The Annals of Probability
2020, Vol. 48, No. 3, 1202–1219
https://doi.org/10.1214/19-AOP1388

NORMAL APPROXIMATION FOR WEIGHTED SUMS UNDER A SECOND-ORDER CORRELATION CONDITION

BY S. G. BOBKOV[1], G. P. CHISTYAKOV[2,*] AND F. GÖTZE[2,**]

[1]*School of Mathematics, University of Minnesota, bobkov@math.umn.edu*
[2]*Department of Mathematics, Bielefeld University, [*]chistyak@math.uni-bielefeld.de; [**]goetze@math.uni-bielefeld.de*

Under correlation-type conditions, we derive an upper bound of order $(\log n)/n$ for the average Kolmogorov distance between the distributions of weighted sums of dependent summands and the normal law. The result is based on improved concentration inequalities on high-dimensional Euclidean spheres. Applications are illustrated on the example of log-concave probability measures.

1. Introduction. Let $X = (X_1, \dots, X_n)$ be an isotropic random vector in $\mathbb{R}^n$ $(n \geq 2)$, that is, with uncorrelated components having mean zero and variance one. We consider the distribution functions $F_\theta(x) = \mathbb{P}\{S_\theta \leq x\}$ of the weighted sums

$$S_\theta = \theta_1 X_1 + \cdots + \theta_n X_n, \quad \theta = (\theta_1, \dots, \theta_n), \theta_1^2 + \cdots + \theta_n^2 = 1,$$

with coefficients taken from the unit sphere S^{n-1} in $\mathbb{R}^n$. Thus, $\mathbb{E} S_\theta = 0$ and $\mathrm{Var}(S_\theta) = 1$ for all $\theta \in S^{n-1}$.

The central limit problem is to determine natural conditions on X and θ which ensure that the random variables S_θ are nearly standard normal. In this case, one would also like to explore the rate of normal approximation in the Kolmogorov distance

$$\rho(F_\theta, \Phi) = \sup_x \left| F_\theta(x) - \Phi(x) \right|,$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2}\, dy, \quad x \in \mathbb{R},$$

is the standard normal distribution function. Let us briefly recall several well-known results in the case of independent components X_k. Here, one of general variants of the central limit theorem asserts that $\rho(F_\theta, \Phi)$ will be small, as long as X_k are identically distributed (the i.i.d. case), while $\max_k |\theta_k|$ is small. Moreover, under the third moment condition this property may be quantified by virtue of the Berry–Esseen bound

$$\rho(F_\theta, \Phi) \leq c \sum_{k=1}^n |\theta_k|^3 \mathbb{E}|X_k|^3. \tag{1.1}$$

Here and below, we denote by c, or by c_j with an integer index j absolute positive constants which may vary from place to place. The inequality (1.1) extends to the non-i.i.d. case as well [28, 29].

It easy to see that the sum in (1.1) is greater than or equal to $1/\sqrt{n}$ for all θ, and that (1.1) leads to this standard $\frac{1}{\sqrt{n}}$-rate in the i.i.d. case, once the coefficients θ_k are equal to each other. For general distributions of X_k, this standard rate cannot be improved by assuming stronger

Received October 2018; revised June 2019.
MSC2010 subject classifications. 60E, 60F.
Key words and phrases. Sudakov's typical distributions, normal approximation.

moment-type conditions. Nevertheless, one may look at the problem from an ensemble point of view in θ asking whether or not $\rho(F_\theta, \Phi)$ will be essentially smaller than $1/\sqrt{n}$ for most of θ on the sphere measured with the uniform probability measure $\mathfrak{s}_{n-1}$ on S^{n-1}. A striking result in this direction was obtained by Klartag and Sodin [21], showing in particular that

$$\mathbb{E}_\theta \rho(F_\theta, \Phi) \le \frac{c}{n} \bar{\beta}_4, \quad \bar{\beta}_4 = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E} X_k^4, \tag{1.2}$$

where we use $\mathbb{E}_\theta$ to denote the average over the measure $\mathfrak{s}_{n-1}$. Large deviation bounds for the set on the sphere where $\rho(F_\theta, \Phi)$ exceeds a multiple of $\frac{1}{n}\bar{\beta}_4$ are derived in [21] as well. Thus, when $\bar{\beta}_4$ is bounded like in the i.i.d. case, the distances $\rho(F_\theta, \Phi)$ turn out to be typically of order $1/n$ in contrast to the classical case of equal coefficients.

The aim of these notes is to extend this interesting phenomenon under a suitable correlation-type condition (and thus for some class of dependent X_k) to isotropic random vectors with a similar $\frac{1}{n}$-rate modulo a logarithmic factor. The scheme of the weighted sums under dependence has already a long history, going back to the seminal work of Sudakov [30]. We will give a short overview of this line of research in Section 10 (partly in Section 7), and now turn to the main result.

We will say that the random vector X satisfies a second-order correlation condition with constant Λ, if for any collection $a_{ij} \in \mathbb{R}$,

$$\mathrm{Var}\Bigg(\sum_{i,j=1}^{n} a_{ij} X_i X_j\Bigg) \le \Lambda \sum_{i,j=1}^{n} a_{ij}^2. \tag{1.3}$$

An optimal value $\Lambda = \Lambda(X)$ is finite as long as $|X|$ has a finite fourth moment, and then it represents the maximal eigenvalue of the covariance matrix associated with the n^2-dimensional random vector $(X_i X_j - \mathbb{E} X_i X_j)_{i,j=1}^n$.

THEOREM 1.1. *Let X be an isotropic random vector in $\mathbb{R}^n$ with a symmetric distribution and a finite constant $\Lambda = \Lambda(X)$. Then*

$$\mathbb{E}_\theta \rho(F_\theta, \Phi) \le \frac{c \log n}{n} \Lambda. \tag{1.4}$$

The characteristic Λ may be bounded, for example, via the relation $\Lambda \le 4/\lambda_1$ in terms of a positive spectral gap, that is in terms of the optimal value $\lambda_1 = \lambda_1(X)$ in the Poincaré-type inequality

$$\lambda_1 \,\mathrm{Var}\big(u(X)\big) \le \mathbb{E}\big|\nabla u(X)\big|^2 \tag{1.5}$$

(with $\lambda_1 > 0$), where u is an arbitrary smooth function u on $\mathbb{R}^n$ (cf. Proposition 3.4 below). In one important particular case, the well-known Kannan–Lovász–Simonovits conjecture asserts that λ_1 is bounded away from zero for the whole class of isotropic log-concave probability distributions on the Euclidean space $\mathbb{R}^n$ of any dimension (for short, K-L-S). Conditional on K-L-S, Theorem 1.1 would hence guarantee the $\frac{\log n}{n}$-rate.

COROLLARY 1.2. *Let X be an isotropic random vector in $\mathbb{R}^n$ with a symmetric log-concave distribution. Assuming the K-L-S hypothesis, we have*

$$\mathbb{E}_\theta \rho(F_\theta, \Phi) \le \frac{c \log n}{n}. \tag{1.6}$$

In fact, modulo a logarithmic factor, the conclusion may be reversed in the sense that (1.6) implies $1/\lambda_1 \le c(\log n)^7$; cf. Section 8.

An unconditional statement in the isotropic log-concave case with a standard rate of normal approximation can be obtained by combining the results of [1] and [2] on the concentration of $F_\theta(x)$ around the average distribution function $F(x) = \mathbb{E}_\theta F(x)$ with respect to the variable θ with a recent bound in the thin-shell problem due to Lee and Vempala [23] on the concentration of the Euclidean length $|X|$ about its average value $\mathbb{E}|X|$ (which is in essence equivalent to the closeness of F to the standard normal distribution function Φ). More details are given in Section 7; one then gets

$$\mathbb{E}_\theta \rho(F_\theta, \Phi) \le \frac{c}{\sqrt{n}}\sqrt{\log n}. \tag{1.7}$$

As for the general (not necessarily log-concave) case, the functional $\Lambda(X)$ turns out to be responsible for both, formally different concentration problems. The proof of Theorem 1.1 is based on results for spherical concentration, which have been recently developed in [7]. They provide improved rates of concentration for smooth functions u on the sphere based on the additional information about the Hessian of u. This naturally leads to the definition of $\Lambda(X)$ as introduced above. The "second-order" concentration inequalities on S^{n-1} may also be used to derive large deviation bounds for $\rho(F_\theta, \Phi)$ considered as random variables on the probability space $(S^{n-1}, \mathfrak{s}_{n-1})$. Moreover, one may remove the symmetry assumption as well, by adding to the right-hand side of (1.4) an additional term responsible for third-order correlations between X_k. We refrain from including these somewhat more technical results here and refer the interested reader to [6] for a full account.

As we will see, there exist several natural classes of probability distributions for which a bound on the parameter Λ can be obtained. Some of them are considered in Section 3, after a brief discussion of general properties of Λ and related functionals in Section 2. Some results about the second-order concentration on the sphere are described in Sections 4, which we apply in Section 5 to explore the concentration of characteristic functions of S_θ with respect to the variable θ. In Section 6, relying upon a general Berry–Esseen-type inequality, we finalize the proof of Theorem 1.1. The relationship of Theorem 1.1 with the K-L-S conjecture and a closely related thin-shell problem in the log-concave case are discussed separately in Sections 7–8.

2. Second-order correlation condition and related functionals. As usual, the Euclidean space $\mathbb{R}^n$ is endowed with the canonical norm $|\cdot|$ and the inner product $\langle\cdot,\cdot\rangle$. We start with preliminary remarks on the second-order correlation condition and related functionals.

Let $X = (X_1, \dots, X_n)$ be a random vector in $\mathbb{R}^n$. With the Hilbert–Schmidt norm of a matrix $A = (a_{ij})_{i,j=1}^n$ given by $\|A\|_{\mathrm{HS}} = (\sum a_{ij}^2)^{1/2}$, the definition (1.3) becomes

$$\operatorname{Var}(\langle AX, X\rangle) \le \Lambda \|A\|_{\mathrm{HS}}^2,$$

where we may restrict ourselves to symmetric matrices A only. This description shows that the functional $\Lambda(X)$ is invariant under linear orthogonal transformations of the space $\mathbb{R}^n$ (just as the Hilbert–Schmidt norm).

Related moment and variance-type functionals are

$$M_p = M_p(X) = \sup_{\theta \in S^{n-1}} \big(\mathbb{E}|S_\theta|^p\big)^{1/p} \quad (p \ge 1),$$

$$\sigma_4^2 = \sigma_4^2(X) = \frac{1}{n}\operatorname{Var}(|X|^2).$$

We are mostly interested in the moments M_p with $p = 2$ and $p = 4$. For example, $M_2 = 1$ in the isotropic case, and $\sigma_4 = 0$, if $|X|$ is constant a.s. These functionals can be controlled in terms of Λ, as the following statement shows.

PROPOSITION 2.1. *We have*

$$\text{(a)} \quad M_4^4 \le M_2^4 + \Lambda; \qquad \text{(b)} \quad \sigma_4^2 \le \Lambda.$$

PROOF. Choosing in (1.3) $a_{ij} = \theta_i \theta_j$, $\theta \in S^{n-1}$, we get $\mathrm{Var}(S_\theta^2) \le \Lambda$. Since $\mathbb{E} S_\theta^2 \le M_2^2$, it follows that $\mathbb{E} S_\theta^4 \le M_2^4 + \Lambda$, that is, (a). Putting $a_{ij} = \delta_{ij}$, we also obtain (b). □

In turn, the M_p-moments may be related to the moments of $|X|$. It is easy to see that

$$\big(\mathbb{E}|X|^p\big)^{1/p} \le M_p \sqrt{n}, \quad p \ge 2,$$

while in the isotropic case, there is an opposite inequality $(\mathbb{E}|X|^p)^{1/p} \ge \sqrt{n}$.

The functionals σ_4^2, M_4, and Λ are useful for the estimation of "small" ball probabilities. For example, if $\mathbb{E}|X|^2 = n$, using an independent copy Y of X, we have

$$\mathbb{P}\left\{|X - Y|^2 \le \frac{1}{4} n\right\} \le \frac{A}{n^2}, \quad A = 256\big(M_4^8 + \sigma_4^4\big).$$

This bound was applied in the proof of Lemma 6.1 below (for details, we refer to [9]). Here, by Proposition 2.1(a) in the isotropic case, $A \le c\Lambda^2$, which is also due to the fact that the functional $\Lambda(X)$ is bounded away from zero for $n \ge 2$ (in contrast to σ_4).

PROPOSITION 2.2. *If X is isotropic, then* $\Lambda \ge \frac{n-1}{n}$.

PROOF. Applying the inequality (1.3) to the matrix A with only one nonzero entry on the (i, j)-place, we get

$$\mathrm{Var}(X_i X_j) = \mathbb{E} X_i^2 X_j^2 - \delta_{ij} \le \Lambda.$$

Summing these bounds over all i, j leads to $\mathbb{E}|X|^4 - n \le n^2 \Lambda$. But $\mathbb{E}|X|^4 \ge (\mathbb{E}|X|^2)^2 = n^2$. □

All of the above definitions extend to complex-valued random variables X_i using complex numbers a_{ij} in the definition (1.3) (of course, a_{ij}^2 should be replaced with $|a_{ij}|^2$). Note that, if ξ is a complex-valued random variable, its variance is defined by

$$\mathrm{Var}(\xi) = \mathbb{E}|\xi - \mathbb{E}\xi|^2 = \mathbb{E}|\xi|^2 - |\mathbb{E}\xi|^2.$$

3. Classes of distributions satisfying second-order correlation condition. Here, we provide a few examples where functionals defined above may be easily evaluated or properly estimated. Bounds are attained for the second-order correlation parameter for the following classes of distributions: i.i.d., coordinatewise symmetric, log-concave and coordinatewise symmetric, and probability measures with a spectral gap.

As before, let $X = (X_1, \dots, X_n)$, $n \ge 2$. The case of independent components may be dealt with by simple calculation.

PROPOSITION 3.1. *If the random variables $X_1, \dots, X_n$ are independent and have mean zero*, *then*

$$\sigma_4^2(X) = \frac{1}{n} \sum_{i=1}^n \mathrm{Var}\big(X_i^2\big), \tag{3.1}$$

$$M_2(X) = \max_i \big(\mathbb{E} X_i^2\big)^{1/2}, \qquad \Lambda(X) \le 2 \max_i \mathbb{E} X_i^4. \tag{3.2}$$

Note that equality (3.1) obviously extends to pairwise independent random variables with mean zero. The proof of the bound of $\Lambda(X)$ in (3.2) is similar to the one in Proposition 3.2 below, so we omit it.

Another class of illustrative examples is given by distributions of random vectors X which are equal to $(\varepsilon_1 X_1, \dots, \varepsilon X_n)$ for arbitrary choices of signs $\varepsilon_i = \pm 1$. We call such distributions coordinatewise symmetric, although in the literature they are also called distributions with unconditional basis. This class includes all symmetric product measures on $\mathbb{R}^n$ and corresponds to the case where the components X_i are i.i.d. random variables with symmetric distributions on the line. It is therefore not surprising that many formulas like those in Proposition 3.1 extend to the coordinatewise symmetric distributions. In particular, the first equality in (3.2) is still valid. As for $\Lambda(X)$, it may be essentially reduced to the moment-type functional

$$V(X) = \sup_{\theta \in S^{n-1}} \operatorname{Var}\big(\theta_1 X_1^2 + \cdots + \theta_n X_n^2\big),$$

representing the maximal eigenvalue of the matrix $\{\operatorname{cov}(X_i^2, X_j^2)\}_{i,j=1}^n$.

PROPOSITION 3.2. *Given a random vector $X = (X_1, \dots, X_n)$ in $\mathbb{R}^n$ with a coordinatewise symmetric distribution, we have*

$$V(X) \le \Lambda(X) \le 2 \max_i \mathbb{E} X_i^4 + V(X). \tag{3.3}$$

If additionally the distribution of X is invariant under permutations of coordinates, then

$$\sigma_4^2(X) \le \Lambda(X) \le 2\mathbb{E} X_1^4 + \sigma_4^2(X), \tag{3.4}$$

where the last term $\sigma_4^2(X)$ may be removed when $\operatorname{cov}(X_1^2, X_2^2) \le 0$.

The proof of this proposition is rather elementary, but technical. So, we postpone it to Section 9.

The following subfamily of coordinate-symmetric distributions admits a uniform bound on Λ. Let us recall that a (Borel) probability measure μ on $\mathbb{R}^n$ is called log-concave, if it satisfies the Brunn–Minkowski-type inequality

$$\mu\big(tA + (1-t)B\big) \ge \mu(A)^t \mu(B)^{1-t}, \quad 0 < t < 1,$$

for all nonempty compact sets A and B in $\mathbb{R}^n$, where $tA + (1-t)B = \{tx + (1-t)y : x \in A, y \in B\}$ denotes the Minkowski weighted sum. An equivalent description was given by Borell [12]: the measure μ should be supported on a closed convex set $V \subset \mathbb{R}$ and have a log-concave density p with respect to the Lebesgue measure λ_V on V of the same dimension as V (i.e., $\log p$ is concave). Note that, if μ is isotropic and log-concave, then necessarily V has dimension n, so that μ_V is the (full) Lebesgue measure.

PROPOSITION 3.3. *Assume that the random vector X in $\mathbb{R}^n$ is isotropic and has a coordinatewise symmetric, log-concave distribution. Then*

$$\sigma_4^2(X) \le \Lambda(X) \le c.$$

PROOF. The distribution of the random vector $(|X_1|, \dots, |X_n|)$ has a log-concave, coordinatewise nonincreasing density. By a theorem due to Klartag [20], the following weighted Poincaré-type inequality holds

$$\operatorname{Var}\big(u(X)\big) \le 4\mathbb{E} \sum_{i=1}^n X_i^2 \big(\partial_i u(X_i)\big)^2$$

for any smooth even function u on $\mathbb{R}^n$. Choosing $u(x) = \theta_1 x_1^2 + \cdots + \theta_n x_n^2$ with $\theta_1^2 + \cdots + \theta_n^2 = 1$, we get

$$\mathrm{Var}\big(u(X)\big) \le 16 \sum_{i=1}^n \theta_i^2 \mathbb{E}X_i^4 \le 16 \max_{i \le n} \mathbb{E}X_i^4.$$

In view of Proposition 3.2, we get

$$\Lambda(X) \le 2 \max_i \mathbb{E}X_i^4 + \sup_{\theta \in S^{n-1}} \mathrm{Var}\big(\theta_1 X_1^2 + \cdots + \theta_n X_n^2\big) \le 18 \max_i \mathbb{E}X_i^4.$$

It remains to recall that L^p-norms of random variables with log-concave distributions are equivalent to each other. In particular, for isotropic log-concave X_i's, we have $\mathbb{E}X_i^4 \le c(\mathbb{E}X_i^2)^2 = c$. □

The above subclass may be potentially enlarged by considering the usual Poincaré-type inequality

$$\lambda_1 \mathrm{Var}\big(u(X)\big) \le \mathbb{E}\big|\nabla u(X)\big|^2. \tag{3.5}$$

PROPOSITION 3.4. *Assume that a mean zero random vector X in $\mathbb{R}^n$ satisfies a Poincaré-type inequality with constant $\lambda_1 > 0$. Then $M_2^2(X) \le 1/\lambda_1$. Moreover,*

$$\sigma_4^2(X) \le \Lambda(X) \le \frac{4}{\lambda_1^2},$$

and if X isotropic, then

$$\sigma_4^2(X) \le \Lambda(X) \le \frac{4}{\lambda_1}.$$

PROOF. Applying (3.5) to the linear functions $f(x) = \langle x, \theta\rangle$, $\theta \in S^{n-1}$, we obtain

$$\lambda_1 \mathrm{Var}\big(\langle X, \theta\rangle\big) \le 1.$$

If X has mean zero, the latter means that $M_2^2(X) \le 1/\lambda_1$. In particular, $\mathbb{E}X_j^2 \le \frac{1}{\lambda_1}$. Taking the quadratic function $u(x) = \sum_{i,j=1}^n a_{ij} x_i x_j$ with $a_{ij} = a_{ji}$, we get, by Cauchy's inequality,

$$\mathrm{Var}\left(\sum_{i,j=1}^n a_{ij} X_i X_j\right) \le \frac{4}{\lambda_1} \sum_{i=1}^n \mathbb{E}\left(\sum_{j=1}^n a_{ij} X_j\right)^2 \le \frac{4}{\lambda_1} \sum_{i,j=1}^n a_{ij}^2 \mathbb{E}X_j^2.$$

Hence, the right-hand side does not exceed $4/\lambda_1^2$ subject to $\sum_{i,j=1}^n a_{ij}^2 \le 1$, and thus $\Lambda(X) \le 4/\lambda_1^2$, while $\Lambda(X) \le 4/\lambda_1$ in the isotropic case. □

4. Second-order concentration on the sphere. Concentration of measure on the sphere means that the range of deviations of any Lipschitz function u on the unit sphere S^{n-1} is essentially of order at most $\frac{1}{\sqrt{n}}$, which may be strengthened as the sub-Gaussian stochastic dominance $|u| \le \frac{c}{\sqrt{n}}|Z|$ where Z denotes a standard normal random variable (cf. [22, 26]). More precisely, there is a sub-Gaussian deviation inequality

$$\mathfrak{s}_{n-1}\big\{\big|u(\theta)\big| \ge r\big\} \le 2e^{-(n-1)r^2/2}, \quad r > 0, \tag{4.1}$$

valid whenever the smooth function u has $\mathfrak{s}_{n-1}$-mean zero and Lipschitz seminorm $\|u\|_{\mathrm{Lip}} \le 1$. This may be partly seen from the Poincaré inequality

$$\int |u|^2 \, d\mathfrak{s}_{n-1} \le \frac{1}{n-1} \int |\nabla u|^2 \, d\mathfrak{s}_{n-1} \tag{4.2}$$

in the class of all smooth complex-valued u with $\mathfrak{s}_{n-1}$-mean zero. Although here there is equality for all linear functions, the spherical concentration phenomenon may be strengthened with respect to the dimension n for a wide subclass of smooth functions. In order to facilitate applications, we shall not use sphere intrinsic gradients but use Euclidean notions induced by the standard embedding of the sphere. Here, functions are defined in an open subset of $\mathbb{R}^n$ and their partial derivatives are understood in the usual sense. We denote by $\nabla^2 u(x)$ the Hessian, that is, the $n \times n$ matrix of second-order partial derivative $\partial_{ij} u(x)$, and by I_n the identity $n \times n$ matrix. The next proposition summarizes several recent results from [7].

PROPOSITION 4.1. *Suppose that a real-valued function u is defined and C^2-smooth in some neighborhood of S^{n-1}. If u is orthogonal to all affine functions in $L^2(\mathfrak{s}_{n-1})$, then*

$$\int u^2 \, d\mathfrak{s}_{n-1} \le \frac{5}{(n-1)^2} \int \|\nabla^2 u - a I_n\|_{\mathrm{HS}}^2 \, d\mathfrak{s}_{n-1} \tag{4.3}$$

for any $a \in \mathbb{R}$. Moreover, if $\|\nabla^2 u - a I_n\| \le 1$ uniformly on S^{n-1} for the operator norm, and the second integral in (4.2) is bounded by b, then

$$\int \exp\left\{\frac{n-1}{2(1+4b)}|u|\right\} d\mathfrak{s}_{n-1} \le 2. \tag{4.4}$$

By Markov's inequality, (4.4) yields a corresponding large deviation bound, which may be stated informally as a subexponential stochastic dominance $|u| \le c_b(\frac{1}{\sqrt{n}} Z)^2$. In particular, this means that the deviations of u are of order at most $1/n$.

The second-order Poincaré-type inequality (4.3) obviously extends to all complex-valued u that are orthogonal to all affine functions on the sphere. In this case, (4.4) may be applied separately to the real and imaginary part of u, which results in

$$\int \exp\left\{\frac{n-1}{4(1+4b)}|u|\right\} d\mathfrak{s}_{n-1} \le 2, \tag{4.5}$$

assuming that $\|\nabla^2 u - a I_n\| \le 1$ on S^{n-1} for some $a \in \mathbb{C}$.

5. Concentration of characteristic functions. Given an isotropic random vector $X = (X_1, \dots, X_n)$ in $\mathbb{R}^n$, introduce the smooth functions

$$u_t(\theta) = f_\theta(t) = \mathbb{E} e^{it\langle X, \theta\rangle}, \quad \theta \in \mathbb{R}^n, \tag{5.1}$$

where $t \ne 0$ serves as a parameter. For any fixed θ, $t \to f_\theta(t)$ represents the characteristic function of the weighted sum $S_\theta = \langle X, \theta\rangle$ with distribution function F_θ, while the $\mathfrak{s}_{n-1}$-mean of u_t,

$$f(t) = \mathbb{E}_\theta f_\theta(t) = \mathbb{E}_\theta \mathbb{E} e^{it\langle X, \theta\rangle},$$

is the characteristic function of the average distribution function

$$F(x) = \int F_\theta(x) \, d\mathfrak{s}_{n-1}(\theta) = \mathbb{E}_\theta \mathbb{P}\{S_\theta \le x\}, \quad x \in \mathbb{R}. \tag{5.2}$$

Let us recall that we use $\mathbb{E}_\theta$ to denote integrals over the unit sphere with respect to the uniform measure $\mathfrak{s}_{n-1}$.

In order to study deviations of the functions u_t from their $\mathfrak{s}_{n-1}$-means $f(t)$ on S^{n-1}, one may start from the Poincare inequality (4.2). Indeed, differentiating the equality (5.1), we get that, for any $\theta' \in S^{n-1}$,

$$\langle \nabla u_t(\theta), \theta'\rangle = it \mathbb{E}\langle X, \theta'\rangle e^{it\langle X, \theta\rangle},$$

which, by Cauchy's inequality, implies

$$\left|\langle \nabla u_t(\theta), \theta' \rangle\right|^2 \le t^2 \mathbb{E}\langle X, \theta' \rangle^2 = t^2.$$

Taking the supremum over all θ', it follows that $|\nabla u_t(\theta)| \le |t|$, which means that u_t has a Lipschitz seminorm $\|u_t\|_{\mathrm{Lip}} \le |t|$ (on the whole space $\mathbb{R}^n$). Therefore, by (4.2),

$$\mathbb{E}_\theta \left| f_\theta(t) - f(t) \right|^2 \le \frac{t^2}{n-1}. \tag{5.3}$$

Thus, the deviations of $f_\theta(t)$ from $f(t)$ with respect to $\theta \in S^{n-1}$ are of order at most $1/\sqrt{n}$—a property which may potentially be transferred to the analogous statement about the deviations of the distribution functions F_θ from F in the sense of certain weak metrics.

In order to obtain better rates, we employ Proposition 4.1, assuming additionally that the random vector X is symmetric and satisfies a second-order correlation condition (1.3) with parameter Λ. To apply the bounds (4.3) and (4.5), we need to choose a suitable value $a \in \mathbb{C}$ and estimate the operator norm $\|\nabla^2 u_t - aI_n\|$ and the Hilbert–Schmidt norm $\|\nabla^2 u_t - aI_n\|_{\mathrm{HS}}$. First note that, by differentiation of (5.1), the Hessian of u_t is given by

$$\left[\nabla^2 u_t(\theta)\right]_{jk} = \frac{\partial^2}{\partial \theta_j \partial \theta_k} f_\theta(t) = -t^2 \mathbb{E} X_j X_k e^{it\langle X, \theta \rangle}$$

for any fixed $t \in \mathbb{R}$. Hence, a good choice could be $a = -t^2 f(t)$ in order to balance the diagonal elements in the matrix of second derivatives of u_t. For any vector $v \in \mathbb{C}^n$ with complex components, using the canonical inner product in the complex n-space, we have

$$\langle \nabla^2 u_t(\theta) v, v \rangle = -t^2 \mathbb{E} \left| \langle X, v \rangle \right|^2 e^{it\langle X, \theta \rangle}.$$

Hence, with this choice of a, by the isotropy assumption,

$$\left| \langle (\nabla^2 u_t(\theta) - aI_n) v, v \rangle \right| \le t^2 \mathbb{E} \left| \langle X, v \rangle \right|^2 + |a| |v|^2 \le 2t^2, \quad |v| = 1.$$

This bound insures that

$$\left\| \nabla^2 u_t(\theta) - aI_n \right\| \le 2t^2. \tag{5.4}$$

In addition, putting $a(\theta) = -t^2 f_\theta(t)$, we have

$$\begin{aligned}
\left\| \nabla^2 u_t(\theta) - a(\theta) I_n \right\|_{\mathrm{HS}}^2 &= \sum_{j,k=1}^n \left| \nabla^2 u_t(\theta)_{jk} - a(\theta)\delta_{jk} \right|^2 \\
&= \sup \left| \sum_{j,k=1}^n z_{jk} \left(\nabla^2 u_t(\theta)_{jk} - a(\theta)\delta_{jk} \right) \right|^2 \\
&= t^4 \sup \left| \mathbb{E} \sum_{j,k=1}^n z_{jk} (X_j X_k - \delta_{jk}) e^{it\langle X, \theta \rangle} \right|^2 \\
&\le t^4 \sup \mathbb{E} \left| \sum_{j,k=1}^n z_{jk} (X_j X_k - \delta_{jk}) \right|^2,
\end{aligned}$$

where the supremum is running over all complex numbers z_{jk} such that $\sum_{j,k=1}^n |z_{jk}|^2 = 1$. But, under this constraint, due to the second-order correlation condition, the last expectation is bounded by Λ, so that

$$\left\| \nabla^2 u_t(\theta) - a(\theta) I_n \right\|_{\mathrm{HS}}^2 \le \Lambda t^4 \tag{5.5}$$

for all θ. On the other hand, by (5.3), and using $\frac{n}{n-1} \le 2$,

$$\mathbb{E}_\theta \big\| \big(a(\theta) - a\big) I_n \big\|_{\mathrm{HS}}^2 = n t^4 \mathbb{E}_\theta \big| f_\theta(t) - f(t) \big|^2 \le 2t^6. \tag{5.6}$$

The two bounds give

$$\mathbb{E}_\theta \big\| \nabla^2 u_t(\theta) - a I_n \big\|_{\mathrm{HS}}^2 \le 2\Lambda t^4 + 4t^6. \tag{5.7}$$

Note that (5.6) is worse in comparison with (5.5) in the variable t. Nevertheless, applying the second-order Poincaré-type inequality, it is possible to improve the resulting inequality (5.7) for reasonably long t-intervals. Since the distribution of X is symmetric about the origin, the characteristic functions $f_\theta(t)$ are even with respect to θ, that is, $f_{-\theta}(t) = f_\theta(t)$. Hence, they are orthogonal in the Hilbert space $L^2(\mathfrak{s}^{n-1})$ to all linear functions on the sphere. Thus, the conditions of Proposition 4.1 are fulfilled for the function $u = u_t$, and using (5.7), the inequality (4.3) gives

$$\mathbb{E}_\theta \big| f_\theta(t) - f(t) \big|^2 \le \frac{5}{(n-1)^2} \big(2\Lambda t^4 + 4t^6\big). \tag{5.8}$$

This bound allows us to improve (5.6) to the form

$$\mathbb{E}_\theta \big\| \big(a(\theta) - a\big) I_n \big\|_{\mathrm{HS}}^2 \le n t^4 \frac{5}{(n-1)^2} \big(2\Lambda t^4 + 4t^6\big) \le \frac{c\Lambda t^4}{n} \big(t^4 + t^6\big).$$

Combining this with (5.5), we therefore obtain that

$$\mathbb{E}_\theta \big\| \nabla^2 u_t(\theta) - a I_n \big\|_{\mathrm{HS}}^2 \le 2\Lambda t^4 + \frac{c\Lambda t^4}{n} \big(t^4 + t^6\big).$$

Here, similar to (5.5), the right-hand side is at most $c\Lambda t^4$ in the interval $|t| \le n^{1/6}$. To enlarge the t-interval, one may repeat the argument. By (4.3),

$$\mathbb{E}_\theta \big| f_\theta(t) - f(t) \big|^2 \le \frac{5}{(n-1)^2} \bigg(2\Lambda t^4 + \frac{c\Lambda t^4}{n} \big(t^4 + t^6\big)\bigg).$$

This bound improves (5.6) to the form

$$\mathbb{E}_\theta \big\| \big(a(\theta) - a\big) I_n \big\|_{\mathrm{HS}}^2 \le n t^4 \frac{5}{(n-1)^2} \bigg(2\Lambda t^4 + \frac{c\Lambda t^4}{n} \big(t^4 + t^6\big)\bigg) \le c\Lambda t^4,$$

where in the last inequality we assume that $|t| \le n^{1/5}$. Combining this with (5.5), we get

$$\mathbb{E}_\theta \big\| \nabla^2 u_t(\theta) - a I_n \big\|_{\mathrm{HS}}^2 \le c\Lambda t^4.$$

In view of (4.3), this already gives the inequality (5.9) below.

To get a stronger deviation inequality, note that, by (5.4), the conditions of Proposition 4.1 (in its second part) are fulfilled for the function

$$u(\theta) = \frac{1}{2t^2} \big(f_\theta(t) - f(t)\big), \quad \theta \in \mathbb{R}^n, 0 < |t| \le n^{1/5},$$

with parameter $b = c\Lambda$. Applying (4.5), we arrive at the following.

COROLLARY 5.1. *Let X be an isotropic random vector in $\mathbb{R}^n$ with a symmetric distribution and finite constant Λ. Then the characteristic functions $f_\theta(t) = \mathbb{E}e^{it\langle X,\theta\rangle}$ satisfy*

$$\mathbb{E}_\theta \big| f_\theta(t) - f(t) \big|^2 \le \frac{c}{n^2} \Lambda t^4 \tag{5.9}$$

whenever $0 < |t| \le n^{1/5}$. Moreover,

$$\mathbb{E}_\theta \exp\bigg\{ \frac{n}{c\Lambda t^2} \big| f_\theta(t) - f(t) \big| \bigg\} \le 2. \tag{5.10}$$

As we have seen, removing the constraint $|t| \le n^{1/5}$, (5.9) may be replaced with a weaker inequality (5.8). When applying the latter to the estimation of $\rho(F_\theta, F)$ via Lemma 6.1 below, we would gain an additional $\log n$ factor in Theorem 1.1. Continuing the iteration process in the proof of Corollary 5.1, one may state (5.9)–(5.10) in the intervals $|t| \le n^\alpha$ with any fixed $\alpha < \frac{1}{4}$. The power $\alpha = \frac{1}{5}$ turns out to be useful in the study of large deviations of $\rho(F_\theta, \Phi)$ above the mean (which we however do not discuss here).

6. Proof of Theorem 1.1. Based on the deviation inequalities (5.9)–(5.10), Fourier analytic tools yield bounds for the closeness of the distribution functions F_θ to the s_{n-1}-mean distribution function F defined in (5.2). The following Berry–Esseen-type bound can be found in [9] (cf. Lemma 6.2), which we state in the case $p = 2$.

LEMMA 6.1. *Suppose that a random vector X in $\mathbb{R}^n$ has a finite moment of order* 4, *with* $\mathbb{E}|X|^2 = n$. *Then, for all $T \ge T_0 > 0$,*

$$\begin{aligned} c\mathbb{E}_\theta \rho(F_\theta, F) \le &\int_0^{T_0} \mathbb{E}_\theta \big| f_\theta(t) - f(t) \big| \frac{dt}{t} \\ &+ \frac{M_4^4 + \sigma_4^2}{n} \left(1 + \log \frac{T}{T_0}\right) + \frac{1}{T} + e^{-T_0^2/16}. \end{aligned} \tag{6.1}$$

PROOF OF THEOREM 1.1. Applying Propositions 2.1–2.2 and using the isotropy assumption, we have $M_4^4 + \sigma_4^2 \le 1 + 2\Lambda \le 4\Lambda$. Hence, (6.1) yields

$$\begin{aligned} c_1\mathbb{E}_\theta \rho(F_\theta, F) \le &\int_0^{T_0} \mathbb{E}_\theta \big| f_\theta(t) - f(t) \big| \frac{dt}{t} \\ &+ \frac{\Lambda}{n} \left(1 + \log \frac{T}{T_0}\right) + \frac{1}{T} + e^{-T_0^2/16}. \end{aligned}$$

Here, the integrand may be estimated by virtue of (5.9), and then we get

$$c_2\mathbb{E}_\theta \rho(F_\theta, F) \le \frac{1}{n} T_0^2 \sqrt{\Lambda} + \frac{1}{n}\left(1 + \log \frac{T}{T_0}\right)\Lambda + \frac{1}{T} + e^{-T_0^2/16},$$

provided that $T_0 \le n^{1/5}$. As a natural choice, take $T_0 = 5\sqrt{\log n}$, $T = 5n$ (assuming that n is large enough), which leads to the bound

$$c_3\mathbb{E}_\theta \rho(F_\theta, F) \le \frac{\log n}{n} \Lambda. \tag{6.2}$$

We finally refer to [8], Theorem 1.1 (cf. also [9], Corollary 4.2) where the estimate

$$\rho(F, \Phi) \le c\frac{1 + \sigma_4^2}{n} \tag{6.3}$$

was derived. Using $\sigma_4^2 \le \Lambda$ and combining (6.2) with the triangle inequality for ρ, we arrive at the desired inequality (1.4). $\square$

REMARK 6.2. Under proper moment assumptions and using the spherical deviation inequality (5.10), one may derive large deviation bounds for $\rho(F_\theta, \Phi)$ as well. In particular, suppose that

$$\mathbb{E}e^{|S_\theta|/\beta} \le 2 \tag{6.4}$$

for all $\theta \in S^{n-1}$ with some $\beta > 0$. Then, in the setting of Theorem 1.1,

$$\mathfrak{s}_{n-1}\left\{\rho(F_\theta, F) \ge \frac{c\log n}{n}(\Lambda + \beta^4)r\right\} \le 3\exp\{-r^{1/8}\}, \quad r \ge 0.$$

In other words, with high $\mathfrak{s}_{n-1}$-probability,

$$\rho(F_\theta, F) \le \frac{c(\log n)^9}{n}(\Lambda + \beta^4).$$

For details, we refer the interested reader to [6].

7. The log-concave case. Specializing to the class of isotropic log-concave distributions on $\mathbb{R}^n$, first let us comment on the unconditional statement with a standard rate of normal approximation as indicated in the inequality (1.7). If the isotropic random vector X has a uniform distribution over a symmetric convex body in $\mathbb{R}^n$, it was shown by Anttila, Ball and Perissinaki that

$$\mathfrak{s}_{n-1}\{\rho(F_\theta, F) \ge r\} \le c_1\sqrt{n}\log n e^{-c_2 n r^2}, \quad r > 0 \tag{7.1}$$

(actually with $c_2 = 50$; cf. [1]). With a different argument, this inequality has been extended to arbitrary isotropic log-concave distributions in [2]. In both papers, as a main step, it was observed that, for every point $x \in \mathbb{R}$, the function $u(\theta) = F_\theta(x)$ has a bounded Lipschitz seminorm on the unit sphere, so that one may apply the spherical concentration inequality (4.1), leading to

$$\mathfrak{s}_{n-1}\{|F_\theta(x) - F(x)| \ge r\} \le 2e^{-cnr^2}, \quad r > 0.$$

Since $\rho(F_\theta, F) \le 1$, (7.1) readily yields an upper bound

$$\mathbb{E}_\theta \rho(F_\theta, F) \le c\sqrt{\frac{\log n}{n}}.$$

Combining it with (6.3) and applying the triangle inequality for the metric ρ, we therefore obtain the normal approximation on average in the form of the relation

$$c\mathbb{E}_\theta \rho(F_\theta, \Phi) \le \sqrt{\frac{\log n}{n}} + \frac{\sigma_4^2}{n}. \tag{7.2}$$

It remains to involve the bound $\sigma_4^2 \le c\sqrt{n}$, which was recently derived by Lee and Vempala [23], and then we arrive at (1.7).

A thin-shell conjecture, raised in [11], asserts that the functional $\sigma_4^2(X)$, or equivalently $\mathrm{Var}(|X|)$, is actually bounded by a dimension-free (and thus universal) constant over the whole class of isotropic log-concave random vectors X in $\mathbb{R}^n$. Specializing to the convex body case, a similar concentration hypothesis was also suggested in [1]. It states that the deviation inequality

$$\mathbb{P}\left\{\left|\frac{|X|}{\sqrt{n}} - 1\right| \ge \varepsilon_n\right\} \le \varepsilon_n$$

holds true with $\varepsilon_n \le c(\log n)/\sqrt{n}$. The boundedness of σ_4^2 allows one to take a slightly thinner shell with $\varepsilon_n = c/\sqrt{n}$. Anyhow, the bound (7.2) subject to the thin-shell conjecture still leads to the standard normal approximation as in (1.7).

Note that, by the Poincaré-type inequality (1.5) applied with $u(x) = |x|^2$, one gets $\sigma_4^2 \le 4/\lambda_1$, so that the thin-shell conjecture is formally weaker than the K-L-S (which is further shown in Proposition 3.4). On the other hand, recently Eldan [15] has developed a new localization technique, in essence reducing the stronger hypothesis to the weaker one modulo a logarithmic factor. It is is therefore possible to state Corollary 1.2 alternatively as follows.

COROLLARY 7.1. *Let X be an isotropic random vector in $\mathbb{R}^n$ with a symmetric log-concave distribution. Assuming that the thin-shell conjecture is true, we have*

$$\mathbb{E}_\theta \rho(F_\theta, \Phi) \le \frac{c(\log n)^3}{n}.$$

PROOF. Combining Theorem 1.1 with Proposition 3.4, we get

$$\mathbb{E}_\theta \rho(F_\theta, \Phi) \le \frac{c}{\lambda_{1,n} n} \log n, \tag{7.3}$$

where $\lambda_1 = \lambda_{1,n}$ is the smallest spectral gap in the Poincaré-type inequality over the class of all isotropic log-concave probability measures on $\mathbb{R}^n$. Assuming the K-L-S conjecture, $\lambda_{1,n}$ is bounded away from zero, which thus leads to the inequality (1.6) of Corollary 1.2. Within the same class, this quantity may be related to the largest value $\sigma_{4,n}^2 = \sup_X \sigma_4^2(X)$. Namely, as shown by Eldan [15],

$$\frac{1}{\lambda_{1,n}} \le c \log n \sum_{k=1}^n \frac{\sigma_{4,k}^2}{k}. \tag{7.4}$$

In particular, the bound of the form $\sigma_{4,n}^2 \le c_1 n^\alpha$ $(0 \le \alpha \le 1)$ implies that

$$\lambda_{1,n}^{-1} \le c\eta_\alpha(n) \tag{7.5}$$

with $\eta_\alpha(n) = \frac{c_1}{\alpha} n^\alpha \log n$ in case $\alpha > 0$ and $\eta_0(n) = 3c_1(\log n)^2$. It remains to apply (7.5) in (7.3) with $\alpha = 0$. $\square$

8. From the normal approximation to the shin shell. To refine the relationship between the central limit theorem and the thin-shell problem, let us complement Corollary 7.1 by the following general statement involving the maximal ψ_1-norm of linear functionals of X.

PROPOSITION 8.1. *Let X be a random vector in $\mathbb{R}^n$ with $\mathbb{E}|X|^2 = n$, satisfying the moment condition* (6.4) *with some $\beta > 0$. Then*

$$c\sigma_4^2(X) \le n(\beta \log n)^4 \mathbb{E}_\theta \rho(F_\theta, \Phi) + \frac{\beta^4}{n^4} + 1. \tag{8.1}$$

In the isotropic log-concave case, the condition (6.4) is fulfilled with some absolute constant β (by the well-known Borell's lemma 3.1 in [12]), and this simplifies (8.1) to

$$c\sigma_4^2(X) \le n(\log n)^4 \mathbb{E}_\theta \rho(F_\theta, \Phi) + 1.$$

Hence, the potential property

$$\mathbb{E}_\theta \rho(F_\theta, \Phi) \le \frac{c \log n}{n} \tag{8.2}$$

as in Corollary 1.2 would imply that

$$\sigma_4^2(X) \le c(\log n)^5, \tag{8.3}$$

assuming additionally that the distribution of X is symmetric about zero. But the symmetry condition may easily be dropped. Indeed, define $X' = (X - Y)/\sqrt{2}$, where Y is an independent copy of a random vector X with an isotropic log-concave distribution on $\mathbb{R}^n$. Then the distribution of X' is isotropic, log-concave and symmetric about zero. Moreover,

$$\begin{aligned}\sigma_4^2(X') &= \frac{1}{2n} \operatorname{Var}\big(|X|^2 + |Y|^2 - 2\langle X, Y\rangle\big) \\ &= \frac{1}{2n} \operatorname{Var}\big(|X|^2\big) + \frac{1}{2n} \operatorname{Var}\big(|Y|^2\big) + \frac{2}{n} \mathbb{E}\langle X, Y\rangle^2 = \sigma_4^2(X) + 2.\end{aligned}$$

Hence, once (8.3) is true for the random vector X', it continues to hold for X as well (with other constant).

Note also that, applying Eldan's inequality (7.4) together with (8.3), from the normal approximation (8.2) we get

$$\lambda_{1,n}^{-1} \le c(\log n)^7.$$

PROOF OF PROPOSITION 8.1. In view of the triangle inequality $\rho(F, \Phi) \le \mathbb{E}_\theta \rho(F_\theta, \Phi)$, it is sufficient to derive (8.1) for $\rho(F, \Phi)$ in place of $\mathbb{E}_\theta \rho(F_\theta, \Phi)$. This means that we need in essence to reverse the inequality (6.3) by using (6.4). To this aim, let us rewrite the definition (5.2) as

$$F(x) = \mathbb{P}\{|X|\theta_1 \le x\}, \quad x \in \mathbb{R},$$

where we assume that X and $\theta = (\theta_1, \dots, \theta_n) \in S^{n-1}$ (as a random vector uniformly distributed on the sphere) are independent. This description yields

$$\int_{-\infty}^{\infty} x^4 \, dF(x) = \mathbb{E}|X|^4 \mathbb{E}_\theta \theta_1^4 = (n^2 + \sigma_4^2 n) \frac{3}{n(n+2)},$$

or equivalently

$$\int_{-\infty}^{\infty} x^4 \, dF(x) - \int_{-\infty}^{\infty} x^4 \, d\Phi(x) = \frac{3}{n+2}(\sigma_4^2 - 2), \tag{8.4}$$

where $\sigma_4^2 = \sigma_4^2(X)$. On the other hand, it follows from (6.4) that

$$\int_{-\infty}^{\infty} e^{|x|/\beta} \, dF(x) \le 2.$$

Using $t^2 \le 4e^{-2}e^t$ $(t \ge 0)$ together with the property $\int_{-\infty}^{\infty} x^2 \, dF(x) = \frac{1}{n}\mathbb{E}|X|^2 = 1$, we have $\beta \ge e/\sqrt{8}$, which can be used to derive the bounds

$$1 - \Phi(x) \le \frac{1}{2} e^{-x^2/2} \le 2e^{-x/\beta}, \quad x \ge 0.$$

In addition, by Markov's inequality, $F(-x) + (1 - F(x)) \le 2e^{-x/\beta}$, so that

$$|F(-x) - \Phi(-x)| + |F(x) - \Phi(x)| \le 6e^{-x/\beta}$$

for all $x \ge 0$. Hence, integrating by parts, we see that, for any $T \ge 6\beta$, the left-hand side of (8.4) does not exceed in absolute value

$$\begin{aligned} &4\int_{-T}^{T} |x|^3 |F(x) - \Phi(x)| \, dx + 24 \int_T^{\infty} x^3 e^{-x/\beta} \, dx \\ &\quad \le 2T^4 \rho(F, \Phi) + 48\beta T^3 e^{-T/\beta}. \end{aligned}$$

Choosing $T = 9\beta \log n$ and recalling (8.4), we get

$$\sigma_4^2 \le 6 + cn\left[\beta^4 (\log n)^4 \rho(F, \Phi) + \frac{\beta^4}{n^9}(\log n)^3\right]. \qquad \square$$

9. Proof of Proposition 3.2. The lower bound on Λ in (3.3) immediately follows from (1.3) by choosing the coefficients to be of the form $a_{ij} = \theta_i \delta_{ij}$. For the upper bound, put $v_i^2 = \mathbb{E}X_i^2$ and define

$$X_{ij}^{(2)} = X_i X_j - \mathbb{E}X_i X_j = X_i X_j - \delta_{ij} v_i^2.$$

The covariances of these mean zero random variables are given by

$$\begin{aligned} \mathbb{E}X_{ij}^{(2)} X_{kl}^{(2)} &= \mathbb{E}(X_i X_j - \delta_{ij} v_i^2) X_k X_l \\ &= \mathbb{E}X_i X_j X_k X_l - \delta_{ij}\delta_{kl} v_i^2 v_k^2. \end{aligned} \tag{9.1}$$

Case 1: $i \neq j$. By the symmetry with respect to the coordinate axes, the right-hand side of (9.1) is vanishing unless $(i, j) = (k, l)$ or $(i, j) = (l, k)$. In both cases, it is equal to

$$\mathbb{E}X_{ij}^{(2)} X_{ij}^{(2)} = \mathbb{E}X_{ij}^{(2)} X_{ji}^{(2)} = \mathbb{E}X_i^2 X_j^2.$$

Case 2: $i = j$. The right-hand side in (9.1) is nonzero only when $k = l$.

Case 2(a): $i = j$, $k = l$, $i \neq k$. The right-hand side in (9.1) is equal to

$$\mathbb{E}X_{ii}^{(2)} X_{kk}^{(2)} = \mathbb{E}X_i^2 X_k^2 - \mathbb{E}X_i^2 \mathbb{E}X_k^2 = \operatorname{cov}(X_i^2, X_k^2).$$

Case 2(b): $i = j = k = l$. The right-hand side is equal to

$$\mathbb{E}X_{ii}^{(2)} X_{ii}^{(2)} = \mathbb{E}X_i^4 - \mathbb{E}X_i^2 \mathbb{E}X_i^2 = \operatorname{Var}(X_i^2).$$

In both subcases, $\mathbb{E}X_{ii}^{(2)} X_{kk}^{(2)} = \operatorname{cov}(X_i^2, X_k^2)$. Therefore, for any collection of real numbers a_{ij} such that $a_{ij} = a_{ji}$ and $\sum_{i,j=1}^n a_{ij}^2 = 1$,

$$\begin{aligned} \operatorname{Var}\left(\sum_{i,j=1}^n a_{ij} X_i X_j\right) &= \sum_{i,j=1}^n \sum_{k,l=1}^n a_{ij} a_{kl} \mathbb{E}X_{ij}^{(2)} X_{kl}^{(2)} \\ &= 2\sum_{i\neq j} a_{ij}^2 \mathbb{E}X_i^2 X_j^2 + \sum_{i,k} a_{ii} a_{kk} \operatorname{cov}(X_i^2, X_k^2). \end{aligned}$$

Here, the first sum on the right-hand side does not exceed

$$\max_{i\neq j} \mathbb{E}X_i^2 X_j^2 \sum_{i\neq j} a_{ij}^2 \leq \max_i \mathbb{E}X_i^4 \sum_{i\neq j} a_{ij}^2 \leq \max_i \mathbb{E}X_i^4$$

(by applying Cauchy's inequality). As for the second sum, it does not exceed $V(X)$, and we obtain

$$\Lambda(X) \leq 2 \max_{i\neq j} \mathbb{E}X_i^2 X_j^2 + V(X), \tag{9.2}$$

from which (3.3) follows immediately.

As for (3.4), recall that the first inequality always holds; cf. Proposition 2.1. For the second one, let us note that

$$\sigma_4^2(X) = \frac{1}{n} \operatorname{Var}(|X|^2) = \operatorname{Var}(X_1^2) + (n-1)\operatorname{cov}(X_1^2, X_2^2) \tag{9.3}$$

and that, for any $\theta = (\theta_1, \dots, \theta_n) \in S^{n-1}$,

$$\operatorname{Var}(\theta_1 X_1^2 + \cdots + \theta_n X_n^2) = \left(\sum_{i=1}^n \theta_i\right)^2 \operatorname{cov}(X_1^2, X_2^2) - \operatorname{cov}(X_1^2, X_2^2) + \operatorname{Var}(X_1^2).$$

Here, in the case $\mathrm{cov}(X_1^2, X_2^2) \geq 0$, the right-hand side is maximized for equal coefficients, and recalling (9.3), we then get

$$\mathrm{Var}(\theta_1 X_1^2 + \cdots + \theta_n X_n^2) \leq (n-1)\,\mathrm{cov}(X_1^2, X_2^2) + \mathrm{Var}(X_1^2) = \sigma_4^2(X).$$

Hence, (9.2) implies (3.4). In the case $\mathrm{cov}(X_1^2, X_2^2) \leq 0$, we similarly conclude that

$$\begin{aligned}\mathrm{Var}(\theta_1 X_1^2 + \cdots + \theta_n X_n^2) &\leq -\mathrm{cov}(X_1^2, X_2^2) + \mathrm{Var}(X_1^2) \\ &= \mathbb{E}X_1^4 - \mathbb{E}X_1^2 X_2^2,\end{aligned}$$

which means that $V(X) \leq \mathbb{E}X_1^4 - \mathbb{E}X_1^2 X_2^2$. Thus, by (9.2),

$$\Lambda(X) \leq 2\mathbb{E}X_1^2 X_2^2 + V(X) \leq \mathbb{E}X_1^2 X_2^2 + \mathbb{E}X_1^4 \leq 2\mathbb{E}X_1^4.$$

Hence, (3.4) follows in this case as well even without the $\sigma_4^2(X)$-functional.

10. Historical remarks. Finally, let us give a short overview on results related to Theorem 1.1 (some account can also be found in the book [13]). It is natural to distinguish between two types of results.

10.1. *Deviations of F_θ from the mean distribution F in different metrics.* The paper by Sudakov [30] starts with the hypothesis

$$\mathbb{E}\left(\sum_{i=1}^n a_i X_i\right)^2 \leq M_2^2 \sum_{i=1}^n a_i^2, \quad a_i \in \mathbb{R},$$

which may be called a first-order correlation condition. Here, an optimal value $M_2 = M_2(X)$ is the same functional we considered in Section 2; equivalently, M_2^2 represents the maximal eigenvalue of the correlation operator for the random vector X. As was shown in [30], if M_2 is bounded, and n is large, then most of F_θ are close to the average distribution F in the sense of the Kantorovich or L^1-distance

$$W_1(F_\theta, F) = \|F_\theta - F\|_{L^1(\mathbb{R},dx)} = \int_{-\infty}^{\infty} \left|F_\theta(x) - F(x)\right| dx.$$

A closely related observation was also made by Diaconis and Freedman [14]. A somewhat different scheme, in which the coefficient vectors are drawn from the Gaussian measure μ_n on $\mathbb{R}^n$ with mean zero and covariance matrix $\frac{1}{n} I_n$, was also considered by Nagaev [27] and von Weizsäcker [31]. In particular, assuming that $M_1 = 1$, [27] contains a quantitative bound

$$\left(\int \|F_\theta - F\|^2_{L^2(\mathbb{R},dx)}\, d\mu_n(\theta)\right)^{1/2} \leq \frac{1}{(\pi n)^{1/4}} \tag{10.1}$$

for the L^2-distance between the distribution functions. When the coefficients have a special structure, similar phenomena were considered in [3, 10].

Returning to the spherical measure $\mathfrak{s}_{n-1}$, the rate as in (10.1) is achieved for the Lévy distance as well. More precisely, there is a general bound

$$\mathbb{E}_\theta L(F_\theta, F) \leq c\frac{\log n}{n^{1/4}},$$

where the constant c depends on M_1 only; cf. [5]. Large deviation bounds on $L(F_\theta, F)$ were given in [2] in the isotropic case. As was already discussed in Section 7, the rate and deviation bounds may be essentially improved and be stated for the stronger Kolmogorov distance, when the random vector X has an isotropic log-concave distribution.

Quantitative variants of Sudakov's theorem for W_1 were studied in [4], where it was shown that, for any $p > 1$,

$$\mathbb{E}_\theta W_1(F_\theta, F) \le \frac{12p}{p-1} M_p n^{-\frac{p-1}{2p}}.$$

The rate is thus approaching $1/\sqrt{n}$ for growing p. Under a stronger assumption (6.4), the above inequality easily implies

$$\mathbb{E}_\theta W_1(F_\theta, F) \le c\beta \frac{\log n}{\sqrt{n}}.$$

Here, the logarithmic term may be removed, if X has an isotropic log-concave distribution (by virtue of Proposition 3.1 in [2]). Note that, in all these results, the rates are not better than a multiple of $1/\sqrt{n}$.

10.2. *Deviations of F_θ from the standard normal distribution function* Φ. To study the approximation of F_θ by the standard normal distribution function for most of θ's, one is led to determine rates for the distance $\rho(F, \Phi)$, which may be reduced to the estimation of $\sigma_4^2(X)$ (via relation (6.3)). In fact, the control of the two functionals, M_4 and σ_4, is sufficient to guarantee a standard rate of normal approximation for F_θ on average. As was shown in [9], we have

$$\mathbb{E}_\theta \rho(F_\theta, \Phi) \le c\big(M_4^3 + \sigma_4^{3/2}\big)\frac{1}{\sqrt{n}}.$$

Note that Theorem 1.1 essentially improves this bound as long as Λ is of the same order as M_4 and σ_4. However, whether or not Λ and even σ_4 is bounded might be a difficult problem for some classes of distributions on $\mathbb{R}^n$ such as the class of isotropic log-concave probability measures. For this class, the property that $\rho(F_\theta, \Phi)$ is small for most of θ (when n is large) was first established by Klartag [18]. In particular, $\mathbb{E}_\theta \rho(F_\theta, \Phi) \le \varepsilon_n \to 0$ as $n \to \infty$ uniformly over the class. For further refinements in this direction, see [16, 17, 19].

There is also a number of results, where the coefficients are fixed, and $\rho(F_\theta, \Phi)$ are bounded by a quantity, which depends on θ as well; cf., for example, [24, 25]. One striking result due to Klartag [20] should be mentioned: If the random vector X in $\mathbb{R}^n$ is isotropic and has a coordinatewise symmetric, isotropic log-concave distribution, then

$$\rho(F_\theta, \Phi) \le c \sum_{k=1}^n \theta_k^4.$$

Moreover, a similar bound holds true for the stronger total variation distance. This is of course more precise in comparison with the average estimate $\mathbb{E}_\theta \rho(F_\theta, \Phi) \le c/n$.

Acknowledgments. The authors would like to thank the two referees for careful reading of the manuscript and valuable comments.

This work was supported by NSF Grant DMS-1855575, the Simons Foundation and CRC 1283 at Bielefeld University.

REFERENCES

[1] ANTTILA, M., BALL, K. and PERISSINAKI, I. (2003). The central limit problem for convex bodies. *Trans. Amer. Math. Soc.* **355** 4723–4735. MR1997580 https://doi.org/10.1090/S0002-9947-03-03085-X

[2] BOBKOV, S. G. (2003). On concentration of distributions of random weighted sums. *Ann. Probab.* **31** 195–215. MR1959791 https://doi.org/10.1214/aop/1046294309

[3] BOBKOV, S. G. (2003). Concentration of distributions of the weighted sums with Bernoullian coefficients. In *Geometric Aspects of Functional Analysis*. *Lecture Notes in Math*. **1807** 27–36. Springer, Berlin. MR2083385 https://doi.org/10.1007/978-3-540-36428-3_3

[4] BOBKOV, S. G. (2004). Concentration of normalized sums and a central limit theorem for noncorrelated random variables. *Ann*. *Probab*. **32** 2884–2907. MR2094433 https://doi.org/10.1214/009117904000000720

[5] BOBKOV, S. G. (2009). On a theorem of V. N. Sudakov on typical distributions. *Zap*. *Nauchn*. *Sem*. *S*.-*Peterburg*. *Otdel*. *Mat*. *Inst*. *Steklov*. (*POMI*) **368** 59–74, 283. MR2749183 https://doi.org/10.1007/s10958-010-9932-y

[6] BOBKOV, S. G., CHISTYAKOV, G. P. and GÖTZE, F. Concentration and Gaussian approximation for randomized sums. In preparation.

[7] BOBKOV, S. G., CHISTYAKOV, G. P. and GÖTZE, F. (2017). Second-order concentration on the sphere. *Commun*. *Contemp*. *Math*. **19** 1650058, 20. MR3670792 https://doi.org/10.1142/S0219199716500589

[8] BOBKOV, S. G., CHISTYAKOV, G. P. and GÖTZE, F. (2017). Gaussian mixtures and normal approximation for V. N. Sudakov's typical distributions. *Zap*. *Nauchn*. *Sem*. *S*.-*Peterburg*. *Otdel*. *Mat*. *Inst*. *Steklov*. (*POMI*) **457** 37–52. MR3723575

[9] BOBKOV, S. G., CHISTYAKOV, G. P. and GÖTZE, F. (2018). Berry–Esseen bounds for typical weighted sums. *Electron*. *J*. *Probab*. **23** Paper No. 92, 22. MR3858920 https://doi.org/10.1214/18-EJP195

[10] BOBKOV, S. G. and GÖTZE, F. (2007). Concentration inequalities and limit theorems for randomized sums. *Probab*. *Theory Related Fields* **137** 49–81. MR2278452 https://doi.org/10.1007/s00440-006-0500-9

[11] BOBKOV, S. G. and KOLDOBSKY, A. (2003). On the central limit property of convex bodies. In *Geometric Aspects of Functional Analysis*. *Lecture Notes in Math*. **1807** 44–52. Springer, Berlin. MR2083387 https://doi.org/10.1007/978-3-540-36428-3_5

[12] BORELL, C. (1974). Convex measures on locally convex spaces. *Ark*. *Mat*. **12** 239–252. MR0388475 https://doi.org/10.1007/BF02384761

[13] BRAZITIKOS, S., GIANNOPOULOS, A., VALETTAS, P. and VRITSIOU, B.-H. (2014). *Geometry of Isotropic Convex Bodies*. *Mathematical Surveys and Monographs* **196**. Amer. Math. Soc., Providence, RI. MR3185453 https://doi.org/10.1090/surv/196

[14] DIACONIS, P. and FREEDMAN, D. (1984). Asymptotics of graphical projection pursuit. *Ann*. *Statist*. **12** 793–815. MR0751274 https://doi.org/10.1214/aos/1176346703

[15] ELDAN, R. (2013). Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geom*. *Funct*. *Anal*. **23** 532–569. MR3053755 https://doi.org/10.1007/s00039-013-0214-y

[16] ELDAN, R. and KLARTAG, B. (2008). Pointwise estimates for marginals of convex bodies. *J*. *Funct*. *Anal*. **254** 2275–2293. MR2402109 https://doi.org/10.1016/j.jfa.2007.08.014

[17] GUÉDON, O. and MILMAN, E. (2011). Interpolating thin-shell and sharp large-deviation estimates for isotropic log-concave measures. *Geom*. *Funct*. *Anal*. **21** 1043–1068. MR2846382 https://doi.org/10.1007/s00039-011-0136-5

[18] KLARTAG, B. (2007). A central limit theorem for convex sets. *Invent*. *Math*. **168** 91–131. MR2285748 https://doi.org/10.1007/s00222-006-0028-8

[19] KLARTAG, B. (2007). Power-law estimates for the central limit theorem for convex sets. *J*. *Funct*. *Anal*. **245** 284–310. MR2311626 https://doi.org/10.1016/j.jfa.2006.12.005

[20] KLARTAG, B. (2009). A Berry–Esseen type inequality for convex bodies with an unconditional basis. *Probab*. *Theory Related Fields* **145** 1–33. MR2520120 https://doi.org/10.1007/s00440-008-0158-6

[21] KLARTAG, B. and SODIN, S. (2011). Variations on the Berry–Esseen theorem. *Teor*. *Veroyatn*. *Primen*. **56** 514–533. MR3136463 https://doi.org/10.4213/tvp4405

[22] LEDOUX, M. (1999). Concentration of measure and logarithmic Sobolev inequalities. In *Séminaire de Probabilités*, *XXXIII*. *Lecture Notes in Math*. **1709** 120–216. Springer, Berlin. MR1767995 https://doi.org/10.1007/BFb0096511

[23] LEE, Y. T. and VEMPALA, S. S. (2017). Eldan's stochastic localization and the KLS hyperplane conjecture: An improved lower bound for expansion. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 998–1007. IEEE Computer Soc., Los Alamitos, CA. MR3734299

[24] MECKES, E. S. and MECKES, M. W. (2007). The central limit problem for random vectors with symmetries. *J*. *Theoret*. *Probab*. **20** 697–720. MR2359052 https://doi.org/10.1007/s10959-007-0119-5

[25] MECKES, M. W. (2009). Gaussian marginals of convex bodies with symmetries. *Beitr*. *Algebra Geom*. **50** 101–118. MR2499783

[26] MILMAN, V. D. and SCHECHTMAN, G. (1986). *Asymptotic Theory of Finite-Dimensional Normed Spaces*. *Lecture Notes in Math*. **1200**. Springer, Berlin. MR0856576

[27] NAGAEV, S. V. (1982). On the distribution of linear functionals in finite-dimensional spaces of large dimension. *Dokl*. *Akad*. *Nauk SSSR* **263** 295–297. MR0650156

[28] PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, New York. Translated from the Russian by A. A. Brown. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82. MR0388499

[29] PETROV, V. V. (1987). *Limit Theorems for Sums of Independent Random Variables* (*Russian*). *Probability Theory and Mathematical Statistics*. Nauka, Moscow. MR0896036

[30] SUDAKOV, V. N. (1978). Typical distributions of linear functionals in finite-dimensional spaces of high dimension. *Dokl*. *Akad*. *Nauk SSSR* **243** 1402–1405. MR0517198

[31] VON WEIZSÄCKER, H. (1997). Sudakov's typical marginals, random linear functionals and a conditional central limit theorem. *Probab. Theory Related Fields* **107** 313–324. MR1440135 https://doi.org/10.1007/s004400050087