained by maximizing the exact likelihood and the conventional SPA. The reconstructions obtained from the cubic and Gaussian approximations are visually indistinguishable at this scale from the two reconstructions shown and are therefore omitted.

DCT model space in (8.24). The resulting image is then normalized so that

$$(8.39) \quad \max_x |m_{\theta_0}(x)| = 1.5/\sqrt{n}.$$

This scaling keeps the signal amplitude comparable to the residual noise level in the averaged image, which is of order $n^{-1/2}$, and hence makes the reconstruction problem nontrivial. The true image and a single realization of all 1000 exposures are kept fixed throughout the experiment.

Figure 4 shows the true image, the averaged noisy image, and the reconstructions obtained by maximizing the exact likelihood and the conventional SPA. The reconstructions obtained from the cubic and Gaussian approximations are visually indistinguishable at this scale from the exact-likelihood and conventional-SPA reconstructions shown in the figure and are therefore omitted. We report their numerical errors below. The relative L^2 errors with respect to m_{θ_0} are

$$(8.40) \quad \frac{\|\bar{y} - m_{\theta_0}\|_2}{\|m_{\theta_0}\|_2} = 0.53307, \quad \frac{\|\hat{m}_M - m_{\theta_0}\|_2}{\|m_{\theta_0}\|_2} = 0.38251, \quad 0.38252, \quad 0.38251, \quad 0.38549,$$

where $M = \text{true}, S, C, G$, respectively, and

$$(8.41) \quad \hat{\theta}^M := \arg \max_{\theta \in \mathbb{R}^{16}} L_n^M(\theta; \bar{y}), \quad \hat{m}^M := B\hat{\theta}^M.$$

Recall that B is defined in (8.24).

We see that all three approximate-likelihood reconstructions give nearly the same point estimate as the exact likelihood in this realization, with the conventional and cubic SPAs agreeing especially closely with it. This indicates that, for this realization, the likelihood maximizer is relatively insensitive to the density-approximation errors. This observation is consistent with general fixed-dimensional results on the accuracy of saddlepoint maximum-likelihood estimators [17].

APPENDIX A. AUXILIARY RESULTS

A.1. Bound on the gradient. Pick any $f \in C^3(\mathbb{R}^d)$ such that

$$(A.1) \quad f(0) = 0, \quad \nabla f(0) = 0, \quad \nabla^2 f(0) = 0.$$

Letting $g(\lambda) := \nabla f(\lambda x) \cdot u$ for any $u \in S^{d-1}$, we have $g(0) = 0$, $g'(0) = \nabla^2 f(0)[x, u] = 0$, and $g''(\lambda) = \nabla^3 f(\lambda x)[x, x, u]$. Taylor's formula with integral remainder yields

$$(A.2) \quad g(1) = \int_0^1 (1-\lambda) g''(\lambda) d\lambda = \int_0^1 (1-\lambda) \nabla^3 f(\lambda x)[x, x, u] d\lambda.$$

Taking the supremum over $u \in S^{d-1}$ yields

$$(A.3) \quad \|\nabla f(x)\| \leq \frac{1}{2} \sup_{\lambda \in [0,1]} \|\nabla^3 f(\lambda x)\|_{\text{op}} \|x\|^2.$$

A.2. A property of the Legendre transform. Consider the equation $\nabla\varphi(\tau) = a$. In the setting of Subsection 4.2, X is isotropic, so $\nabla\varphi(0) = 0$ and $\nabla^2\varphi(0) = I_d$. Similarly to (A.2),

$$(A.4) \quad \langle \nabla\varphi(\tau), u \rangle = \langle \tau, u \rangle + \int_0^1 (1-\lambda) \nabla^3\varphi(\lambda\tau)[\tau, \tau, u] d\lambda = \langle a, u \rangle.$$

This implies

$$(A.5) \quad \|\tau - a\| \leq \frac{1}{2} \max_{\lambda \in [0,1]} \|\nabla^3\varphi(\lambda\tau)\|_{\text{op}} \|\tau\|^2.$$

Assuming $\max_{\lambda \in [0,1]} \|\nabla^3\varphi(\lambda\tau)\|_{\text{op}} \|\tau\| \leq 1$, we get $\|\tau\| \leq 2\|a\|$.

Further, a simple manipulation shows that

$$(A.6) \quad \varphi^*(a) - \frac{1}{2}\|a\|^2 = -\frac{1}{2}\|\tau - a\|^2 - \frac{1}{6}\nabla^3\varphi(\lambda\tau)[\tau, \tau, \tau]$$

for some $\lambda \in (0, 1)$. Combining with the preceding results proves (4.25).

Note that (4.26) implies that the assumption $\max_{\lambda \in [0,1]} \|\nabla^3\varphi(\lambda\tau)\|_{\text{op}} \|\tau\| \leq 1$ is satisfied. This and the existence of a solution τ in the set $\|\tau\| \leq 2\|a\|$ for any a that satisfies (4.26) follows by removing u from (A.4) and applying Brouwer's fixed-point theorem to the resulting equation.

Indeed, define the matrix function:

$$(A.7) \quad B(\tau) = \int_0^1 (1-\lambda) \nabla^3\varphi(\lambda\tau) \tau d\lambda.$$

By construction, $\|B(\tau)\| \leq 1/2$ whenever $\|\tau\| \leq 2\|a\|$, $a \in U_d$, and (4.26) holds. Under these conditions, (A.4) is equivalent to the fixed-point equation $\tau = D(\tau)$, where $D(\tau) := (I + B(\tau))^{-1}a$, defined on the ball $\|\tau\| \leq 2\|a\|$. Since $\|(I + B(\tau))^{-1}\| \leq 2$, the map $D(\tau)$ sends this ball into itself. By Brouwer's fixed-point theorem, a solution exists. Uniqueness follows from the strict positivity of $\nabla^2\varphi(\tau)$, $\tau \in V_d$, cf. Assumption 2.1(2).

A.3. Proof of Corollary 4.3. Set $a = x/\sqrt{n}$. On the region $\|x\| \leq L\sqrt{d}$,

$$(A.8) \quad \|a\| \leq L(d/n)^{1/2} \rightarrow 0$$

when $d^2/n \rightarrow 0$.

By Lemma 2.2, there exists a fixed $0 < r' \leq r$ such that parts (1), (2), and (4) of Assumption 2.1 hold on $V'_d = \{\|\tau\| < r'\}$, and c_3, c_4 are uniformly bounded. Part (3) holds there by restriction from $\{\|\tau\| < r\}$.

Since (A.8) and the uniform local bounds on $H(\tau)$ in (2.12) imply $\|\tau\| \lesssim \|a\|$, the corresponding saddlepoints $\tau(a)$ converge to zero uniformly for $\|x\| \leq L\sqrt{d}$. We used here that X is isotropic and that (2.9) holds; together these imply (2.12), by the argument following (2.12). Hence all $\tau(a)$ belong to V'_d for d, n sufficiently large, and Theorem 3.1 may be applied on V'_d .

Taylor expansion gives

$$(A.9) \quad \varphi(\tau) = \frac{1}{2}\|\tau\|^2 + \frac{1}{6}T_3[\tau, \tau, \tau] + O(\|\tau\|^4)$$

uniformly in d , and therefore

$$(A.10) \quad \nabla\varphi(\tau) = \tau + \frac{1}{2}T_3[\tau, \tau, \cdot] + O(\|\tau\|^3).$$

Solving $\nabla\varphi(\tau) = a$ by one substitution yields

$$(A.11) \quad \tau = a - \frac{1}{2}T_3[a, a, \cdot] + O(\|a\|^3).$$

Substitution into $\varphi^*(a) = \langle \tau, a \rangle - \varphi(\tau)$ gives

$$(A.12) \quad \varphi^*(a) = \frac{1}{2}\|a\|^2 - \frac{1}{6}T_3[a, a, a] + O(\|a\|^4).$$

Since $n\|a\|^4 = O_L(d^2/n)$, we obtain

$$(A.13) \quad n(\varphi^*(a) - \frac{1}{2}\|a\|^2) = -\frac{1}{6\sqrt{n}}T_3[x, x, x] + O_L\left(\frac{d^2}{n}\right).$$

Similarly,

$$(A.14) \quad H(\tau) = I_d + T_3[\tau, \cdot, \cdot] + O(\|\tau\|^2),$$

and (A.11) gives

$$(A.15) \quad H(\tau) = I_d + T_3[a, \cdot, \cdot] + O(\|a\|^2).$$

Expanding the logarithm of the determinant around I_d therefore gives

$$(A.16) \quad \log \det H(\tau) = \operatorname{tr} T_3[a, \cdot, \cdot] + O(d\|a\|^2) = \frac{1}{\sqrt{n}} \operatorname{tr} T_3[x, \cdot, \cdot] + O_L\left(\frac{d^2}{n}\right).$$

The density of S_n is related to the density ρ_n of the sample mean by

$$(A.17) \quad f_n(x) = n^{-d/2} \rho_n(x/\sqrt{n}).$$

Using (2.22), (A.13), (A.16), and Theorem 3.1 now gives (4.33).

A.4. A lower bound for $\|(nH_r)^{-1/2}\|_\infty$. We show that the factor $d^{1/2}$ in (5.27) cannot in general be removed under (5.4). It suffices to consider $z = \tau = 0$ and $\Sigma = I_d$. Since $A(0) = 1$, (5.9) gives

$$(A.18) \quad H(0) = s(I_d + bb^\top).$$

Choose

$$(A.19) \quad b = 2(1, (d-1)^{-1/2}, \dots, (d-1)^{-1/2})^\top.$$

Then $\|b\|^2 = 8$, so this choice satisfies (5.4) uniformly in d . By an easy calculation,

$$(A.20) \quad (I_d + bb^\top)^{-1/2} = I_d - \frac{1}{12}bb^\top.$$

Since $H_r(0) = H(0)$, it follows that

$$(A.21) \quad (nH_r(0))^{-1/2} = (ns)^{-1/2}(I_d - \frac{1}{12}bb^\top).$$

Consider the first row of the matrix in (A.21). Its diagonal entry equals

$$(A.22) \quad 1 - \frac{1}{12}b_1^2 = \frac{2}{3},$$

while each of its $d-1$ off-diagonal entries has absolute value

$$(A.23) \quad \frac{1}{12}|b_1b_j| = \frac{1}{3\sqrt{d-1}}, \quad j = 2, \dots, d.$$

Therefore

$$(A.24) \quad \|(nH_r(0))^{-1/2}\|_\infty \geq \frac{1}{\sqrt{ns}} \left(\frac{2}{3} + \frac{1}{3}\sqrt{d-1} \right) \gtrsim (d/n)^{1/2}.$$

APPENDIX B. DERIVATION OF THE CUBIC LIKELIHOOD IN SECTION 8

We give the details of the calculations leading to (8.19), (8.20). From (8.5),

$$(B.1) \quad X = \sigma(S\mu + \sqrt{c_S}G_\alpha Z), \quad Z \sim \mathcal{N}(0, I_d).$$

Here S has the distribution in (8.4),

$$(B.2) \quad \mathbb{E}S = 0, \quad \mathbb{E}S^2 = 1, \quad \mathbb{E}c_S = 1,$$

and Z is independent of S . It follows immediately that $\mathbb{E}X = 0$ and

$$(B.3) \quad \operatorname{Var}(X) = \sigma^2(C_\alpha + \mu\mu^\top) = \Sigma,$$

as in (8.8).

Since X is centered,

$$(B.4) \quad T_3 = \nabla^3 \varphi(0) = \mathbb{E}X^{\otimes 3}.$$

Expanding $X^{\otimes 3}$ using (B.1), all terms containing an odd number of Gaussian factors vanish. Therefore,

$$(B.5) \quad \begin{aligned} T_3 &= \sigma^3 [(\mathbb{E}S^3)\mu^{\otimes 3} + \mathbb{E}(Sc_S) \operatorname{Sym}(\mu \otimes C_\alpha)] \\ &= \sigma^3 [\kappa_3 \mu^{\otimes 3} + \beta \operatorname{Sym}(\mu \otimes C_\alpha)], \end{aligned}$$

where

$$(B.6) \quad \begin{aligned} \text{Sym}(\mu \otimes C_\alpha)[u, u, u] &= 3\langle \mu, u \rangle \langle u, C_\alpha u \rangle, \quad u \in \mathbb{R}^d, \\ \text{Sym}(\mu \otimes C)_{ijk} &= \mu_i C_{jk} + \mu_j C_{ik} + \mu_k C_{ij}. \end{aligned}$$

This proves the first equation in (8.20).

We next derive the vector q in (8.20). This vector appears in the expansion of the determinant term below and it is defined by the property

$$(B.7) \quad \text{tr}\{AT_3[u, \cdot, \cdot]\} = \sum_{i=1}^d \left(\sum_{j,k=1}^d A_{jk}(T_3)_{ijk} \right) u_i = \langle q, u \rangle,$$

where $A = \Sigma^{-1}$. We have

$$(B.8) \quad \sum_{j,k=1}^d \mu_i \mu_j \mu_k A_{jk} = \langle \mu, A\mu \rangle \mu_i$$

and

$$(B.9) \quad \begin{aligned} \sum_{j,k=1}^d (\mu_i (C_\alpha)_{jk} + \mu_j (C_\alpha)_{ik} + \mu_k (C_\alpha)_{ij}) A_{jk} \\ = \text{tr}(C_\alpha A) \mu_i + 2(C_\alpha A \mu)_i. \end{aligned}$$

Combining (B.5) with (B.7)–(B.9) therefore gives

$$(B.10) \quad q = \sigma^3 [\kappa_3 \langle \mu, A\mu \rangle \mu + \beta \{ \text{tr}(C_\alpha A) \mu + 2C_\alpha A \mu \}],$$

which is (8.20).

It remains to derive the cubic approximation to the likelihood. This is the non-isotropic version of the calculation leading to (A.11)–(A.16). Indeed, since $\mathbb{E}X = 0$ and $\nabla^2 \varphi(0) = \Sigma$, similar Taylor expansions, with $A = \Sigma^{-1}$ and $u = Aa$, give

$$(B.11) \quad \begin{aligned} \tau &= u - \frac{1}{2} AT_3[u, u, \cdot] + O(\|a\|^3), \\ \varphi^*(a) &= \frac{1}{2} \langle a, Aa \rangle - \frac{1}{6} T_3[u, u, u] + O(\|a\|^4). \end{aligned}$$

Similarly,

$$(B.12) \quad H(\tau) = \Sigma + T_3[u, \cdot, \cdot] + O(\|a\|^2),$$

and hence

$$(B.13) \quad \begin{aligned} \log \det H(\tau) &= \log \det \Sigma + \text{tr}\{AT_3[u, \cdot, \cdot]\} + O(d\|a\|^2) \\ &= \log \det \Sigma + \langle q, u \rangle + O(d\|a\|^2), \end{aligned}$$

where the last equality follows from (B.7).

The logarithm of the conventional SPA for the density of $n^{-1} \sum_{j=1}^n X_j$ is

$$(B.14) \quad \frac{d}{2} \log n - \frac{d}{2} \log(2\pi) - \frac{1}{2} \log \det H(\tau) - n\varphi^*(a).$$

Substitution of the preceding expansions for $\varphi^*(a)$ and $H(\tau)$ gives

$$(B.15) \quad C_{n,\Sigma} - \frac{n}{2} \langle a, Aa \rangle + \frac{n}{6} T_3[u, u, u] - \frac{1}{2} \langle q, u \rangle,$$

which is (8.19).

Finally, in the central region (8.21), $\|a\| = O_R((d/n)^{1/2})$ under the uniform spectral bounds on Σ . Consequently both $nO(\|a\|^4)$ and $O(d\|a\|^2)$ are $O_R(d^2/n)$. Together with Theorem 3.1, this gives (8.22).

APPENDIX C. EVALUATION OF c_3 AND c_4 IN THE NUMERICAL EXPERIMENT

We explain the numerical evaluation of the derivative constants used to assess the bound in Theorem 3.1 at $a = 0$. The corresponding saddlepoint is $\tau = 0$. Set

$$(C.1) \quad M := H(0)^{-1/2}, \quad v := M\mu.$$

Replacing τ with iMt in (8.9) gives

$$(C.2) \quad \varphi(iMt) = F(\langle v, t \rangle, \langle Mt, C_\alpha Mt \rangle)$$

for an explicit scalar function F . We have $H(0) = \sigma^2(C_\alpha + \mu\mu^\top)$; see (8.8) and, for its derivation, (B.3). Therefore,

$$(C.3) \quad \sigma^2(MC_\alpha M + vv^\top) = I_d.$$

Consequently,

$$(C.4) \quad \varphi(iMt) = F(\langle v, t \rangle, \sigma^{-2}\|t\|^2 - \langle v, t \rangle^2).$$

To compute the tensor operator norms at a fixed t , let $\Theta \in S^{d-1}$ and replace t with $t + h\Theta$ in (C.4). Set

$$(C.5) \quad x := \langle v, t \rangle, \quad r := \|t\|, \quad u := \langle v, \Theta \rangle, \quad w := \langle t, \Theta \rangle.$$

Then

$$(C.6) \quad \varphi(iM(t + h\Theta)) = F(x + hu, \sigma^{-2}(r^2 + 2hw + h^2) - (x + hu)^2).$$

Thus the directional derivatives of orders three and four with respect to h at $h = 0$ depend only on the four scalar quantities x, r, u, w . For each fixed t , the computation of the tensor operator norm therefore reduces to a two-dimensional optimization over the geometrically feasible pairs (u, w) , with $|u| \leq \|v\|$ and $|w| \leq r$, while the supremum over the ball $\{\|t\| < 2.5\epsilon^{1/2}\}$ in (2.8) reduces to a two-dimensional optimization in (x, r) , with $|x| \leq \|v\|r$. Hence the full computation of $c_{3,0}$ and $c_{4,0}$ is a four-dimensional compact optimization problem.

DECLARATION OF GENERATIVE AI USE

OpenAI's ChatGPT was used in the preparation of this manuscript for substantial assistance with the Python implementation of the numerical experiments, for routine algebraic calculations, and for English-language editing and stylistic suggestions. The Python code was reviewed, tested, and validated by the author. The mathematical arguments and conclusions are the author's own, and the author takes full responsibility for the content of the manuscript and the reported numerical results.

REFERENCES

- [1] N. H. Anderson, P. Hall, and D. Titterton. Edgeworth expansions in very-high-dimensional problems. *Journal of Statistical Planning and Inference*, 70(1):1–18, July 1998.
- [2] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and Geometry of Markov Diffusion Operators*, volume 348 of *Grundlehren Der Mathematischen Wissenschaften*. Springer International Publishing, Cham, 2014.
- [3] A. Belloni and V. Chernozhukov. Posterior inference in curved exponential families under increasing dimensions. *The Econometrics Journal*, 17(2):S75–S100, June 2014.
- [4] V. Bentkus. On the dependence of the Berry–Esseen bound on dimension. *Journal of Statistical Planning and Inference*, 113(2):385–402, May 2003.
- [5] V. Bentkus. A Lyapunov-type Bound in $\mathbb{R}^d$. *Theory of Probability & Its Applications*, 49(2):311–323, Jan. 2005.
- [6] V. Bentkus and F. Götze. Optimal rates of convergence in the CLT for quadratic forms. *The Annals of Probability*, 24(1), Jan. 1996.
- [7] R. N. Bhattacharya and R. Ranga Rao. *Normal Approximation and Asymptotic Expansions*. Number 64 in *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, 2010.
- [8] P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley, Hoboken, New Jersey, anniversary edition, 2012.
- [9] S. Bobkov, G. Chistyakov, and F. Götze. Bounds for characteristic functions in terms of quantiles and entropy. *Electronic Communications in Probability*, 17, Jan. 2012.

- [10] S. Bobkov and M. Ledoux. From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities. *Geometric and Functional Analysis*, 10(5):1028–1052, Dec. 2000.
- [11] S. G. Bobkov and F. Götze. Berry–Esseen bounds in the local limit theorems. *Lithuanian Mathematical Journal*, 65(1):50–66, Jan. 2025.
- [12] S. A. Broda and M. S. Paoletta. Saddlepoint Approximations: A Review and Some New Applications. In J. E. Gentle, W. K. Härdle, and Y. Mori, editors, *Handbook of Computational Statistics*, pages 953–983. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
- [13] R. W. Butler. *Saddlepoint Approximations with Applications*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2007.
- [14] N. Chakraborty and S. N. Lahiri. The Central Limit Theorem in High-Dimensions and Its Applications. In S. Ghosal and A. Roy, editors, *Frontiers of Statistics and Data Science*, pages 81–100. Springer Nature Singapore, Singapore, 2025.
- [15] V. Chernozhukov, D. Chetverikov, and K. Kato. Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, 45(4), July 2017.
- [16] H. E. Daniels. Saddlepoint Approximations in Statistics. *The Annals of Mathematical Statistics*, 25(4):631–650, 1954.
- [17] J. Goodman. Asymptotic accuracy of the saddlepoint approximation for maximum likelihood estimation. *The Annals of Statistics*, 50(4), Aug. 2022.
- [18] F. Götze. On the Rate of Convergence in the Multivariate CLT. *The Annals of Probability*, 19(2), Apr. 1991.
- [19] D. D. Haroske and C. Schneider. Besov spaces with positive smoothness on $\mathbb{R}^n$, embeddings and growth envelopes. *Journal of Approximation Theory*, 161(2):723–747, Dec. 2009.
- [20] J. L. Jensen. Uniform Saddlepoint Approximations and Log-Concave Densities. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 53(1):157–172, Sept. 1991.
- [21] J. L. Jensen. *Saddlepoint Approximations*. Oxford University Press, Oxford, May 1995.
- [22] J. L. Jensen. On the Use of Saddlepoint Approximations in High Dimensional Inference. *Sankhya A*, 83(1):379–392, Feb. 2021.
- [23] M. J. Kasprzak, R. Giordano, and T. Broderick. How good is your Laplace approximation of the Bayesian posterior? Finite-sample computable error bounds for a variety of useful divergences. *Journal of Machine Learning Research*, 26:1–81, 2025.
- [24] A. Katsevich. The Laplace approximation accuracy in high dimensions: A refined analysis and new skew adjustment, June 2024. arXiv:2306.07262 [math].
- [25] A. Katsevich. Improved dimension dependence in the Bernstein–von Mises theorem via a new Laplace approximation bound. *Information and Inference: A Journal of the IMA*, 14(3):iaaf020, 2025.
- [26] A. Katsevich. The Laplace asymptotic expansion in high dimensions, June 2025. arXiv:2406.12706 [math].
- [27] A. Katsevich and A. Katsevich. High-dimensional Laplace asymptotics up to the concentration threshold, 2026. arXiv:2602.23151v2.
- [28] J. Kolassa and J. Li. Multivariate saddlepoint approximations in tail probability and conditional inference. *Bernoulli*, 16(4):1191–1207, 2010.
- [29] J. E. Kolassa. Multivariate saddlepoint tail probability approximations. *Annals of Statistics*, 31(1):274–286, 2003.
- [30] T. J. Kozubowski, K. Podgórski, and I. Rychlik. Multivariate generalized Laplace distribution and related random fields. *Journal of Multivariate Analysis*, 113:59–72, Jan. 2013.
- [31] A. Lewis. Galaxy shear estimation from stacked images. *Monthly Notices of the Royal Astronomical Society*, 398(1):471–476, Sept. 2009.
- [32] L. Lovász and S. Vempala. The geometry of logconcave functions and sampling algorithms. *Random Structures and Algorithms*, 30(3):307–358, 2007.
- [33] A. Maurer and M. Pontil. Empirical Bernstein bounds and sample variance penalization. In *22nd Annual Conference on Learning Theory (COLT)*, 2009.
- [34] S. Portnoy. On the central limit theorem in R^p when $p \rightarrow \infty$. *Probability Theory and Related Fields*, 73(4):571–583, Nov. 1986.
- [35] Y. Tang and N. Reid. Laplace and saddlepoint approximations in high dimensions. *Bernoulli*, 31(3), Aug. 2025.
- [36] R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, Sept. 2018.
- [37] C. S. Withers. Edgeworth Expansions When the Parameter Dimension Increases with Sample Size. *Econometrics*, 14(2):21, Apr. 2026.
- [38] W. Zhang, M. V. Bravington, and R. M. Fewster. Fast Likelihood-Based Inference for Latent Count Models Using the Saddlepoint Approximation. *Biometrics*, 75(3):723–733, Sept. 2019.
- [39] X. Zhang, C. Ling, and L. Qi. The Best Rank-1 Approximation of a Symmetric Tensor and Related Spherical Optimization Problems. *SIAM Journal on Matrix Analysis and Applications*, 33(3):806–821, Jan. 2012.
- [40] M. Zhilova. New Edgeworth-type expansions with finite sample guarantees. *The Annals of Statistics*, 50(5), Oct. 2022.