GAUSSIAN APPROXIMATIONS FOR RANDOM VECTORS

P.-L. MÉLIOT AND A. NIKEGHBALI

ABSTRACT. We present several refinements on the fluctuations of sequences of random vectors (with values in the Euclidean space $\mathbb{R}^d$) which converge after normalization to a multidimensional Gaussian distribution. More precisely we refine such results in two directions: first we give conditions under which one can obtain bounds on the speed of convergence to the multidimensional Gaussian distribution, and then we provide a setting in which one can obtain precise moderate or large deviations (in particular we see at which scale the Gaussian approximation for the tails ceases to hold and how the symmetry of the Gaussian tails is then broken). These results extend some of our earlier works obtained for real valued random variables, but they are not simple extensions, as some new phenomena are observed that could not be visible in one dimension. Even for very simple objects such as the symmetric random walk in $\mathbb{Z}^d$, we observe a loss of symmetry that we can quantify for walks conditioned to be far away from the origin. Also, unlike the one dimensional case where the Kolmogorov distance is natural, in the multidimensional case there is no more such a canonical distance. We choose to work with the so-called convex distance, and as a consequence the geometry of the Borel measurable sets that we consider shall play an important role (also making the proofs more complicated). We illustrate our results with some examples such as correlated random walks, the characteristic polynomials of random unitary matrices, or pattern countings in random graphs.

CONTENTS

1. Introduction 2
1.1. Notational conventions 3
1.2. Mod-Gaussian convergence 4
1.3. The one-dimensional case 7
1.4. Main results and outline of the article 8
Acknowledgement 9
2. Estimates on the speed of convergence 9
2.1. Convex distance between probability measures 10
2.2. Smoothing techniques 13
2.3. Distance between Fourier transforms 16
2.4. Berry–Esseen type estimates 18
3. Large deviation results 25
3.1. Sum of a Gaussian noise and an independent random variable 25
3.2. Asymptotics of the probabilities of cones 28
3.3. Approximation of spherical sectors by unions of cones 30
4. Examples of multi-dimensional convergent sequences 36
4.1. First examples 36
4.2. The method of cumulants 39
4.3. Sparse dependency graphs 48
4.4. Empirical measures of Markov chains 59
References 61

Date: March 3, 2022.

1. Introduction

Let $d \geq 1$ be a positive integer, and $(\mathbf{X}_n)_{n\in\mathbb{N}}$ be an arbitrary sequence of random vectors with values in the finite-dimensional vector space $\mathbb{R}^d$. For many probabilistic models, there exists a renormalisation $\mathbf{Y}_n = \alpha_n \mathbf{X}_n$ that admits a limit in law: $\mathbf{Y}_n \rightharpoonup_{n\to\infty} \mathbf{Y}$. For instance, if $\mathbf{X}_n = \sum_{i=1}^n \mathbf{A}_i$ is a sum of independent and identically distributed random vectors with a second moment, then the multi-dimensional central limit theorem ensures the convergence in law of

$$\mathbf{Y}_n = \frac{1}{\sqrt{n}}\,(\mathbf{X}_n - n\,\mathbb{E}[\mathbf{A}_1])$$

to a Gaussian distribution with covariance matrix $K = (\mathrm{cov}(\mathbf{A}_1^{(i)}, \mathbf{A}_1^{(j)}))_{1\leq i,j\leq d}$; see [Fel71, Section VIII.4, Theorem 2] or [Vaa00, Example 2.18]. Assuming that K is non-degenerate, this means that for any Borel set $A \subset \mathbb{R}^d$ whose topological boundary has zero Lebesgue measure,

$$\lim_{n\to\infty} \mathbb{P}[\mathbf{Y}_n \in A] = \frac{1}{\sqrt{(2\pi)^d \det K}} \int_{\mathbb{R}^d} 1_{\mathbf{x}\in A}\, \mathrm{e}^{-\frac{\mathbf{x}^t K^{-1} \mathbf{x}}{2}}\, d\mathbf{x}. \tag{1}$$

There are various ways to go beyond this convergence in law. A first direction consists in describing what happens at the edge of the distribution of $\mathbf{Y}_n$, that is with Borel subsets $A = A_n$ that vary with n and that grow in size. For instance, one can try to estimate the probability of A being in a spherical sector

$$\mathbb{P}[\mathbf{Y}_n \in [R_n, +\infty)\, B]\,,$$

where B is a Borel subset of the Euclidean sphere $\mathbb{S}^{d-1}$, and $R_n \to +\infty$. If $(R_n)_{n\in\mathbb{N}}$ does not grow too fast, one can guess that the normal approximation will still be valid, whereas if R_n is too large, it will need to be corrected. These results belong to the theory of *large deviations*, and for sums of i.i.d. random variables, the large deviations are described by Cramer's theorem [Cra38]; see [DZ98, Section 2.2.2] for an exposition of the multi-dimensional version of this result.

Another way to make precise Equation (1) is by computing bounds on the difference between $\mathbb{P}[\mathbf{Y}_n \in A]$ and its limit. In the one-dimensional case, if $A = [a, b]$ is an interval and $(A_i)_{i\geq 1}$ is as before a sequence of i.i.d. random variables, then the *Berry–Esseen estimates* [Ber41; Ess45] yield

$$\left|\mathbb{P}[Y_n \in [a,b]] - \frac{1}{\sqrt{2\pi\,\mathrm{var}(A_1)}} \int_a^b \mathrm{e}^{-\frac{x^2}{2\,\mathrm{var}(A_1)}}\, dx\right| \leq \frac{C}{\sqrt{n}},$$

assuming that A_1 has a moment of order 3; see [Fel71, Section XVI.5]. The multi-dimensional analogue of this result is much less straightforward: one way to state it is by replacing intervals by convex sets, and this approach is followed in [BR10].

In two previous works [FMN16; FMN19], we proposed a general framework that enabled us to prove large deviation results and estimates on the speed of convergence for sequences $(X_n)_{n\in\mathbb{N}}$ of one-dimensional random variables: the framework of *mod-ϕ convergence*. The interest of this theory is that it allowed us to deal with examples that were much more general than sums of i.i.d. random variables. In this setting we obtained asymptotic results for statistics of random graphs, functionals of Markov chains, arithmetic functions of random integers, characteristic polynomials of random matrices, *etc.*; see the aforementioned works and [BMN19; Chh+20; FMN20; MN22]. In this article, we shall extend this theory to a multi-dimensional setting. This extension is not trivial, as new phenomenons that cannot occur in dimension 1 take place when $d \geq 2$. In the next paragraphs of this introduction, we fix our notation and we present the hypotheses of multi-dimensional mod-Gaussian convergence. We then give an outline of the main results that we shall establish under these hypotheses (Subsection 1.4).

1.1. **Notational conventions.** We start by fixing some notation that will be used throughout the article. Since we shall work with sequences of random vectors, we need to pay special care to the exponents and indices of the quantities that we manipulate. We shall use the following conventions:

- The numbers in $\mathbb{R}$ are written with regular letters $x, y, X, Y, \ldots$; vectors in $\mathbb{R}^d$ are written with bold letters $\mathbf{x}, \mathbf{y}, \mathbf{X}, \mathbf{Y}, \ldots$
- Random numbers $X, Y, \ldots$ and random vectors $\mathbf{X}, \mathbf{Y}, \ldots$ are indicated by capital letters. We shall also use capital letters for matrices.
- The coordinates of a vector $\mathbf{x}$ are written as exponents in parentheses:
$$\mathbf{x} = (x^{(1)}, x^{(2)}, \ldots, x^{(d)}).$$
- The sequences of (random) numbers or vectors are labelled by an index $n \in \mathbb{N}$; thus, for instance, $(\mathbf{V}_n)_{n\in\mathbb{N}}$ is a sequence of random vectors in $\mathbb{R}^d$. The previous convention allows one to avoid any ambiguity between the label of a coordinate (exponent) and the label of the sequence (index).
- Without loss of generality, we assume that our (random) vectors take their values in $\mathbb{R}^d$ with $d \geq 2$. Though all our results hold also when $d = 1$, one can give shorter, and sometimes more precise proofs in this particular case, see our previous works [FMN16; FMN19].

The vector space $\mathbb{R}^d$ is endowed with the three equivalent norms

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{d} |x^{(i)}|;$$
$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{d} |x^{(i)}|^2};$$
$$\|\mathbf{x}\|_\infty = \max_{i\in[\![1,d]\!]} |x^{(i)}|,$$

and we denote:

- $B^d_{(\mathbf{x},\varepsilon)} = \{\mathbf{y} \in \mathbb{R}^d \mid \|\mathbf{x}-\mathbf{y}\|_2 \leq \varepsilon\}$ the Euclidean ball of radius ε and center $\mathbf{x}$;
- and $C^d_{(\mathbf{x},\varepsilon)} = \{\mathbf{y} \in \mathbb{R}^d \mid \|\mathbf{x}-\mathbf{y}\|_\infty \leq \varepsilon\} = \prod_{i=1}^d [x^{(i)} - \varepsilon, x^{(i)} + \varepsilon]$ the hypercube of edge 2ε and centered at $\mathbf{x}$.

As we shall mostly use the Euclidean norm, we simply note $\|\mathbf{x}\|_2 = \|\mathbf{x}\|$. The corresponding Euclidean distance is denoted by $\|\mathbf{x}-\mathbf{y}\| = d(\mathbf{x},\mathbf{y})$. Finally, $\mathbb{S}^{d-1}$ is the unit sphere $\{\mathbf{x} \in \mathbb{R}^d \mid \|x\| = 1\}$.

If $\mathbf{X}$ is a random vector in $\mathbb{R}^d$, we denote its complex Laplace transform

$$\varphi_{\mathbf{X}}(\mathbf{z}) = \mathbb{E}[\mathrm{e}^{\langle \mathbf{z} \,|\, \mathbf{X}\rangle}] = \mathbb{E}\left[\mathrm{e}^{\sum_{i=1}^d z^{(i)} X^{(i)}}\right] \qquad \text{for } \mathbf{z} = (z^{(1)}, \ldots, z^{(d)}) \in \mathbb{C}^d.$$

This quantity might not be well defined for certain values of $\mathbf{z}$, for instance if the coordinates of $\mathbf{z}$ have too large real parts; on the other hand the Fourier transform

$$\phi_{\mathbf{X}}(\boldsymbol{\zeta}) = \varphi_{\mathbf{X}}(\mathrm{i}\boldsymbol{\zeta}) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\sum_{i=1}^d \zeta^{(i)} X^{(i)}}\right]$$

is well defined for any vector $\boldsymbol{\zeta}$ in $\mathbb{R}^d$. In this article, we shall compare the distribution of random vectors $\mathbf{V}_n$ with a reference Gaussian distribution $\mathcal{N}_{\mathbb{R}^d}(\mathbf{m}, K)$; let us recall briefly what needs to be known about them. We denote $\mathrm{S}_+(d,\mathbb{R})$ the set of positive-definite symmetric matrices of size $d \times d$, and we fix $\mathbf{m} \in \mathbb{R}^d$ and $K \in \mathrm{S}_+(d,\mathbb{R})$. The Gaussian distribution with mean $\mathbf{m}$ and covariance matrix K on $\mathbb{R}^d$ is the probability distribution $\mathcal{N}_{\mathbb{R}^d}(\mathbf{m}, K)$ whose density with respect to Lebesgue measure is

$$\frac{1}{\sqrt{(2\pi)^d\,\det K}}\,\mathrm{e}^{-\frac{(\mathbf{x}-\mathbf{m})^t K^{-1}(\mathbf{x}-\mathbf{m})}{2}}\,d\mathbf{x}.$$

The Laplace transform of a random vector $\mathbf{X}$ with Gaussian law $\mathcal{N}_{\mathbb{R}^d}(\mathbf{m}, K)$ is given over the complex domain $\mathbb{C}^d$ by the formula

$$\varphi_{\mathbf{X}}(\mathbf{z}) = \mathrm{e}^{\langle \mathbf{m}\,|\,\mathbf{z}\rangle + \frac{\mathbf{z}^t K \mathbf{z}}{2}}.$$

In particular, the Fourier transform of $\mathbf{X}$ is $\phi_{\mathbf{X}}(\boldsymbol{\zeta}) = \mathrm{e}^{\mathrm{i}\langle \mathbf{m}\,|\,\boldsymbol{\zeta}\rangle - \frac{\boldsymbol{\zeta}^t K \boldsymbol{\zeta}}{2}}$ for any vector $\boldsymbol{\zeta}$ in $\mathbb{R}^d$. In our computations, the positive eigenvalues $k^{(1)} \leq k^{(2)} \leq \cdots \leq k^{(d)}$ of $K \in \mathrm{S}_+(d,\mathbb{R})$ will play an important role. For any vector $\mathbf{x} \in \mathbb{R}^d$,

$$k^{(1)}\|\mathbf{x}\|^2 \leq \mathbf{x}^t K \mathbf{x} \leq k^{(d)}\|\mathbf{x}\|^2.$$

The largest eigenvalue $k^{(d)}$ is the spectral norm $\rho(K^{1/2})$ of $K^{1/2}$, and the smallest eigenvalue $k^{(1)}$ is related to the spectral norm of $K^{-1/2}$ by the relation $k^{(1)} = (\rho(K^{-1/2}))^{-1}$.

1.2. **Mod-Gaussian convergence.** We now present our main hypothesis on the sequences of random vectors $(\mathbf{X}_n)_{n\in\mathbb{N}}$.

Definition 1.1 (Mod-Gaussian convergence)**.** *Let $(\mathbf{X}_n)_{n\in\mathbb{N}}$ be a sequence of random vectors in $\mathbb{R}^d$, and $K \in \mathrm{S}_+(d,\mathbb{R})$. We say that the sequence is mod-Gaussian convergent in the Laplace sense on $\mathbb{C}^d$ with parameters t_nK, $t_n \to +\infty$, and limiting function ψ if, locally uniformly on compact sets of $\mathbb{C}^d$,*

$$\mathbb{E}[\mathrm{e}^{\langle \mathbf{z}\,|\,\mathbf{X}_n\rangle}]\ \exp\left(-t_n\,\frac{\mathbf{z}^t K \mathbf{z}}{2}\right) = \psi_n(\mathbf{z}) \to \psi(\mathbf{z}).$$

Here, $\psi(\mathbf{z})$ is a continuous function of $\mathbf{z} \in \mathbb{C}^d$, with $\psi(\mathbf{0}) = 1$. We say that the sequence is mod-Gaussian convergent in the Fourier sense on $\mathbb{R}^d$ with the same parameters if the convergence takes place on $\mathbb{R}^d$, that is to say, if there exists a continuous function $\theta(\boldsymbol{\zeta})$ of $\boldsymbol{\zeta} \in \mathbb{R}^d$ such that

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}\langle \boldsymbol{\zeta}\,|\,\mathbf{X}_n\rangle}]\ \exp\left(t_n\,\frac{\boldsymbol{\zeta}^t K \boldsymbol{\zeta}}{2}\right) = \theta_n(\boldsymbol{\zeta}) \to \theta(\boldsymbol{\zeta})$$

locally uniformly on compact sets of $\mathbb{R}^d$.

Remark 1.2. For the mod-Gaussian convergence in the Laplace sense, it might happen that the convergence only holds on a domain $D \subset \mathbb{C}^d$ of the form $D = \mathcal{S}_{\mathbf{a},\mathbf{b}}$, where $\mathcal{S}_{\mathbf{a},\mathbf{b}}$ is the multi-strip $\mathcal{S}_{(a^{(1)},b^{(1)})} \times \cdots \times \mathcal{S}_{(a^{(d)},b^{(d)})}$, with

$$\mathcal{S}_{(a,b)} = \{z \in \mathbb{C}\,|\, a < \mathrm{Re}(z) < b\}, \quad a,b \in \mathbb{R} \sqcup \{\pm\infty\}.$$

If so, we shall still speak of mod-Gaussian convergence, but making precise this domain of convergence.

In the one-dimensional case, the notion of mod-Gaussian convergence was introduced in [JKN11]; the multi-dimensional definition first appeared in [KN12]. Let us give two important examples:

Example 1.3 (Sums of i.i.d. random vectors). Consider again a sequence of i.i.d. random vectors $(\mathbf{A}_i)_{i\geq 1}$, which we assume to be centered, with non-degenerate covariance matrix and with a third moment:

$$\mathbb{E}[\mathbf{A}_1] = 0 \quad ; \quad \left(\mathrm{cov}(\mathbf{A}_1^{(i)}, \mathbf{A}_1^{(j)})\right)_{1\leq i,j\leq d} \in \mathrm{S}_+(d,\mathbb{R}) \quad ; \quad \mathbb{E}[\|\mathbf{A}_1\|^3] < +\infty.$$

A Taylor expansion of the Fourier transform of $\mathbf{A}_1$ shows that, if

$$\mathbf{X}_n = \frac{1}{n^{1/3}} \sum_{i=1}^{n} \mathbf{A}_i,$$

then $(\mathbf{X}_n)_{n\in\mathbb{N}}$ is mod-Gaussian convergent in the Fourier sense, with parameters $n^{1/3}\,K$ with $K = \mathrm{cov}(\mathbf{A}_1)$, and limit

$$\theta(\boldsymbol{\zeta}) = \exp\left(-\frac{\mathrm{i}}{6} \sum_{i,j,k=1}^{d} \mathbb{E}[\mathbf{A}_1^{(i)}\mathbf{A}_1^{(j)}\mathbf{A}_1^{(k)}]\,\zeta^{(i)}\zeta^{(j)}\zeta^{(k)}\right).$$

If $\mathbf{A}_1$ has a convergent Laplace transform, then one has in fact a mod-Gaussian convergence in the Laplace sense on $\mathbb{C}^d$.

Example 1.4 (Characteristic polynomials of random unitary matrices). Let U_n be a random matrix in the unitary group $\mathrm{U}(n)$, taken according to the Haar measure of this Lie group. We set

$$\mathbf{X}_n = \log\det(I_n - U_n) = \sum_{i=1}^{n} \log(1 - \mathrm{e}^{\mathrm{i}\theta_i}),$$

where $\mathrm{e}^{\mathrm{i}\theta_1}, \ldots, \mathrm{e}^{\mathrm{i}\theta_n}$ are the eigenvalues of U_n, all belonging to the unit circle. Notice that on the open unit disk $\{z \mid |z| < 1\}$, the complex logarithm is given by its Taylor series:

$$\log(1-z) = -\sum_{k=1}^{\infty} \frac{z^k}{k}.$$

We extend the definition to the unit circle by setting $\log(1 - \mathrm{e}^{\mathrm{i}\theta}) = \lim_{r\to 1} \log(1 - r\mathrm{e}^{\mathrm{i}\theta})$; the limit exists and is finite if $\mathrm{e}^{\mathrm{i}\theta} \neq 1$. Since the eigenvalues are all different from 1 with probability 1 on $\mathrm{U}(n)$, $\mathbf{X}_n$ is correctly defined. The random variables $\mathbf{X}_n$ take their values in $\mathbb{C}$, which we identify with $\mathbb{R}^2$: $\mathbf{X}_n = \mathbf{X}_n^{(1)} + \mathrm{i}\mathbf{X}_n^{(2)}$. The Laplace transform of $\mathbf{X}_n$ is given by the Selberg integral

$$\mathbb{E}\left[\mathrm{e}^{z^{(1)}\mathbf{X}_n^{(1)} + z^{(2)}\mathbf{X}_n^{(2)}}\right] = \prod_{j=1}^{n} \frac{\Gamma(j)\,\Gamma(j+z^{(1)})}{\Gamma\left(j + \frac{z^{(1)}+\mathrm{i}z^{(2)}}{2}\right)\,\Gamma\left(j + \frac{z^{(1)}-\mathrm{i}z^{(2)}}{2}\right)}$$

for $\mathrm{Re}(z^{(1)}) > -1$ and $|\mathrm{Im}(z^{(2)})| < 1$; see [KS00, Formula (71)]. Since the Γ function has no zero on the complex plane, the right-hand side of this formula is a biholomorphic function on the multi-strip $\mathcal{S}_{(-1,+\infty)} \times \mathbb{C}$, and therefore, the formula above for the Laplace transform of $\mathbf{X}_n$ actually holds without restriction on the imaginary part of $z^{(2)}$. Let us introduce Barnes' G-function

$$G(z) = (2\pi)^{\frac{z}{2}} \exp\left(-z + \frac{z^2(1+\gamma)}{2}\right) \prod_{k=1}^{\infty} \left(1 + \frac{z}{k}\right)^k \exp\left(\frac{z^2}{2k} - z\right),$$

γ being the Euler constant. The function $G(z)$ is entire, and it is the solution of the functional equation

$$G(1) = 1 \quad ; \quad G(z+1) = G(z)\,\Gamma(z).$$

Then, one can rewrite the Laplace transform of $\mathbf{X}_n$ as

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{e}^{z^{(1)}\mathbf{X}_n^{(1)}+z^{(2)}\mathbf{X}_n^{(2)}}\right] \\
&= \frac{G\left(1+\frac{z^{(1)}+\mathrm{i}z^{(2)}}{2}\right)\,G\left(1+\frac{z^{(1)}-\mathrm{i}z^{(2)}}{2}\right)}{G(1+z^{(1)})}\,\frac{G(n+1)\,G(z^{(1)}+n+1)}{G\left(\frac{z^{(1)}+\mathrm{i}z^{(2)}}{2}+n+1\right)\,G\left(\frac{z^{(1)}-\mathrm{i}z^{(2)}}{2}+n+1\right)} \\
&= \frac{G\left(1+\frac{z^{(1)}+\mathrm{i}z^{(2)}}{2}\right)\,G\left(1+\frac{z^{(1)}-\mathrm{i}z^{(2)}}{2}\right)}{G(1+z^{(1)})}\,n^{\frac{(z^{(1)})^2+(z^{(2)})^2}{4}}\,(1+o(1)),
\end{aligned}$$

see the details in [KN12, Section 3] for the asymptotics of the ratio of Barnes' functions. Therefore, the sequence of random vectors $(\mathbf{X}_n)_{n\in\mathbb{N}}$ is mod-Gaussian convergent in the Laplace sense, on the domain $\mathcal{S}_{(-1,+\infty)}\times\mathbb{C}$, with parameters $t_n I_2 = \frac{\log n}{2}\,I_2$ and limiting function

$$\psi(\mathbf{z}) = \frac{G\left(1+\frac{z^{(1)}+\mathrm{i}z^{(2)}}{2}\right)\,G\left(1+\frac{z^{(1)}-\mathrm{i}z^{(2)}}{2}\right)}{G(1+z^{(1)})}.$$

The real part of $\mathbf{X}_n$ has been studied extensively in [MN22]; its mod-Gaussian convergence leads to asymptotic formulæ for $\mathbb{P}[\mathrm{Re}(\mathbf{X}_n)\geq x_n]$ for x_n in a large range of values, up to $O(n)$. The techniques that we shall develop hereafter only deal with the fluctuations of the random vectors $\mathbf{X}_n$ in plane domains of size $O(\log n)$, but we shall already see in this setting phenomena that are specific to the multi-dimensional mod-Gaussian convergence.

Many other examples of mod-Gaussian convergent sequences will be provided in Section 4. In the two previous examples, an adequate renormalisation $\mathbf{Y}_n$ of $\mathbf{X}_n$ admits a limit in law which is a Gaussian distribution. This is a consequence of the following general simple statement:

Proposition 1.5 (Central limit theorem)**.** *Let $(\mathbf{X}_n)_{n\in\mathbb{N}}$ be a sequence of random vectors that is mod-Gaussian convergent in the Fourier sense with parameters t_nK. The rescaled random variables*

$$\mathbf{Y}_n = \frac{\mathbf{X}_n}{\sqrt{t_n}}$$

converge in law towards $\mathcal{N}_{\mathbb{R}^d}(\mathbf{0},K)$.

Proof. The Fourier transform of $\mathbf{Y}_n$ is

$$\phi_{\mathbf{Y}_n}(\zeta) = \phi_{\mathbf{X}_n}\left(\frac{\zeta}{\sqrt{t_n}}\right) = \mathrm{e}^{-\frac{\zeta^t K\zeta}{2}}\,\theta_n\left(\frac{\zeta}{\sqrt{t_n}}\right),$$

and the local uniform convergence of the residue θ_n ensures that $\theta_n(\zeta/\sqrt{t_n})$ converges towards $\theta(\mathbf{0})=1$ for any $\zeta\in\mathbb{R}^d$. The convergence of the characteristic functions imply the convergence in law by Lévy's continuity theorem. □

The purpose of our article is to improve on Proposition 1.5 in the two aforementioned directions: estimates of the speed of convergence (Section 2) and large deviation results (Section 3). Let us mention that a third direction, which is opposite to the one of large deviations and concern estimates of probabilities $\mathbb{P}[\mathbf{Y}_n\in A_n]$ with Borel subsets A_n of size going to zero, can be pursued in the framework of multi-dimensional mod-Gaussian convergence. One then obtains *local limit theorems*: see [KN12; DKN15; BMN19].

Remark 1.6. Let $(\mathbf{X}_n)_{n\in\mathbb{N}}$ be a sequence of random vectors that converges mod-Gaussian, for instance in the Laplace sense on $\mathbb{C}^n$, with parameters t_nK and limiting function $\psi(\mathbf{z})$. Then, if $\mathbf{Y}_n = K^{-1/2}\mathbf{X}_n$, one has

$$\begin{aligned}\mathbb{E}\left[\mathrm{e}^{\langle \mathbf{z}\,|\,\mathbf{Y}_n\rangle}\right] &= \mathbb{E}\left[\mathrm{e}^{\left\langle K^{-1/2}\mathbf{z}\,\middle|\,\mathbf{X}_n\right\rangle}\right] \\ &= \mathrm{e}^{t_n\frac{(K^{-1/2}\mathbf{z})^t\,K\,(K^{-1/2}\mathbf{z})}{2}}\,\psi_n(K^{-1/2}\mathbf{z}) \\ &= \mathrm{e}^{t_n\frac{\|\mathbf{z}\|^2}{2}}\,\psi(K^{-1/2}\mathbf{z})\,(1+o(1)).\end{aligned}$$

So, $(\mathbf{Y}_n)_{n\in\mathbb{N}}$ converges in the mod-Gaussian sense with parameters t_nI_d and limiting function $\psi(K^{-1/2}\mathbf{z})$. Thus, every multi-dimensional mod-Gaussian convergence in the sense of Definition 1.1 can be assumed to have as parameters t_n times the identity matrix, $(t_n)_{n\in\mathbb{N}}$ being a sequence increasing to $+\infty$. However, this reduction is not necessarily the most interesting thing to do. Indeed, most of the time, the matrix t_nK is given by the first order asymptotics of the covariance matrix of $\mathbf{X}_n$; and it will be more convenient to deal with the joint moments of the coordinates of $\mathbf{X}_n$ than with those of $\mathbf{Y}_n$.

Remark 1.7. In [FMN16; FMN19] which focused on the one-dimensional case, we worked in a framework which was much more general than mod-*Gaussian* convergence. Hence, for any *infinitely divisible distribution* ϕ on $\mathbb{R}$, one can define mod-ϕ convergence by replacing in our Definition 1.1 the Lévy exponent $-\zeta^2/2$ of the Gaussian law by an arbitrary Lévy exponent of an infinitely divisible distribution; see [Sat99, Chapter 2] for an exposition of the Lévy–Khintchine formula. The same thing can be done in dimension $d \geq 2$, since the infinitely divisible distributions have exactly the same classification in dimension 1 and in higher dimensions [Sat99, Theorem 8.1]. However, if one wants in this setting large deviation estimates, then the techniques that we shall use require the infinitely divisible reference law to have a convergent Laplace transform on the complex plane; and this happens only for Gaussian distributions. Similarly, if one wants to have estimates on the speed of convergence, then we shall need sufficiently many moments for the reference law, and again we shall be restricted to the Gaussian case. This explains why we only focus on mod-Gaussian convergence when $d \geq 2$. Also, we do not know of many interesting examples of multi-dimensional mod-ϕ convergent sequences with $\phi \neq \mathcal{N}_{\mathbb{R}^d}(\mathbf{0}, K)$. One interesting example would be related to the convergence in distribution of the marginal law of a process of Ornstein–Uhlenbeck type towards a self-decomposable distribution, see [Sat99, Chapter 3, Section 17].

1.3. **The one-dimensional case.** Let us briefly recall the results of [FMN16; FMN19] that we aim to extend to higher dimensions. We start with the large deviation results, see [FMN16, Theorem 4.2.1 and Proposition 4.4.1]:

Theorem 1.8 (Large deviations in dimension $d = 1$)**.** *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of real-valued random variables that is mod-Gaussian convergent in the Laplace sense, with domain $\mathcal{S}_{(a,b)}$ with $a < 0 < b$, parameters $(t_n)_{n\in\mathbb{N}}$ and limit $\psi(z)$. We set $Y_n = \frac{X_n}{\sqrt{t_n}}$; $Y_n \rightharpoonup \mathcal{N}_{\mathbb{R}}(0,1)$. For any $x > 0$,*

$$\mathbb{P}[X_n \geq t_nx] = \mathbb{P}\left[Y_n \geq \sqrt{t_n}x\right] = \frac{\mathrm{e}^{-\frac{t_nx^2}{2}}}{x\sqrt{2\pi t_n}}\,\psi(x)\,(1+o(1)).$$

The estimate remains true as soon as $y = \sqrt{t_n}x$ is a $O(\sqrt{t_n})$ and grows to infinity. If $y = o(\sqrt{t_n})$, then the correction $\psi(x)$ to the Gaussian tail is asymptotic to 1 and does not appear.

Thus, up to the scale $y = o(\sqrt{t_n})$, the probability $\mathbb{P}[Y_n \geq y]$ is asymptotically the same as the Gaussian tail; one says that $o(\sqrt{t_n})$ is the *normality zone* of the sequence $(Y_n)_{n\in\mathbb{N}}$.

When $y = O(\sqrt{t_n})$, this is not true anymore and the limiting residue ψ of mod-Gaussian convergence appears as a correction of the Gaussian tail.

To get estimates of Berry–Esseen type, one only needs a mod-Gaussian convergence in the Fourier sense, but sometimes with precisions on the speed and zone of convergence of

$$\theta_n(\zeta) = \mathbb{E}[\mathrm{e}^{\mathrm{i}\zeta X_n}]\,\mathrm{e}^{\frac{t_n\zeta^2}{2}} \tag{2}$$

towards its limit $\theta(\zeta)$. Let us start with a result without additional hypotheses [FMN16, Proposition 4.1.1]:

Theorem 1.9 (General Berry–Esseen estimates in dimension $d = 1$)**.** *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of real-valued random variables that is mod-Gaussian convergent in the Fourier sense, with parameters $(t_n)_{n\in\mathbb{N}}$ and limit $\theta(\zeta)$. We set $Y_n = \frac{X_n}{\sqrt{t_n}}$. Then,*

$$d_{\mathrm{Kol}}(Y_n, \mathcal{N}_{\mathbb{R}}(0,1)) = \sup_{y\in\mathbb{R}} \left|\mathbb{P}[Y_n \leq y] - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} \mathrm{e}^{-\frac{x^2}{2}}\,dx\right| = O\left(\frac{1}{\sqrt{t_n}}\right).$$

This result is usually not optimal, and one can get better estimates by controlling the size of $\theta_n(\zeta)$ on a zone that grows with n. The precise hypotheses are the following [FMN19, Theorem 15]:

Theorem 1.10 (Berry–Esseen estimates with a zone of control)**.** *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of real-valued random variables. We suppose that there exists a zone $[-K(t_n)^\gamma, K(t_n)^\gamma]$, such that the residue given by Equation (2) satisfies for any ζ in this zone:*

$$|\theta_n(\zeta) - 1| \leq K_1|\zeta|^v \exp(K_2|\zeta|^w),$$

with $w \geq 2$, $v > 0$ and $-\frac{1}{2} < \gamma \leq \min(\frac{v-1}{2}, \frac{1}{w-2})$. Then, there is a constant $C(\gamma, K, K_1, K_2)$ such that

$$d_{\mathrm{Kol}}(Y_n, \mathcal{N}_{\mathbb{R}}(0,1)) \leq C(\gamma, K, K_1, K_2)\,\frac{1}{(t_n)^{\frac{1}{2}+\gamma}}.$$

In many situations including sums of i.i.d. random variables, one has a zone of control with $v = w = 3$ and $\gamma = 1$, leading to a Berry–Esseen estimate of size $O(\frac{1}{(t_n)^{3/2}})$.

Our goal will be to give analogues of Theorems 1.8, 1.9 and 1.10 in higher dimensions. Our hypotheses when $d \geq 2$ will be extremely similar to the previous ones, but the results that we shall obtain are somewhat more subtle, and their proofs will be much more complicated.

1.4. Main results and outline of the article.

In Section 2, we focus on the speed of convergence, and we introduce a distance between probability measures on $\mathbb{R}^d$ that is akin to the Kolmogorov distance, and that can be controlled under the assumption of mod-Gaussian convergence. This distance allows one to compare the distribution of $\mathbf{Y}_n$ and the Gaussian limit $\mathcal{N}_{\mathbb{R}^d}(\mathbf{0}, K)$ on any Borel measurable *convex* subset of $\mathbb{R}^d$. In general, we obtain a bound of order $O((t_n)^{-1/2})$, see Theorem 2.17. However, if the limit ψ of mod-Gaussian convergence has a certain form, and if one knows more about the convergence of the residues $\psi_n(\mathbf{z}) \to \psi(\mathbf{z})$, then one can prove bounds of order $o((t_n)^{-1/2})$, or even $O((t_n)^{-3/2})$ (Theorems 2.20 and 4.8).

In Section 3, we study the large deviations of a mod-Gaussian convergent sequence of random vectors. In the multi-dimensional setting, the open intervals $t_n[b, +\infty)$ of Theorem 1.8 are replaced by spherical or ellipsoidal sectors $t_n(S \times [b, +\infty))$, where S is a part of the K-sphere $\{\mathbf{x} \in \mathbb{R}^d \,|\, \mathbf{x}^t K^{-1} \mathbf{x} = 1\}$. If S is a sufficiently regular subset of this sphere (Jordan measurable), then one obtains in Theorem 3.6 a large deviation principle, which involves the surface integral of the residue ψ on this subset S. Therefore, the residue ψ

measures intuitively a loss of symmetry when looking at the fluctuations of a mod-Gaussian convergent sequence $(\mathbf{X}_n)_{n\in\mathbb{N}}$ at the scale t_n, and when comparing these fluctuations with the Gaussian fluctuations. The two examples presented in Section 1.2 will show clearly this phenomenon.

Finally, in Section 4, we study examples and applications of the main theorems of Sections 2-3. A large class of examples relies on the method of cumulants developed in [FMN16, Sections 5 and 9] and [FMN19, Sections 4-5], and of which we propose a multi-dimensional generalisation in Section 4.2. The method of cumulants allows one to study the fluctuations of d-dimensional random walks with dependent steps (Section 4.3), and those of the empirical measures of finite Markov chains (Section 4.4).

Acknowledgement. We are extremely grateful to Valentin Féray for countless discussions on the subject and we thank him for sharing so generously with us his very helpful insights.

2. Estimates on the speed of convergence

If μ and ν are two probability measures on $\mathbb{R}^d$, then a general way to measure a distance between μ and ν consists in fixing a class $\mathscr{F}$ of bounded measurable functions, and considering

$$d_{\mathscr{F}}(\mu,\nu) = \sup_{f\in\mathscr{F}} |\mu(f)-\nu(f)| = \sup_{f\in\mathscr{F}} \left| \int_{\mathbb{R}^d} f(x)\,\mu(dx) - \int_{\mathbb{R}^d} f(x)\,\nu(dx)\right|.$$

In particular, given a family $\mathscr{B}$ of Borel subsets of $\mathbb{R}^d$, one can consider the distance associated to the class of indicator functions $\mathscr{F} = \{1_B,\ B\in\mathscr{B}\}$:

$$d_{\mathscr{B}}(\mu,\nu) = \sup_{B\in\mathscr{B}} |\mu(B)-\nu(B)|.$$

In dimension 1, the Kolmogorov distance d_{Kol} is defined by this mean with the family $\mathscr{B} = \{(-\infty,s],\ s\in\mathbb{R}\}$. In higher dimension $d\geq 2$, the analogous distance

$$d_{\mathrm{Kol},\mathbb{R}^d}(\mu,\nu) = \sup_{(s^{(1)},\ldots,s^{(d)})\in\mathbb{R}^d} \left| \int_{-\infty}^{s^{(1)}} \cdots \int_{-\infty}^{s^{(d)}} \mu(d\mathbf{x}) - \nu(d\mathbf{x}) \right|$$

is not very convenient to deal with: for instance, it says nothing of the differences

$$\mu(B^d_{(\mathbf{x},\varepsilon)}) - \nu(B^d_{(\mathbf{x},\varepsilon)})$$

of the measures of Euclidean balls, since these probabilities are not directly accessible from the d-dimensional cumulative distribution functions. In the setting of mod-Gaussian convergence, a much better distance is given by the class of *convex Borel sets*. If ν is a probability measure that is absolutely continuous with respect to the Lebesgue measure and sufficiently isotropic, then the convex sets define a distance d_{convex} that is compatible with the convergence in law to ν, see [BR10, Theorems 2.11 and 3.1] and Section 2.1 hereafter. We shall then prove that, given a mod-Gaussian convergent sequence of random vectors $(\mathbf{X}_n)_{n\in\mathbb{N}}$, the convex distance between $\mathbf{Y}_n = \mathbf{X}_n/\sqrt{t_n}$ and its limiting distribution $\mathcal{N}_{\mathbb{R}^d}(\mathbf{0},K)$ can be bounded by a negative power of t_n (Section 2.4, Theorems 2.17 and 2.20). The proof of these results rely on various inequalities relating Fourier transforms and the convex distance to a Gaussian distribution (Sections 2.2 and 2.3).

2.1. **Convex distance between probability measures.** Recall that a subset $C \subset \mathbb{R}^d$ is called convex if, for all $\mathbf{x}, \mathbf{y} \in C$, the whole segment $[\mathbf{x}, \mathbf{y}] = \{t\mathbf{x} + (1-t)\mathbf{y}\,,\, t \in [0,1]\}$ is included into C. Let μ and ν be probability measures on $\mathbb{R}^d$.

Definition 2.1 (Convex distance)**.** *The convex distance between μ and ν by*

$$d_{\mathrm{convex}}(\mu,\nu) = \sup_{\substack{C \text{ convex Borel} \\ \text{subset of } \mathbb{R}^d}} |\mu(C) - \nu(C)|.$$

If $d = 1$, then the convex sets of $\mathbb{R}$ are the intervals. Therefore, when $d = 1$, the convex distance is a metric on probability measures that is equivalent to the Kolmogorov distance:

$$d_{\mathrm{Kol}}(\mu,\nu) \leq d_{\mathrm{convex}}(\mu,\nu) \leq 2\, d_{\mathrm{Kol}}(\mu,\nu).$$

For the reasons stated in the introduction of this section, we consider the convex distance to be the correct multi-dimensional generalisation of the Kolmogorov distance. This convex distance controls the convergence in law:

Proposition 2.2 (Convex distance and weak convergence, direct implication)**.** *Consider probability measures $(\mu_n)_{n\in\mathbb{N}}$ and ν on $\mathbb{R}^d$, such that $d_{\mathrm{convex}}(\mu_n,\nu) \to 0$. Then, μ_n converges in law to ν.*

Proof. The set of open hypercubes $\prod_{i=1}^d (a^{(i)}, b^{(i)})$ is a π-system $\mathscr{H}$ in $\mathbb{R}^d$, and every open set of $\mathbb{R}^d$ is a countable union of such hypercubes. If $d_{\mathrm{convex}}(\mu_n,\nu) \to 0$, then for every $H \in \mathscr{H}$, $\mu_n(H) \to \nu(H)$ since H is convex. Then, by [Bil99, Theorem 2.2], $\mu_n \rightharpoonup \nu$. □

Though the converse statement is false, it becomes true if ν is a probability measure that is *regular with respect to the class $\mathscr{C}$ of Borel convex sets*. Let us explain this notion of regularity. If $\varepsilon > 0$ and $C \subset \mathbb{R}^d$, we set

$$C^\varepsilon = \{\mathbf{x} \in \mathbb{R}^d \mid d(\mathbf{x}, C) \leq \varepsilon\}.$$

If C is a convex set, then C^ε is a closed convex set. We also define

$$C^{-\varepsilon} = \mathbb{R}^d \setminus (\mathbb{R}^d \setminus C)^\varepsilon = \{\mathbf{x} \in \mathbb{R}^d \mid d(\mathbf{x}, \mathbb{R}^d \setminus C) > \varepsilon\};$$

this is a part of C, and if C is a convex set, then $C^{-\varepsilon}$ is an open convex set. The *ε-boundary* of a (convex) set is $\partial^\varepsilon C = C^\varepsilon \setminus C^{-\varepsilon}$: it is a closed set, which contains for every $\varepsilon > 0$ the topological boundary ∂C of C (see Figure 1).

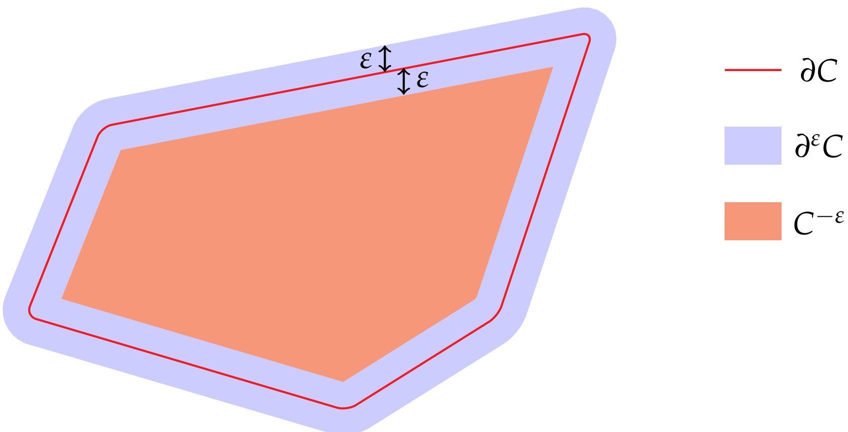

FIGURE 1. The ε-boundary $\partial^\varepsilon C$ of a convex set C.

Definition 2.3 (Regular probability measures). *A probability measure μ on $\mathbb{R}^d$ is said regular with respect to the class of convex sets if there exists a constant $R \geq 0$ such that, for every Borel convex set C,*

$$\mu(\partial^\varepsilon C) \leq R\varepsilon.$$

Example 2.4 (Regular measures in dimension $d = 1$). In dimension 1, a measure is regular with respect to the class of convex sets if and only it is absolutely continuous with respect to the Lebesgue measure, and with bounded density. Indeed, if $\mu \in \mathscr{P}(\mathbb{R})$ and $\frac{d\mu}{dx} \leq \frac{R}{4}$, then for any interval $[a,b]$,

$$\mu(\partial^\varepsilon[a,b]) \leq \mu([a-\varepsilon,a+\varepsilon]) + \mu([b-\varepsilon,b+\varepsilon]) \leq 4\,\frac{R}{4}\,\varepsilon = R\varepsilon,$$

so the measure is regular with respect to the class of convex sets. Conversely, suppose that μ is regular with respect to the class of convex sets, of constant R; and denote F_μ its cumulative distribution function. Notice then that F_μ is an absolutely continuous function: if $([a_i,b_i])_{i\in I}$ is a finite family of disjoint intervals, then for any $\varepsilon > 0$,

$$\sum_{i\in I} |F_\mu(b_i) - F_\mu(a_i)| \leq \sum_{i\in I} \mu\left(\partial^{\frac{b_i-a_i+\varepsilon}{2}}\left\{\frac{a_i+b_i}{2}\right\}\right) \leq R\sum_{i\in I}\frac{b_i-a_i+\varepsilon}{2},$$

so making ε go to 0,

$$\sum_{i\in I} |F_\mu(b_i) - F_\mu(a_i)| \leq \frac{R}{2}\sum_{i\in I}|b_i - a_i|.$$

As a consequence, F_μ is almost everywhere derivable, and

$$\frac{dF_\mu(x)}{dx} = \lim_{\varepsilon\to 0}\frac{F_\mu(x+\varepsilon)-F_\mu(x-\varepsilon)}{2\varepsilon} \leq \frac{R}{2},$$

so μ is absolutely continuous with respect to the Lebesgue measure, with density bounded by $\frac{R}{2}$.

Example 2.5 (Non-regular measures in dimension $d \geq 2$). In dimension $d \geq 2$, one can construct probability measures that have bounded density with respect to the Lebesgue measure, but that are not regular with respect to the class of convex sets. Let g be a continuous non-negative function on $\mathbb{R}_+$, such that:

- $\int_0^\infty g(r)\,r^{d-1}\,dr < +\infty$.
- g is bounded by some constant M on $\mathbb{R}_+$, and there exists a level $L > 0$ such that $\limsup_{r\to\infty} g(r) > L$.

Hence, we want both $g(r)$ to integrate r^{d-1} and to reach a fixed level $L > 0$ an infinite number of times (on very small intervals). It is easily seen that such functions do exist. Let μ be the measure on $\mathbb{R}^d$ with density $g(\|\mathbf{x}\|)\,d\mathbf{x}$; up to a renormalisation of g, one can assume μ to be a probability measure. Then, μ has density bounded by M. On the other hand, there exists a sequence $r_n \to +\infty$ such that $g(r_n) \geq L$ for every n. Since g is continuous, one can find for every r_n an interval $[r_n - \varepsilon_n, r_n + \varepsilon_n]$ such that $g(r) \geq \frac{L}{2}$ for every r in this interval. Then,

$$\mu\left(\partial^{\varepsilon_n} B^d_{(\mathbf{0},r_n)}\right) = \mathrm{vol}(\mathbb{S}^{d-1})\int_{r_n-\varepsilon_n}^{r_n+\varepsilon_n} g(r)\,r^{d-1}\,dr \geq L\,\mathrm{vol}(\mathbb{S}^{d-1})\,(r_n)^{d-1}\,\varepsilon_n.$$

Since $(r_n)^{d-1}$ can be taken as large as wanted, μ is not a regular measure with respect to convex sets.

The interest of the notion of regularity comes from:

Proposition 2.6 (Convex distance and weak convergence, reciprocal implication)**.** *Let ν be a probability on $\mathbb{R}^d$ that is regular with respect to the class of convex sets. If $(\mu_n)_{n\in\mathbb{N}}$ is a sequence of probability measures with $\mu_n \rightharpoonup \nu$ (convergence in law), then $d_{\mathrm{convex}}(\mu_n,\nu) \to 0$.*

Proof. See [BR10, Theorem 2.11]. Later, we shall see that if ν is regular, and if sufficiently many derivatives of $\widehat{\mu}_n(\zeta)$ converge to those of $\widehat{\nu}(\zeta)$ locally uniformly, then we have indeed $d_{\mathrm{convex}}(\mu_n,\nu) \to 0$. We shall even be able to quantify this; see the remark at the end of Section 2.3. □

In the sequel, we shall need to know that the Gaussian distributions $\mathcal{N}_{\mathbb{R}^d}(\mathbf{m},K)$ are always regular with respect to the class of convex sets. This is a consequence of the following:

Lemma 2.7. *Consider a probability measure $\nu(d\mathbf{x}) = g(\|\mathbf{x}\|)\,d\mathbf{x}$, with $\int_0^\infty r^{d-1}\,|g'(r)|\,dr = c_1 < +\infty$. Then, for any convex set $C \subset \mathbb{R}^d$, and any $\varepsilon > 0$*

$$\nu(\partial^\varepsilon C) \leq 2\,c_1\,\mathrm{vol}(\mathbb{S}^{d-1})\,\varepsilon = \frac{4\pi^{d/2}}{\Gamma(d/2)}\,c_1\,\varepsilon.$$

Proof. See [BR10, Theorem 3.1]. □

Corollary 2.8 (Regularity of the Gaussian distributions)**.** *If ν is a Gaussian distribution on $\mathbb{R}^d$ with covariance matrix K, then ν is regular with constant $R = 2\sqrt{(d+1)\,\rho(K^{-1})}$: for any convex Borel set $C \subset \mathbb{R}^d$ and any $\varepsilon > 0$,*

$$\nu(\partial^\varepsilon C) \leq 2\sqrt{(d+1)\,\rho(K^{-1})}\,\varepsilon.$$

Proof. If ν is the standard Gaussian distribution on $\mathbb{R}^d$ ($K = I_d$), then

$$c_1 = \frac{1}{(2\pi)^{d/2}}\int_0^\infty r^d\,\mathrm{e}^{-\frac{r^2}{2}}\,dr = \frac{2^{(d-1)/2}}{(2\pi)^{d/2}}\int_0^\infty u^{\frac{d+1}{2}-1}\,\mathrm{e}^{-u}\,du = \frac{1}{\sqrt{2}\,\pi^{d/2}}\,\Gamma\left(\frac{d+1}{2}\right).$$

Therefore, ν is regular with respect to convex sets, of constant

$$R \leq 2\sqrt{2}\,\frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} \leq 2\sqrt{d+1}.$$

More generally, suppose that $\nu = \mathcal{N}_{\mathbb{R}^d}(\mathbf{0},K)$ is an arbitrary non-degenerated Gaussian distribution on $\mathbb{R}^d$. If $\|\mathbf{x}-\mathbf{y}\| \leq \varepsilon$, then

$$\|K^{-1/2}(\mathbf{x}-\mathbf{y})\| \leq (k^{(1)})^{-1/2}\|\mathbf{x}-\mathbf{y}\| \leq (k^{(1)})^{-1/2}\,\varepsilon,$$

where $k^{(1)} \leq k^{(2)} \leq \cdots \leq k^{(d)}$ are the positive eigenvalues of K. It follows that $K^{-1/2}(C^\varepsilon) \subset (K^{-1/2}C)^{\varepsilon(k^{(1)})^{-1/2}}$, and similarly, $K^{-1/2}(C^{-\varepsilon}) \supset (K^{-1/2}C)^{-\varepsilon(k^{(1)})^{-1/2}}$. As a consequence, if $\nu = \mathcal{N}_{\mathbb{R}^d}(\mathbf{0},K)$ and $\widetilde{\nu} = \mathcal{N}_{\mathbb{R}^d}(\mathbf{0},I_d)$, then

$$\begin{aligned}
\nu(\partial^\varepsilon C) &= \frac{1}{\sqrt{(2\pi)^d\,\det K}}\left(\int_{C^\varepsilon} - \int_{C^{-\varepsilon}}\right)\left(\mathrm{e}^{-\frac{\mathbf{x}^t K^{-1}\mathbf{x}}{2}}\,d\mathbf{x}\right) \\
&\leq \frac{1}{\sqrt{(2\pi)^d}}\left(\int_{(K^{-1/2}C)^{\varepsilon(k^{(1)})^{-1/2}}} - \int_{(K^{-1/2}C)^{-\varepsilon(k^{(1)})^{-1/2}}}\right)\left(\mathrm{e}^{-\frac{\|\mathbf{y}\|^2}{2}}\,d\mathbf{y}\right) \\
&\leq \widetilde{\nu}\left(\partial^{\varepsilon(k^{(1)})^{-1/2}}(K^{-1/2}C)\right) \leq 2\sqrt{d+1}\,(k^{(1)})^{-1/2}\,\varepsilon.
\end{aligned}$$

Indeed, if C is a convex set, then $K^{-1/2}C$ is also a convex set. □

2.2. **Smoothing techniques.** The regularity of the Gaussian distributions enables us to use smoothing techniques in order to compute $d_{\mathrm{convex}}(\mu, \mathcal{N}_{\mathbb{R}}(\mathbf{0}, K))$. We first introduce an adequate smoothing kernel, which is the multi-dimensional analogue of the kernel of [FMN19, Lemma 17]:

Lemma 2.9. *There exists a kernel ρ on $\mathbb{R}^d$ such that:*

(1) *The kernel is non-negative, $\int_{\mathbb{R}^d} \rho(\mathbf{x})\, d\mathbf{x} = 1$, and $\int_{B^d_{(\mathbf{0},(2d+2)^{3/2})}} \rho(\mathbf{x})\, d\mathbf{x} \geq c_2$ with*

$$2c_2 - 1 \geq 1 - \frac{4}{9\pi}.$$

(2) *The kernel $\rho(\mathbf{x})$ integrates $(\|\mathbf{x}\|_1)^{d+1}$:*

$$\int_{\mathbb{R}^d} \rho(\mathbf{x})\, (\|\mathbf{x}\|_1)^{d+1}\, d\mathbf{x} < +\infty.$$

(3) *The Fourier transform $\widehat{\rho}(\boldsymbol{\zeta}) = \int_{\mathbb{R}^d} \mathrm{e}^{\mathrm{i}\langle \zeta \,|\, \mathbf{x}\rangle} \rho(d\mathbf{x})$ of ρ is compactly supported on $[-1,1]^d$.*

(4) *The Fourier transform of ρ is of class $\mathscr{C}^{2d}$, and for any multi-index $\boldsymbol{\beta}$ of weight $|\boldsymbol{\beta}| = \sum_{i=1}^d \beta^{(i)} \leq 2d$,*

$$\left\| \frac{\partial^{|\boldsymbol{\beta}|} \widehat{\rho}(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}^{\boldsymbol{\beta}}} \right\|_\infty \leq 2^{1+\frac{d}{2}}\, \pi^{-\frac{d}{2}}\, (2d+2)^{|\boldsymbol{\beta}|+\frac{d}{2}}.$$

Proof. We set $\mathrm{sinc}(x) = \frac{\sin x}{x}$, $\mathrm{sinc}_{2d+2}(x) = (\mathrm{sinc}(\frac{x}{2d+2}))^{2d+2}$, and finally

$$\rho(\mathbf{x}) = \frac{\prod_{i=1}^d \mathrm{sinc}_{2d+2}(x^{(i)})}{\int_{\mathbb{R}^d} \prod_{i=1}^d \mathrm{sinc}_{2d+2}(x^{(i)})\, d\mathbf{x}}.$$

(1) Since $|\tan x| \geq |x|$ on $[-\frac{\pi}{2}, \frac{\pi}{2}]$, one has

$$\int_{\mathbb{R}} (\mathrm{sinc}(x))^{2d+2}\, dx \geq \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\cos x)^{2d+2}\, dx = \sqrt{\pi}\, \frac{\Gamma(d+\frac{3}{2})}{\Gamma(d+2)}.$$

Then, since the Euclidean ball $B^d_{(\mathbf{0},\varepsilon)}$ contains the hypercube $C^d_{(\mathbf{0},\frac{\varepsilon}{\sqrt{d}})}$,

$$\begin{aligned}
\frac{\int_{\mathbb{R}^d \setminus B^d_{(\mathbf{0},\varepsilon)}} \prod_{i=1}^d \mathrm{sinc}_{2d+2}(x^{(i)})\, d\mathbf{x}}{\int_{\mathbb{R}^d} \prod_{i=1}^d \mathrm{sinc}_{2d+2}(x^{(i)})\, d\mathbf{x}} &\leq \frac{\int_{\mathbb{R}^d \setminus C^d_{(\mathbf{0},\frac{\varepsilon}{\sqrt{d}})}} \prod_{i=1}^d \mathrm{sinc}_{2d+2}(x^{(i)})\, d\mathbf{x}}{\int_{\mathbb{R}^d} \prod_{i=1}^d \mathrm{sinc}_{2d+2}(x^{(i)})\, d\mathbf{x}} \\
&\leq d\, \frac{\int_{\mathbb{R} \setminus [-\frac{\varepsilon}{\sqrt{d}}, \frac{\varepsilon}{\sqrt{d}}]} \mathrm{sinc}_{2d+2}(x)\, dx}{\int_{\mathbb{R}} \mathrm{sinc}_{2d+2}(x)\, dx} \\
&\leq \frac{2d\, \Gamma(d+2)}{\sqrt{\pi}\, \Gamma(d+\frac{3}{2})} \int_{\frac{\varepsilon}{\sqrt{d}\,(2d+2)}}^{+\infty} \frac{dy}{y^{2d+2}} \\
&\leq \frac{2d\, \Gamma(d+2)}{\sqrt{\pi}\, (2d+1)\, \Gamma(d+\frac{3}{2})} \left(\frac{\sqrt{d}\,(2d+2)}{\varepsilon} \right)^{2d+1} \\
&\leq \frac{2d\, \Gamma(d+2)}{\sqrt{\pi}\, (2d+1)\, \Gamma(d+\frac{3}{2})} \left(\frac{d}{2d+2} \right)^{d+\frac{1}{2}}
\end{aligned}$$

if $\varepsilon = (2d+2)^{3/2}$. This last expression takes its maximal value $\frac{2}{9\pi}$ for $d = 1$. Therefore,

$$\int_{B^d_{(\mathbf{0},(2d+2)^{3/2})}} \rho(\mathbf{x})\,d\mathbf{x} \geq 1 - \frac{2}{9\pi} \qquad ; \qquad 2c_2 - 1 \geq 1 - \frac{4}{9\pi}.$$

(2) Note that

$$\begin{aligned}\prod_{i=1}^d \mathrm{sinc}_{2d+2}(x^{(i)}) &\leq \mathrm{sinc}_{2d+2}\left(\max_{1\leq i\leq d} |x^{(i)}|\right) \\ &\leq \frac{(2d+2)^{2d+2}}{\max_{1\leq i\leq d} |x^{(i)}|^{2d+2}} \leq \frac{(d(2d+2))^{2d+2}}{(\|\mathbf{x}\|_1)^{2d+2}}.\end{aligned}$$

Therefore, the kernel indeed integrates $(\|\mathbf{x}\|_1)^{d+1}$.

(3) One has

$$\begin{aligned}\widehat{\rho}(\boldsymbol{\zeta}) &= Z \prod_{i=1}^d \left(\int_{\mathbb{R}} \mathrm{sinc}_{2d+2}(x)\,\mathrm{e}^{\mathrm{i}x\zeta^{(i)}}\,dx\right) \\ &= Z' \prod_{i=1}^d \left(\int_{\mathbb{R}} (\mathrm{sinc}(x))^{2d+2}\,\mathrm{e}^{\mathrm{i}x\frac{\zeta^{(i)}}{2d+2}}\,dx\right) \\ &= Z' \prod_{i=1}^d \left(\widehat{\mathrm{sinc}}\right)^{*(2d+2)}\left(\frac{\zeta^{(i)}}{2d+2}\right),\end{aligned}$$

where Z and Z' are normalisation constants, and $f^{*(2d+2)}$ is the $(2d+2)$-th convolution power of an integrable function f. However, the Fourier transform of the sine cardinal is $\widehat{\mathrm{sinc}}(\zeta) = \pi\,1_{|\zeta|\leq 1}$, which is supported on $[-1,1]$. Therefore, $(\widehat{\mathrm{sinc}})^{*(2d+2)}$ is supported on $[-(2d+2),(2d+2)]$, and $\widehat{\rho}$ is indeed supported on $[-1,1]^d$.

(4) If $f(x) = (\mathrm{sinc}(x))^{2d+2}$, then $\frac{\partial^k \widehat{f}}{\partial\zeta^k}(\zeta) = \mathrm{i}^k\,\widehat{(x^k f)}(\zeta)$, so

$$\begin{aligned}\left\|\frac{\partial^k \widehat{f}}{\partial\zeta^k}(\zeta)\right\|_\infty &\leq \int_{\mathbb{R}} \frac{\sin^{2d+2}(x)}{|x|^{2d+2-k}}\,dx \leq \int_{-1}^1 |x|^k\,dx + 2\int_1^\infty \frac{1}{x^{2d+2-k}}\,dx \\ &\leq 2\left(\frac{1}{k+1} + \frac{1}{2d+1-k}\right) \leq 2\,\frac{2d+2}{2d+1}\end{aligned}$$

for any $k \in [\![0,2d]\!]$. Since $\widehat{\rho}(\boldsymbol{\zeta}) = \prod_{i=1}^d \frac{\widehat{f}((2d+2)\zeta^{(i)})}{\widehat{f}(0)}$, it follows that

$$\begin{aligned}\left\|\frac{\partial^{|\beta|}\widehat{\rho}(\boldsymbol{\zeta})}{\partial\boldsymbol{\zeta}^{\boldsymbol{\beta}}}\right\|_\infty &\leq (2d+2)^{|\beta|}\left(\frac{2(2d+2)\,\Gamma(d+2)}{\sqrt{\pi}\,(2d+1)\,\Gamma(d+\frac{3}{2})}\right)^d \\ &\leq 2^{1+\frac{d}{2}}\,\pi^{-\frac{d}{2}}\,(2d+2)^{|\beta|+\frac{d}{2}}\end{aligned}$$

by using Stirling estimates on the last line. □

For $\varepsilon > 0$, we note

$$\rho_\varepsilon(\mathbf{x}) = \left(\frac{(2d+2)^{3/2}}{\varepsilon}\right)^d \rho\left(\frac{(2d+2)^{3/2}\,\mathbf{x}}{\varepsilon}\right).$$

For every $d \geq 1$, the two first points in Lemma 2.9 translate into the following properties for the smoothing kernel ρ_ε:

(SK1) $\rho_\varepsilon(\mathbf{x}) \geq 0$ on $\mathbb{R}^d$, and $\int_{\mathbb{R}^d} \rho_\varepsilon(\mathbf{x})\,d\mathbf{x} = 1$.

(SK2) ρ_ε gives a mass close to 1 to the Euclidean ball $B^d_{(\mathbf{0},\varepsilon)}$:

$$\int_{B^d_{(\mathbf{0},\varepsilon)}} \rho_\varepsilon(\mathbf{x})\,d\mathbf{x} = c_2 > \frac{1}{2}.$$

(SK3) the Fourier transform of ρ_ε is compactly supported on $\left[-\frac{(2d+2)^{3/2}}{\varepsilon}, \frac{(2d+2)^{3/2}}{\varepsilon}\right]^d$.

The discussion hereafter holds for any kernel with the two properties (SK1) and (SK2). Let f be a bounded measurable function on $\mathbb{R}^d$. It is well known that, if the *oscillations* of f are not too large with respect to a probability measure ν, then one can use smoothing techniques in order to control $\int_{\mathbb{R}^d} f(\mathbf{x})\,(\mu(d\mathbf{x}) - \nu(d\mathbf{x}))$. More precisely, consider the following quantities, all related to the oscillations of the function f:

$$\begin{aligned} M(f,\mathbf{x},\varepsilon) &= \sup\{f(\mathbf{y}) \,|\, d(\mathbf{x},\mathbf{y}) \leq \varepsilon\}; \\ m(f,\mathbf{x},\varepsilon) &= \inf\{f(\mathbf{y}) \,|\, d(\mathbf{x},\mathbf{y}) \leq \varepsilon\}; \\ \omega(f,\mathbf{x},\varepsilon) &= M(f,\mathbf{x},\varepsilon) - m(f,\mathbf{x},\varepsilon). \end{aligned}$$

We then denote $f_{\mathbf{t}}(\mathbf{x}) = f(\mathbf{x}+\mathbf{t})$ the translate of f by a vector $\mathbf{t} \in \mathbb{R}^d$, and

$$\begin{aligned} \omega(f) &= \sup\{|f(\mathbf{x}) - f(\mathbf{y})|,\, \mathbf{x},\mathbf{y} \in \mathbb{R}^d\}; \\ \omega(f,\varepsilon) &= \int_{\mathbb{R}^d} \omega(f,\mathbf{x},\varepsilon)\,\nu(d\mathbf{x}); \\ \omega^*(f,\varepsilon) &= \sup\{\omega(f_t,\varepsilon) \,|\, t \in \mathbb{R}^d\}. \end{aligned}$$

Lemma 2.10. *For any kernel ρ_ε that satisfies the two properties (SK1) and (SK2), one has*

$$\left|\int_{\mathbb{R}^d} f(\mathbf{x})\,(\mu(d\mathbf{x}) - \nu(d\mathbf{x}))\right| \leq \frac{1}{2c_2-1}\left(\frac{\omega(f)}{2}\,\|(\mu-\nu)*\rho_\varepsilon\| + \omega^*(f,2\varepsilon)\right),$$

where $\|(\mu-\nu)*\rho_\varepsilon\| = ((\mu-\nu)*\rho_\varepsilon)_+(\mathbb{R}^d) + ((\mu-\nu)*\rho_\varepsilon)_-(\mathbb{R}^d)$ is the norm of the signed measure $(\mu-\nu)*\rho_\varepsilon$.

Proof. Cf. [BR10, Corollary 11.5]. □

Suppose that $f = 1_C$ is the indicator function of a Borel convex set $C \subset \mathbb{R}^d$, and that ν is a probability measure that is regular with constant R. Then, $\omega(f) = 1$, and $\omega(f,\mathbf{x},\varepsilon) = 1_{\mathbf{x}\in\partial^\varepsilon C}$. As a consequence,

$$\omega^*(f,2\varepsilon) = \int_{\mathbb{R}^d} 1_{\mathbf{x}\in\partial^{2\varepsilon}C}\,\nu(d\mathbf{x}) \leq 2R\,\varepsilon.$$

Combining the previous lemma with the estimate of c_2 given in Lemma 2.9, we conclude that:

Proposition 2.11 (Convex distance to a regular measure)**.** *Let μ be an arbitrary probability measure on $\mathbb{R}^d$, and ν a probability measure that is regular with respect to the class of convex sets, with constant R. One has the following inequality:*

$$d_{\text{convex}}(\mu,\nu) \leq \frac{1}{1-\frac{4}{9\pi}}\left(\frac{\|(\mu-\nu)*\rho_\varepsilon\|}{2} + 2R\,\varepsilon\right).$$

In particular, if ν a Gaussian distribution with covariance matrix $K > 0$, then

$$d_{\text{convex}}(\mu,\nu) \leq \frac{1}{1-\frac{4}{9\pi}}\left(\frac{\|(\mu-\nu)*\rho_\varepsilon\|}{2} + 4\sqrt{(d+1)\,\rho(K^{-1})}\,\varepsilon\right).$$

2.3. **Distance between Fourier transforms.** Given two probability measures μ and ν on $\mathbb{R}^d$, we now explain how to relate the total variation norm $\|(\mu-\nu)*\rho_\varepsilon\|$ to certain properties of the Fourier transforms $\widehat{\mu}(\boldsymbol{\zeta})$ and $\widehat{\nu}(\boldsymbol{\zeta})$. In the remainder of this paragraph, we assume that μ and ν have moments of all order (or at least of all order $\leq d+1$); therefore, $\widehat{\mu}$ and $\widehat{\nu}$ have derivatives of all order. We then denote

$$\Delta_\varepsilon(\widehat{\mu},\widehat{\nu}) = \max_{|\boldsymbol{\beta}|\in[\![0,d+1]\!]} \int_{\left[-\frac{(2d+2)^{3/2}}{\varepsilon},\frac{(2d+2)^{3/2}}{\varepsilon}\right]^d} \left|\frac{\partial^{|\boldsymbol{\beta}|}(\widehat{\mu}-\widehat{\nu})}{\partial\boldsymbol{\zeta}^{\boldsymbol{\beta}}}(\boldsymbol{\zeta})\right| d\boldsymbol{\zeta},$$

the maximum being taken over multi-indices $\boldsymbol{\beta}=(\beta^{(1)},\ldots,\beta^{(d)})$ such that $|\boldsymbol{\beta}|\in[\![0,d+1]\!]$. We also abbreviate

$$C^d_{(\mathbf{0},(2d+2)^{3/2}/\varepsilon)} = \left[-\frac{(2d+2)^{3/2}}{\varepsilon},\frac{(2d+2)^{3/2}}{\varepsilon}\right]^d = D^d_\varepsilon.$$

Proposition 2.12 (Variation norm and distance between Fourier transforms)**.** *For any probability measures μ and ν with moments of all order, if $\varepsilon\leq\frac{1}{\sqrt{2d+2}}$, then*

$$\|(\mu-\nu)*\rho_\varepsilon\| \leq 4\,(d+1)^{\frac{d+1}{2}}\,\Delta_\varepsilon(\widehat{\mu},\widehat{\nu}).$$

Lemma 2.13. *Let $g:\mathbb{R}^d\to\mathbb{R}$ be a function such that $\int_{\mathbb{R}^d}|g(\mathbf{x})|\,(\|\mathbf{x}\|_1)^{d+1}\,d\mathbf{x}<+\infty$. Then,*

$$\int_{\mathbb{R}^d}|g(\mathbf{x})|\,d\mathbf{x} \leq \sqrt{\frac{\pi d}{2}}\left(\frac{\mathrm{e}}{\pi}\right)^d \max_{|\boldsymbol{\beta}|\in\{0,d+1\}} \int_{\mathbb{R}^d}\left|\frac{\partial^{|\boldsymbol{\beta}|}\widehat{g}}{\partial\boldsymbol{\zeta}^{\boldsymbol{\beta}}}(\boldsymbol{\zeta})\right| d\boldsymbol{\zeta}.$$

Proof. We follow the proof of [BR10, Lemma 11.6], but we make the constants more explicit. Let $G_+=\{\mathbf{x}\in\mathbb{R}^d\,|\,g(\mathbf{x})\geq 0\}$, and $G_-=\mathbb{R}^d\setminus G_+$. For any sequences of signs $\boldsymbol{\alpha}\in\{\pm1\}^d$, we denote $Q_{\boldsymbol{\alpha}}=\{\mathbf{x}\in\mathbb{R}^d\,|\,\forall i\in[\![1,d]\!]\,,\ \mathrm{sgn}(x^{(i)})=\alpha^{(i)}\}$ the corresponding quadrant of $\mathbb{R}^d$, with by convention $\mathrm{sgn}(0)=+1$. We then write:

$$\begin{aligned}
&\int_{\mathbb{R}^d}|g(\mathbf{x})|\,d\mathbf{x}\\
&\quad= \sum_{\boldsymbol{\alpha}\in\{\pm1\}^d}\left(\int_{Q_{\boldsymbol{\alpha}}\cap G_+}-\int_{Q_{\boldsymbol{\alpha}}\cap G_-}\right) g(\mathbf{x})\,(d^{d+1}+(\|\mathbf{x}\|_1)^{d+1})\,\frac{1}{d^{d+1}+(\|\mathbf{x}\|_1)^{d+1}}\,d\mathbf{x}\\
&\quad= \sum_{\boldsymbol{\alpha}\in\{\pm1\}^d}\left(\int_{Q_{\boldsymbol{\alpha}}\cap G_+}-\int_{Q_{\boldsymbol{\alpha}}\cap G_-}\right)\left(\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\widehat{h_{\boldsymbol{\alpha}}}(\boldsymbol{\zeta})\,\mathrm{e}^{-\mathrm{i}\langle\mathbf{x}\,|\,\boldsymbol{\zeta}\rangle}\,d\boldsymbol{\zeta}\right)\frac{1}{d^{d+1}+(\|\mathbf{x}\|_1)^{d+1}}\,d\mathbf{x},
\end{aligned}$$

where

$$h_{\boldsymbol{\alpha}}(\mathbf{x}) = \left(d^{d+1}+\left(\sum_{i=1}^d \alpha^{(i)}\,x^{(i)}\right)^{d+1}\right) g(\mathbf{x}).$$

The Fourier transform of $h_{\boldsymbol{\alpha}}(\mathbf{x})$ is:

$$\widehat{h_{\boldsymbol{\alpha}}}(\boldsymbol{\zeta}) = \left(d^{d+1}+(-\mathrm{i})^{d+1}\sum_{|\boldsymbol{\beta}|=d+1}\binom{d+1}{\beta^{(1)},\ldots,\beta^{(d)}}\boldsymbol{\alpha}^{\boldsymbol{\beta}}\,\frac{\partial^{|\boldsymbol{\beta}|}}{\partial\boldsymbol{\zeta}^{\boldsymbol{\beta}}}\right)\widehat{g}(\boldsymbol{\zeta}).$$

Therefore,

$$\left|\int_{\mathbb{R}^d}\widehat{h_{\boldsymbol{\alpha}}}(\boldsymbol{\zeta})\,\mathrm{e}^{-\mathrm{i}\langle\mathbf{x}\,|\,\boldsymbol{\zeta}\rangle}\,d\boldsymbol{\zeta}\right| \leq 2\,d^{d+1}\max_{|\boldsymbol{\beta}|\in\{0,d+1\}}\left\|\frac{\partial^{|\boldsymbol{\beta}|}\widehat{g}(\boldsymbol{\zeta})}{\partial\boldsymbol{\zeta}^{\boldsymbol{\beta}}}\right\|_{\mathrm{L}^1}.$$

It follows that

$$\|g\|_{\mathrm{L}^1} \leq \frac{2\,d^{d+1}}{(2\pi)^d}\left(\int_{\mathbb{R}^d}\frac{1}{d^{d+1}+(\|\mathbf{x}\|_1)^{d+1}}\,d\mathbf{x}\right)\max_{|\boldsymbol{\beta}|\in\{0,d+1\}}\left\|\frac{\partial^{|\boldsymbol{\beta}|}\widehat{g}(\boldsymbol{\zeta})}{\partial\boldsymbol{\zeta}^{\boldsymbol{\beta}}}\right\|_{\mathrm{L}^1}.$$

Last we compute $\int_{\mathbb{R}^d} \frac{1}{d^{d+1}+(\|\mathbf{x}\|_1)^{d+1}}\,d\mathbf{x}$ as follows:

$$\begin{aligned}\int_{\mathbb{R}^d} \frac{1}{d^{d+1}+(\|\mathbf{x}\|_1)^{d+1}}\,d\mathbf{x} &= \frac{2^d}{d^{d+1}} \int_{(\mathbb{R}_+)^d} \frac{1}{1+\left(\frac{x^{(1)}+\cdots+x^{(d)}}{d}\right)^{d+1}}\,d\mathbf{x} \\ &= \frac{2^d}{d} \int_{(\mathbb{R}_+)^d} \frac{1}{1+(y^{(1)}+\cdots+y^{(d)})^{d+1}}\,d\mathbf{y} \\ &= \frac{2^d}{d!} \int_{t=0}^{+\infty} \frac{t^{d-1}}{1+t^{d+1}}\,dt = \frac{2^d\,\pi}{(d+1)!\,\sin(\frac{\pi}{d+1})}.\end{aligned}$$

We conclude by using Stirling's approximation for $(d+1)!$:

$$\|g\|_{\mathrm{L}^1} \leq \sqrt{\frac{\pi d}{2}}\left(\frac{\mathrm{e}}{\pi}\right)^d \max_{|\boldsymbol{\beta}|\in\{0,d+1\}} \left\| \frac{\partial^{|\boldsymbol{\beta}|}\widehat{g}(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}^{\boldsymbol{\beta}}} \right\|_{\mathrm{L}^1}. \qquad \square$$

Proof of Proposition 2.12. Since ρ_ε is a smoothing kernel, the signed measure obtained by the convolution $(\mu-\nu)*\rho_\varepsilon$ has a (smooth) density g with respect to the Lebesgue measure, and one can apply to it the previous Lemma:

$$\|(\mu-\nu)*\rho_\varepsilon\| = \|g\|_{\mathrm{L}^1} \leq C(d) \max_{|\boldsymbol{\beta}|\in\{0,d+1\}} \left\| \frac{\partial^{|\boldsymbol{\beta}|}}{\partial \boldsymbol{\zeta}^{\boldsymbol{\beta}}} \left((\widehat{\mu}-\widehat{\nu})\,\widehat{\rho_\varepsilon}\right) \right\|_{\mathrm{L}^1}.$$

Indeed,

$$\begin{aligned}&\int_{\mathbb{R}^d} |g(\mathbf{x})|\,(\|\mathbf{x}\|_1)^{d+1}\,d\mathbf{x} \\ &\leq \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} (\mu+\nu)(d\mathbf{y})\,\rho_\varepsilon(\mathbf{x}-\mathbf{y})\,(\|\mathbf{x}\|_1)^{d+1}\,d\mathbf{x} \\ &\leq 2^{d+1}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} (\mu+\nu)(d\mathbf{y})\,\rho_\varepsilon(\mathbf{x}-\mathbf{y})((\|\mathbf{x}-\mathbf{y}\|_1)^{d+1}+(\|\mathbf{y}\|_1)^{d+1})\,d\mathbf{x} \\ &\leq 2^{d+2}\left(\int_{\mathbb{R}^d}(\|\mathbf{x}\|_1)^{d+1}\,\rho_\varepsilon(\mathbf{x})\,d\mathbf{x}\right) + 2^{d+1}\left(\int_{\mathbb{R}^d}(\|\mathbf{x}\|_1)^{d+1}\,(\mu+\nu)(d\mathbf{x})\right)\end{aligned}$$

which is finite since μ and ν have moments of all order, and ρ_ε integrates $(\|\mathbf{x}\|_1)^{d+1}$. Now, if $|\boldsymbol{\beta}|$ is a fixed multi-index of total weight $|\boldsymbol{\beta}| = d+1$, then

$$\begin{aligned}&\left\| \frac{\partial^{|\boldsymbol{\beta}|}}{\partial \boldsymbol{\zeta}^{\boldsymbol{\beta}}}\left((\widehat{\mu}-\widehat{\nu})\,\widehat{\rho_\varepsilon}\right)\right\|_{\mathrm{L}^1} \\ &\leq \sum_{\alpha^{(1)}=0}^{\beta^{(1)}}\cdots\sum_{\alpha^{(d)}=0}^{\beta^{(d)}} \prod_{i=1}^{d}\binom{\beta^{(i)}}{\alpha^{(i)}} \int_{D_\varepsilon^d} \left| \frac{\partial^{|\boldsymbol{\beta}-\boldsymbol{\alpha}|}(\widehat{\mu}-\widehat{\nu})(\boldsymbol{\zeta})}{\partial\boldsymbol{\zeta}^{\boldsymbol{\beta}-\boldsymbol{\alpha}}}\,\frac{\partial^{|\boldsymbol{\alpha}|}\widehat{\rho_\varepsilon}(\boldsymbol{\zeta})}{\partial\boldsymbol{\zeta}^{\boldsymbol{\alpha}}}\right|\,d\boldsymbol{\zeta} \\ &\leq \sum_{\alpha^{(1)}=0}^{\beta^{(1)}}\cdots\sum_{\alpha^{(d)}=0}^{\beta^{(d)}} \prod_{i=1}^{d}\binom{\beta^{(i)}}{\alpha^{(i)}} \left\|\frac{\partial^{|\boldsymbol{\alpha}|}\widehat{\rho_\varepsilon}}{\partial\boldsymbol{\zeta}^{\boldsymbol{\alpha}}}\right\|_\infty \Delta_\varepsilon(\widehat{\mu},\widehat{\nu}) \\ &\leq 2\left(\sqrt{\frac{4d+4}{\pi}}\right)^d\left(1+\frac{\varepsilon}{\sqrt{2d+2}}\right)^{d+1}\Delta_\varepsilon(\widehat{\mu},\widehat{\nu}) \\ &\leq 2\mathrm{e}^{1/2}\left(\sqrt{\frac{4d+4}{\pi}}\right)^d\Delta_\varepsilon(\widehat{\mu},\widehat{\nu})\end{aligned}$$

by using the bound on the derivatives of $\widehat{\rho_\varepsilon}$ given by Lemma 2.9. Since the constant before $\Delta_\varepsilon(\widehat{\mu},\widehat{\nu})$ is always larger than 1, the same estimate holds when $|\boldsymbol{\beta}|=0$, so we get

$$\begin{aligned}\|(\mu-\nu)*\rho_\varepsilon\| &\leq 2\mathrm{e}^{1/2}\sqrt{\frac{\pi d}{2}}\left(\frac{\mathrm{e}}{\pi}\sqrt{\frac{4d+4}{\pi}}\right)^d \Delta_\varepsilon(\widehat{\mu},\widehat{\nu}) \\ &\leq \frac{\pi^2}{\sqrt{2\mathrm{e}}}\left(\frac{2\mathrm{e}}{\pi^{3/2}}\sqrt{d+1}\right)^{d+1}\Delta_\varepsilon(\widehat{\mu},\widehat{\nu}) \\ &\leq 4\left(\sqrt{d+1}\right)^{d+1}\Delta_\varepsilon(\widehat{\mu},\widehat{\nu}),\end{aligned}$$

by simplyfing a bit the constants for the last inequality. □

Corollary 2.14 (Convex distance and distance between Fourier transforms)**.** *Let μ be an arbitrary probability measure on $\mathbb{R}^d$, and ν be a probability distribution that is regular with respect to the class of convex sets, with constant R. We assume that μ and ν have moments of any order smaller than $d+1$. Then,*

$$d_{\mathrm{convex}}(\mu,\nu) \leq \frac{2}{1-\frac{4}{9\pi}}\left((d+1)^{\frac{d+1}{2}}\Delta_\varepsilon(\widehat{\mu},\widehat{\nu})+R\,\varepsilon\right)$$

for any $\varepsilon < \frac{1}{\sqrt{2d+2}}$.

In particular,

$$d_{\mathrm{convex}}(\mu,\mathcal{N}_{\mathbb{R}^d}(\mathbf{0},K)) \leq \frac{2}{1-\frac{4}{9\pi}}\left((d+1)^{\frac{d+1}{2}}\Delta_\varepsilon\left(\widehat{\mu},\mathrm{e}^{-\frac{\zeta^tK\zeta}{2}}\right)+2\sqrt{(d+1)\,\rho(K^{-1})\,\varepsilon}\right).$$

Remark 2.15. Corollary 2.14 implies the reciprocal of Proposition 2.2 if one assumes that, in addition to the convergence in law $\mu_n \rightharpoonup \nu$, which is equivalent to the local uniform convergence $\widehat{\mu}_n(\zeta) \to \widehat{\nu}(\zeta)$, one also has convergence of the partial derivatives of these Fourier transforms up to order $d+1$. Actually, given a measure ν regular with respect to convex sets, one does not need this additional assumption to have $d_{\mathrm{convex}}(\mu_n,\nu) \to 0$, but this is required if one wants quantitative estimates.

2.4. **Berry–Esseen type estimates.** We now apply the preliminary results of the previous paragraphs to a mod-Gaussian convergent sequence of random variables $(\mathbf{X}_n)_{n\in\mathbb{N}}$.

2.4.1. *Normal approximation of a mod-Gaussian convergent sequence.* We start with the following general hypotheses:

(BE1) The sequence of random vectors $(\mathbf{X}_n)_{n\in\mathbb{N}}$ is mod-Gaussian convergent in the Laplace sense, on $\mathbb{C}^d$ or on a multi-strip $\mathcal{S}_{(\mathbf{a},\mathbf{b})}$ such that $\mathbf{0} \in \prod_{i=1}^d (a^{(i)},b^{(i)})$, with parameters t_nK and limit $\psi(\mathbf{z})$. We denote as before $\theta_n(\zeta)=\psi_n(\mathrm{i}\zeta)$ and $\theta(\zeta)=\psi(\mathrm{i}\zeta)$ for $\zeta\in\mathbb{R}^d$.

(BE2) Or, the sequence $(\mathbf{X}_n)_{n\in\mathbb{N}}$ is mod-Gaussian convergent in a *strong* Fourier sense: the residue θ_n *and all its partial derivatives up to order* $d+1$ converge locally uniformly on $\mathbb{R}$ towards θ and all its partial derivatives up to order $d+1$.

The hypothesis (BE1) is stronger than (BE2). Indeed, if the holomorphic functions $\psi_n(\mathbf{z})$ converge locally uniformly towards $\psi(\mathbf{z})$, then $\psi(\mathbf{z})$ is holomorphic on the domain of convergence, and one has automatically a local uniform convergence of the complex derivatives up to any order. In particular, by restriction to the domain $D=(\mathrm{i}\mathbb{R})^d$, one gets the content of Hypothesis (BE2), which is itself just a bit stronger than Fourier mod-Gaussian convergence in the sense of Definition 1.1.

Until the end of this paragraph, $(\mathbf{X}_n)_{n\in\mathbb{N}}$ is a fixed mod-Gaussian convergent sequence that satisfies Hypothesis (BE2). We set

$$M(B) = M((\theta_n)_{n\in\mathbb{N}}, B) = \sup_{n\in\mathbb{N}} \max_{|\boldsymbol{\alpha}|\in[\![0,d+1]\!]} \max_{\zeta\in C^d_{(\mathbf{0},B(2d+2)^{3/2})}} \left|\frac{\partial^{|\boldsymbol{\alpha}|}\theta_n(\zeta)}{\partial\zeta^{\boldsymbol{\alpha}}}\right|.$$

Because $\theta_n \to \theta$ uniformly on $C^d_{(\mathbf{0},B(2d+2)^{3/2})}$, as well as their derivatives up to order $d+1$, $M(B)$ is finite. On the other hand, we denote μ_n the law of $\mathbf{X}_n/\sqrt{t_n}$, and $\nu = \mathcal{N}_{\mathbb{R}^d}(\mathbf{0}, K)$.

Lemma 2.16. *Under the hypothesis (BE2), assuming also that $d \geq 2$ and that $\det K = 1$, one has*

$$\Delta_{\frac{1}{B\sqrt{t_n}}}(\widehat{\mu}_n, \widehat{\nu}) \leq \left(\frac{2\pi}{\mathrm{e}}\right)^{\frac{d}{2}} (d+1)^{d+\frac{11}{4}} (\rho(K^{-1}))^{\frac{d}{2}+1} \frac{M(B)}{\sqrt{t_n}}$$

for any fixed $B > 0$, and for n large enough.

Before proving this result, let us discuss the assumption that K is *normalised*:

$$\det K = 1.$$

First, this assumption can be made without loss of generality, by replacing K by $K_{\mathrm{norm}} = (\det K)^{-1/d} K$, and t_n by $(\det K)^{1/d} t_n$. On the other hand, since $\det K = k^{(1)}k^{(2)}\cdots k^{(d)}$ is the product of the eigenvalues of $\det K$, if $\det K = 1$, then $\rho(K) \geq 1$ and $\rho(K^{-1}) \geq 1$. For $K \in \mathrm{S}_+(d,\mathbb{R})$, set $\tau(K) = \rho(K^{-1})\,\rho(K)$. This quantity is invariant by the transformations $K \mapsto \lambda K$, and it is always larger than 1, since

$$1 \leq \rho((K_{\mathrm{norm}})^{-1}) = (\det K)^{1/d}\,\rho(K^{-1}) \leq \rho(K)\,\rho(K^{-1}) = \tau(K).$$

Because of this sequence of inequalities, if an upper bound involving $\rho(K^{-1})$ is proven under the additional assumption $\det K = 1$, then most of the time, an equivalent upper bound will hold without the assumption $\det K = 1$, replacing $\rho(K^{-1})$ by $\tau(K)$.

Proof of Lemma 2.16. Fix a multi-index $\boldsymbol{\beta}$ of weight $\leq d+1$, and set $\varepsilon = 1/(B\sqrt{t_n})$. We introduce the Hermite polynomials

$$H_n(x) = (-1)^n\,\mathrm{e}^{\frac{x^2}{2}}\,\frac{d^n}{dx^n}\left(\mathrm{e}^{-\frac{x^2}{2}}\right), \quad n \geq 0.$$

If $\boldsymbol{\alpha}$ is a multi-index, then

$$\mathrm{e}^{\frac{\|\zeta\|^2}{2}}\,\frac{\partial^{\boldsymbol{\alpha}}}{\partial\zeta^{\boldsymbol{\alpha}}}\left(\mathrm{e}^{-\frac{\|\zeta\|^2}{2}}\right) = (-1)^{|\boldsymbol{\alpha}|}\prod_{i=1}^{d} H_{\alpha^{(i)}}(\zeta^{(i)}),$$

which we abbreviate as $(-1)^{|\boldsymbol{\alpha}|}H_{\boldsymbol{\alpha}}(\zeta)$. In the following, given two multi-indices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we write $\boldsymbol{\alpha} \leq \boldsymbol{\beta}$ if $\alpha^{(i)} \leq \beta^{(i)}$ for all $i \in [\![1,d]\!]$. We then set $\binom{\boldsymbol{\beta}}{\boldsymbol{\alpha}} = \prod_{i=1}^{d}\binom{\beta^{(i)}}{\alpha^{(i)}}$. One has

$$\begin{aligned}
&\int_{D^d_\varepsilon}\left|\frac{\partial^{|\boldsymbol{\beta}|}(\widehat{\mu}_n - \widehat{\nu})}{\partial\zeta^{\boldsymbol{\beta}}}(\zeta)\right| d\zeta \\
&\leq \int_{D^d_{1/B\sqrt{t_n}}}\left|\frac{\partial^{|\boldsymbol{\beta}|}\widehat{\nu}(\zeta)}{\partial\zeta^{\boldsymbol{\beta}}}\right|\,\left|\theta_n\left(\frac{\zeta}{\sqrt{t_n}}\right) - 1\right| d\zeta \\
&\quad + \sum_{0<\boldsymbol{\alpha}\leq\boldsymbol{\beta}}\left(\frac{1}{\sqrt{t_n}}\right)^{|\boldsymbol{\alpha}|}\binom{\boldsymbol{\beta}}{\boldsymbol{\alpha}}\int_{D^d_{1/B\sqrt{t_n}}}\left|\frac{\partial^{|\boldsymbol{\beta}-\boldsymbol{\alpha}|}\widehat{\nu}}{\partial\zeta^{\boldsymbol{\beta}-\boldsymbol{\alpha}}}(\zeta)\right|\,\left|\frac{\partial^{|\boldsymbol{\alpha}|}\theta_n}{\partial\zeta^{\boldsymbol{\alpha}}}\left(\frac{\zeta}{\sqrt{t_n}}\right)\right| d\zeta \\
&\leq \frac{M(B)}{\sqrt{t_n}}\int_{\mathbb{R}^d}\left|\frac{\partial^{|\boldsymbol{\beta}|}\widehat{\nu}(\zeta)}{\partial\zeta^{\boldsymbol{\beta}}}\right|\,\|\zeta\|_1\,d\zeta \;+\; M(B)\sum_{0<\boldsymbol{\alpha}\leq\boldsymbol{\beta}}\left(\frac{1}{\sqrt{t_n}}\right)^{|\boldsymbol{\alpha}|}\binom{\boldsymbol{\beta}}{\boldsymbol{\alpha}}\int_{\mathbb{R}^d}\left|\frac{\partial^{|\boldsymbol{\beta}-\boldsymbol{\alpha}|}\widehat{\nu}}{\partial\zeta^{\boldsymbol{\beta}-\boldsymbol{\alpha}}}(\zeta)\right| d\zeta.
\end{aligned}$$

Recall that $\widehat{\nu}(\boldsymbol{\zeta}) = \mathrm{e}^{-\frac{\zeta^t K \zeta}{2}}$. To evaluate the integrals above, we make the change of variables $\boldsymbol{\xi} = K^{1/2}\boldsymbol{\zeta}$. If M is an invertible matrix in $\mathrm{GL}(\mathbb{R}^d)$, and $\boldsymbol{\zeta} = M\,\boldsymbol{\xi}$, then

$$\frac{\partial f}{\partial \zeta^{(i)}} = \sum_{j=1}^d M_{ij}\,\frac{\partial f}{\partial \xi^{(j)}}.$$

Therefore, with $M = K^{-1/2}$,

$$\begin{aligned}\left|\frac{\partial^{|\boldsymbol{\beta}|}\widehat{\nu}(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}^{\boldsymbol{\beta}}}\right| = \left|\left(\prod_{i=1}^d \frac{\partial^{\beta^{(i)}}}{\partial \zeta^{\beta^{(i)}}}\right)\mathrm{e}^{-\frac{\|\boldsymbol{\xi}\|^2}{2}}\right| &\leq \sum_{\substack{j_{11},\ldots,j_{1\beta^{(1)}}\\ \vdots \\ j_{d1},\ldots,j_{d\beta^{(d)}}}} \left|\left(\prod_{i=1}^d\prod_{k=1}^{\beta^{(i)}} M_{ij_{ik}}\frac{\partial}{\partial \xi^{(j_{ik})}}\right)\mathrm{e}^{-\frac{\|\boldsymbol{\xi}\|^2}{2}}\right| \\ &\leq \sum_{\substack{j_{11},\ldots,j_{1\beta^{(1)}}\\ \vdots \\ j_{d1},\ldots,j_{d\beta^{(d)}}}} \left(\prod_{i=1}^d\prod_{k=1}^{\beta^{(i)}} |M_{ij_{ik}}|\right)\,|H_{\boldsymbol{\alpha}(\mathbf{j})}(\boldsymbol{\xi})|\,\mathrm{e}^{-\frac{\|\boldsymbol{\xi}\|^2}{2}},\end{aligned}$$

where the multi-indices $\boldsymbol{\alpha}(\mathbf{j})$ depend on the choice of the indices $j_{11},\ldots,j_{d\beta^{(d)}} \in [\![1,d]\!]$, and are all of total weight $|\boldsymbol{\beta}|$. However, for any multi-indices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$,

$$\begin{aligned}\int_{\mathbb{R}^d} |H_{\boldsymbol{\alpha}}(\boldsymbol{\xi})||H_{\boldsymbol{\beta}}(\boldsymbol{\xi})|\,\mathrm{e}^{-\frac{\|\boldsymbol{\xi}\|^2}{2}}\,d\boldsymbol{\xi} &\leq \sqrt{\int_{\mathbb{R}^d}(H_{\boldsymbol{\alpha}}(\boldsymbol{\xi}))^2\,\mathrm{e}^{-\frac{\|\boldsymbol{\xi}\|^2}{2}}\,d\boldsymbol{\xi}}\sqrt{\int_{\mathbb{R}^d}(H_{\boldsymbol{\beta}}(\boldsymbol{\xi}))^2\,\mathrm{e}^{-\frac{\|\boldsymbol{\xi}\|^2}{2}}\,d\boldsymbol{\xi}} \\ &\leq \sqrt{(2\pi)^d\,\boldsymbol{\alpha}!\,\boldsymbol{\beta}!}\end{aligned}$$

where $\boldsymbol{\alpha}! = \prod_{i=1}^d (\alpha^{(i)})!$ if $\boldsymbol{\alpha}$ is a multi-index of integers. Notice that if $\boldsymbol{\alpha}$ is of total weight $|\boldsymbol{\alpha}|$, then $\boldsymbol{\alpha}! \leq |\boldsymbol{\alpha}|!$. So,

$$\begin{aligned}\int_{\mathbb{R}^d}\left|\frac{\partial^{|\boldsymbol{\beta}|}\widehat{\nu}(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}^{\boldsymbol{\beta}}}\right|\|\boldsymbol{\zeta}\|_1\,d\boldsymbol{\zeta} &\leq \sum_{k,l}\sum_{j_{11},\ldots,j_{d\beta^{(d)}}} |M_{kl}|\left(\prod_{i=1}^d\prod_{k=1}^{\beta^{(i)}}|M_{ij_{ik}}|\right)\int_{\mathbb{R}^d}|H_{\boldsymbol{\alpha}(\mathbf{j})}(\boldsymbol{\xi})|\,|\boldsymbol{\xi}^{(l)}|\,\mathrm{e}^{-\frac{\|\boldsymbol{\xi}\|^2}{2}}\,d\boldsymbol{\xi} \\ &\leq d\left(\max_{i\in[\![1,d]\!]}\sum_{j=1}^d |M_{ij}|\right)^{1+|\boldsymbol{\beta}|}\sqrt{(2\pi)^d\,|\boldsymbol{\beta}|!}\end{aligned}$$

Moreover,

$$\max_{i\in[\![1,d]\!]}\sum_{j=1}^d |M_{ij}| = \max_{\|\mathbf{v}\|_\infty\leq 1}\|M\mathbf{v}\|_\infty \leq \max_{\|\mathbf{v}\|_2\leq\sqrt{d}}\|M\mathbf{v}\|_2 = \sqrt{d}\,\rho(K^{-1/2}).$$

Thus,

$$\begin{aligned}\int_{\mathbb{R}^d}\left|\frac{\partial^{|\boldsymbol{\beta}|}\widehat{\nu}(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}^{\boldsymbol{\beta}}}\right|\|\boldsymbol{\zeta}\|_1\,d\boldsymbol{\zeta} &\leq d^{\frac{3+|\boldsymbol{\beta}|}{2}}\left(\rho(K^{-1})\right)^{\frac{1+|\boldsymbol{\beta}|}{2}}\sqrt{(2\pi)^d\,|\boldsymbol{\beta}|!}, \\ \int_{\mathbb{R}^d}\left|\frac{\partial^{|\boldsymbol{\beta}-\boldsymbol{\alpha}|}\widehat{\nu}}{\partial \boldsymbol{\zeta}^{\boldsymbol{\beta}-\boldsymbol{\alpha}}}(\boldsymbol{\zeta})\right|d\boldsymbol{\zeta} &\leq \left(d\,\rho(K^{-1})\right)^{\frac{|\boldsymbol{\beta}-\boldsymbol{\alpha}|}{2}}\sqrt{(2\pi)^d\,|\boldsymbol{\beta}-\boldsymbol{\alpha}|!}\,.\end{aligned}$$

We conclude that

$$\begin{aligned}
&\int_{D^d_\varepsilon} \left|\frac{\partial^{|\beta|}(\widehat{\mu}_n-\widehat{\nu})}{\partial\zeta^\beta}(\zeta)\right|\,d\zeta \\
&\leq \sqrt{(2\pi)^d\,|\beta|!}\,\left(d\,\rho(K^{-1})\right)^{\frac{|\beta|}{2}}\,M(B)\,\left(\frac{d^{\frac{3}{2}}\rho(K^{-\frac{1}{2}})}{\sqrt{t_n}}+\left(1+\frac{1}{\sqrt{t_n\,d\,\rho(K^{-1})}}\right)^{|\beta|}-1\right) \\
&\leq_{n\to\infty} \sqrt{(2\pi)^d\,|\beta|!}\,\left(d\,\rho(K^{-1})\right)^{\frac{|\beta|}{2}}\,M(B)\,\left(\frac{d^{\frac{3}{2}}\rho(K^{-\frac{1}{2}})}{\sqrt{t_n}}+\frac{|\beta|+1}{\sqrt{t_n\,d\,\rho(K^{-1})}}\right) \\
&\leq_{n\to\infty} \sqrt{(2\pi)^d\,(d+1)!}\,\left(d\,\rho(K^{-1})\right)^{\frac{d+2}{2}}\,M(B)\,\frac{d+\frac{d+2}{d}}{\sqrt{t_n}}
\end{aligned}$$

where the symbol $\leq_{n\to\infty}$ means that the inequality holds for n large enough. The inequality follows by using Stirling's estimate

$$(d+1)! \leq \sqrt{2\pi}\,\frac{(d+1)^{d+\frac{3}{2}}}{\mathrm{e}^d},$$

and by simplifying a bit the constants. □

Theorem 2.17 (General Berry–Esseen estimates). *Let $(\mathbf{X}_n)_{n\in\mathbb{N}}$ be a sequence of random variables that satisfies the hypothesis (BE2). We denote μ_n the law of $\frac{\mathbf{X}_n}{\sqrt{t_n}}$ and $\nu=\mathcal{N}_{\mathbb{R}^d}(\mathbf{0},K)$. One has*

$$d_{\mathrm{convex}}(\mu_n,\nu)=O\left(\frac{1}{\sqrt{t_n}}\right).$$

If $d\geq 2$, one can take for constant in the $O(\cdot)$

$$\frac{2\,\rho(K^{-1/2})}{1-\frac{4}{9\pi}}\left(\left(\frac{2\pi}{\mathrm{e}}\right)^{\frac{d}{2}}(d+1)^{\frac{3d}{2}+\frac{13}{4}}\,(\tau(K))^{\frac{d}{2}+1}\,M(B)+2\sqrt{(d+1)\,\tau(K)}\,\frac{1}{B}\right)$$

for any $B>0$.

Proof. Suppose first that $\det K=1$. Then, combining the previous Lemma 2.16 with Corollary 2.14, and setting $\varepsilon=\frac{1}{B\sqrt{t_n}}$, one obtains

$$\begin{aligned}
&d_{\mathrm{convex}}(\mu_n,\nu) \\
&\leq \frac{2}{1-\frac{4}{9\pi}}\left(\left(\frac{2\pi}{\mathrm{e}}\right)^{\frac{d}{2}}(d+1)^{\frac{3d}{2}+\frac{13}{4}}\,(\rho(K^{-1}))^{\frac{d}{2}+1}\,M(B)+2\sqrt{(d+1)\,\rho(K^{-1})}\,\frac{1}{B}\right)\frac{1}{\sqrt{t_n}}
\end{aligned}$$

for any $B>0$.

If $\det K\neq 1$, then

$$\begin{aligned}
&\sqrt{t_n\,(\det K)^{1/d}}\;d_{\mathrm{convex}}\left(\frac{\mathbf{X}_n}{\sqrt{t_n}},\mathcal{N}_{\mathbb{R}^d}(\mathbf{0},K)\right) \\
&=\sqrt{t_n\,(\det K)^{1/d}}\;d_{\mathrm{convex}}\left(\frac{\mathbf{X}_n}{\sqrt{t_n\,(\det K)^{1/d}}},\mathcal{N}_{\mathbb{R}^d}(\mathbf{0},K_{\mathrm{norm}})\right) \\
&\leq \frac{2}{1-\frac{4}{9\pi}}\left(\left(\frac{2\pi}{\mathrm{e}}\right)^{\frac{d}{2}}(d+1)^{\frac{3d}{2}+\frac{13}{4}}\,(\tau(K))^{\frac{d}{2}+1}\,M(B)+2\sqrt{(d+1)\,\tau(K)}\,\frac{1}{B}\right) \\
&\leq \frac{2}{1-\frac{4}{9\pi}}\left(\left(\frac{2\pi}{\mathrm{e}}\right)^{\frac{d}{2}}(d+1)^{\frac{3d}{2}+\frac{13}{4}}\,(\tau(K))^{\frac{d}{2}+1}\,M(B)+2\sqrt{(d+1)\,\tau(K)}\,\frac{1}{B}\right)
\end{aligned}$$

for any $B > 0$, by using the inequality $\rho((K_{\text{norm}})^{-1}) \leq \tau(K)$ previously mentioned. Finally, $(\det K)^{1/d} \geq \frac{1}{\rho(K^{-1})}$. ☐

2.4.2. *Modification of the normal approximation.* In [FMN16], the one-dimensional equivalent of Theorem 2.17 was an important tool in the proof of the large deviation results (Theorem 1.8). We shall use in Section 3 the same ideas, but we shall then need a slightly better Berry–Esseen bound $o((t_n)^{-1/2})$, instead of $O((t_n)^{-1/2})$. Under the general hypothesis (BE2), it is possible to write such an upper bound, but then one needs to replace $\nu = \mathcal{N}_{\mathbb{R}^d}(\mathbf{0}, K)$ by a *signed* measure ν_n, which is a small modification of the Gaussian distribution ν. This paragraph is devoted to the construction of the measures ν_n, and to the proof of the upper bound $d_{\text{convex}}(\mu_n, \nu_n) = o((t_n)^{-1/2})$. Given a function f of $\boldsymbol{\zeta} \in \mathbb{R}^d$, we denote

$$(\nabla f)(\zeta) = \left(\frac{\partial f(\zeta)}{\partial \zeta^{(1)}}, \ldots, \frac{\partial f(\zeta)}{\partial \zeta^{(d)}}\right).$$

Given a sequence $(\mathbf{X}_n)_{n\in\mathbb{N}}$ that satisfies Hypothesis (BE2), notice that

$$\frac{\partial \theta_n}{\partial \zeta^{(i)}}(\mathbf{0}) = \frac{\partial}{\partial \zeta^{(i)}} \left(\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\langle \zeta \,|\, \mathbf{X}_n\rangle}\right] \mathrm{e}^{\frac{\zeta^t K \zeta}{2}}\right)_{|\zeta = \mathbf{0}} = \mathrm{i}\, \mathbb{E}\left[\mathbf{X}_n^{(i)}\right]$$

is a purely imaginary number. Therefore, the gradient vector $\nabla\theta(\mathbf{0})$ belongs to $(\mathrm{i}\mathbb{R})^d$. We set

$$\frac{d\nu_n}{d\mathbf{x}} = \frac{1}{\sqrt{(2\pi)^d \det K}}\, \mathrm{e}^{-\frac{\mathbf{x}^t K^{-1} \mathbf{x}}{2}} \left(1 - \mathrm{i}\, \frac{\mathbf{x}^t\, K^{-1} \nabla\theta(\mathbf{0})}{\sqrt{t_n}}\right).$$

This is the density of a signed measure of total mass 1 and with Fourier transform

$$\int_{\mathbb{R}^d} \mathrm{e}^{\mathrm{i}\langle \zeta \,|\, \mathbf{x}\rangle}\, \nu_n(d\mathbf{x}) = \mathrm{e}^{-\frac{\zeta^t K \zeta}{2}} \left(1 + \frac{\zeta^t\, \nabla\theta(\mathbf{0})}{\sqrt{t_n}}\right).$$

Lemma 2.18. *The modified Gaussian distribution ν_n is regular with respect to the class of Borel convex sets, with constant*

$$R \leq_{n\to\infty} 2\sqrt{(d+1)\, \rho(K^{-1})}.$$

Proof. Suppose first that $K = I_d$. The density of ν_n is then bounded in absolute value by

$$(2\pi)^{-\frac{d}{2}}\, \mathrm{e}^{-\frac{\|\mathbf{x}\|^2}{2}} \left(1 + \frac{\|\mathbf{x}\|\, \|\nabla\theta(\mathbf{0})\|}{\sqrt{t_n}}\right),$$

which satisfies the hypotheses of Lemma 2.7:

$$\begin{aligned}\int_{r=0}^{\infty} |g'(r)|\, r^{d-1}\, dr &\leq \frac{1}{(2\pi)^{d/2}} \int_0^\infty \mathrm{e}^{-\frac{r^2}{2}} \left(r + \frac{(r^2+1)\, \|\nabla\theta(\mathbf{0})\|}{\sqrt{t_n}}\right) r^{d-1}\, dr \\ &= \frac{1}{\sqrt{2}\, \pi^{d/2}}\, \Gamma\left(\frac{d+1}{2}\right) + O\left(\frac{1}{\sqrt{t_n}}\right).\end{aligned}$$

Hence, ν_n is regular with constant

$$R = 2\sqrt{2}\, \frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} + O\left(\frac{1}{\sqrt{t_n}}\right),$$

which is smaller than $2\sqrt{d+1}$ for t_n large enough. The general case is then obtained by the same techniques of change of variables as in the proof of Lemma 2.7. ☐

Lemma 2.19. *Under the hypothesis (BE2), assuming also* $\det K = 1$*, one has*

$$\Delta_{\frac{1}{B\sqrt{t_n}}}(\widehat{\mu}_n, \widehat{\nu}_n) \leq C_1(d, \rho(K^{-1}))\, \frac{\|\nabla\theta_n(\mathbf{0}) - \nabla\theta(\mathbf{0})\|_\infty}{\sqrt{t_n}} + C_2(d, \rho(K^{-1}))\, \frac{M(B)}{t_n}$$

for any $B > 0$*, where* $C_1(d, \rho(K^{-1}))$ *and* $C_2(d, \rho(K^{-1}))$ *are some constants that depend only on* d *and on the spectral radius* $\rho(K^{-1})$*.*

Proof. As in Lemma 2.16, we fix a multi-index $\boldsymbol{\beta}$ of total weight smaller than $d+1$, and we set $\varepsilon = 1/(B\sqrt{t_n})$. We have

$$\widehat{\mu}_n(\boldsymbol{\zeta}) - \widehat{\nu}_n(\boldsymbol{\zeta}) = \mathrm{e}^{-\frac{\boldsymbol{\zeta}^t K \boldsymbol{\zeta}}{2}} \left(\theta_n\left(\frac{\boldsymbol{\zeta}}{\sqrt{t_n}}\right) - 1 - \frac{\boldsymbol{\zeta}^t\, \nabla\theta(\mathbf{0})}{\sqrt{t_n}}\right) = \widehat{\nu}(\boldsymbol{\zeta})\, \delta_n\left(\frac{\boldsymbol{\zeta}}{\sqrt{t_n}}\right).$$

In the expansion of

$$\frac{\partial^{|\boldsymbol{\beta}|}(\widehat{\mu}_n - \widehat{\nu}_n)}{\partial \boldsymbol{\zeta}^{\boldsymbol{\beta}}} = \sum_{\mathbf{0} \leq \boldsymbol{\alpha} \leq \boldsymbol{\beta}} \left(\frac{1}{\sqrt{t_n}}\right)^{|\boldsymbol{\alpha}|} \binom{\boldsymbol{\beta}}{\boldsymbol{\alpha}} \left(\frac{\partial^{|\boldsymbol{\beta}-\boldsymbol{\alpha}|}\widehat{\nu}}{\partial \boldsymbol{\zeta}^{\boldsymbol{\beta}-\boldsymbol{\alpha}}}(\boldsymbol{\zeta})\right) \left(\frac{\partial^{|\boldsymbol{\alpha}|}\delta_n}{\partial \boldsymbol{\zeta}^{\boldsymbol{\alpha}}}\left(\frac{\boldsymbol{\zeta}}{\sqrt{t_n}}\right)\right),$$

we separate the multi-indices $\boldsymbol{\alpha}$ in three categories:

(0) If $\boldsymbol{\alpha} = \mathbf{0}$, then

$$\begin{aligned}
&\int_{D_\varepsilon^d} \left|\frac{\partial^{|\boldsymbol{\beta}|}\widehat{\nu}}{\partial \boldsymbol{\zeta}^{\boldsymbol{\beta}}}(\boldsymbol{\zeta})\right| \left|\delta_n\left(\frac{\boldsymbol{\zeta}}{\sqrt{t_n}}\right)\right| d\boldsymbol{\zeta} \\
&\leq \int_{D^d_{1/B\sqrt{t_n}}} \left|\frac{\partial^{|\boldsymbol{\beta}|}\widehat{\nu}}{\partial \boldsymbol{\zeta}^{\boldsymbol{\beta}}}(\boldsymbol{\zeta})\right| \left(\left|\frac{\boldsymbol{\zeta}^t(\nabla\theta_n(\mathbf{0}) - \nabla\theta(\mathbf{0}))}{\sqrt{t_n}}\right| + \frac{M(B)\,(\|\boldsymbol{\zeta}\|_1)^2}{2t_n}\right) d\boldsymbol{\zeta} \\
&\leq \frac{\|\nabla\theta_n(\mathbf{0}) - \nabla\theta(\mathbf{0})\|_\infty}{\sqrt{t_n}} \int_{\mathbb{R}^d} \left|\frac{\partial^{|\boldsymbol{\beta}|}\widehat{\nu}}{\partial \boldsymbol{\zeta}^{\boldsymbol{\beta}}}(\boldsymbol{\zeta})\right| \|\boldsymbol{\zeta}\|_1\, d\boldsymbol{\zeta} + \frac{d\, M(B)}{2t_n} \int_{\mathbb{R}^d} \left|\frac{\partial^{|\boldsymbol{\beta}|}\widehat{\nu}}{\partial \boldsymbol{\zeta}^{\boldsymbol{\beta}}}(\boldsymbol{\zeta})\right| (\|\boldsymbol{\zeta}\|_2)^2\, d\boldsymbol{\zeta}
\end{aligned}$$

by using a Taylor–Lagrange expansion of order 2 on the second line. The first integral has been bounded in the proof of Lemma 2.16 by

$$d^{\frac{d+4}{2}} \left(\rho(K^{-1})\right)^{\frac{d+2}{2}} \sqrt{(2\pi)^d\,(d+1)!}\,.$$

On the other hand, setting $M = K^{-1/2}$ and $\boldsymbol{\zeta} = M\boldsymbol{\xi}$, the second integral is smaller than

$$\begin{aligned}
&\rho(K^{-1}) \sum_{l} \sum_{j_{11},\ldots,j_{d\beta^{(d)}}} \left(\prod_{i=1}^{d} \prod_{k=1}^{\beta^{(i)}} |M_{i j_{ik}}|\right) \int_{\mathbb{R}^d} |H_{\boldsymbol{\alpha}(\mathbf{j})}(\boldsymbol{\xi})|\, |\xi^{(l)}|^2\, \mathrm{e}^{-\frac{\|\boldsymbol{\xi}\|^2}{2}}\, d\boldsymbol{\xi} \\
&\leq d\,\rho(K^{-1}) \sum_{j_{11},\ldots,j_{d\beta^{(d)}}} \left(\prod_{i=1}^{d} \prod_{k=1}^{\beta^{(i)}} |M_{i j_{ik}}|\right) \sqrt{(2\pi)^d\,(d+1)!} \\
&\leq d^{\frac{d+3}{2}} \left(\rho(K^{-1})\right)^{\frac{d+3}{2}} \sqrt{(2\pi)^d\,(d+1)!}\,.
\end{aligned}$$

So, the term with multi-index $\boldsymbol{\alpha} = \mathbf{0}$ gives a contribution smaller than

$$\left(\frac{\|\nabla\theta_n(\mathbf{0}) - \nabla\theta(\mathbf{0})\|_\infty}{\sqrt{t_n}} + \frac{\sqrt{d\,\rho(K^{-1})}\, M(B)}{2t_n}\right) d^{\frac{d+4}{2}} \left(\rho(K^{-1})\right)^{\frac{d+2}{2}} \sqrt{(2\pi)^d\,(d+1)!}\,.$$

(1) If $|\boldsymbol{\alpha}| = 1$, then $\boldsymbol{\alpha} = (0,\ldots,0,1_i,0,\ldots,0)$ for some $i \in [\![1,d]\!]$. The derivative of δ_n is then

$$\left|\frac{\partial \delta_n(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}^{\boldsymbol{\alpha}}}\right| = \left|\frac{\partial \theta_n(\boldsymbol{\zeta})}{\partial \zeta^{(i)}} - \frac{\partial \theta(\mathbf{0})}{\partial \zeta^{(i)}}\right| \leq \|\nabla\theta_n(\mathbf{0}) - \nabla\theta(\mathbf{0})\|_\infty + M(B)\,\|\boldsymbol{\zeta}\|_1$$

if $\boldsymbol{\zeta} \in C^d_{(\mathbf{0},B(2d+2)^{3/2})}$. Therefore, the term with multi-index $\boldsymbol{\alpha} = (0,\ldots,1_i,\ldots,0)$ gives a contribution to the sum smaller than

$$\begin{aligned}&\frac{\beta^{(i)}}{\sqrt{t_n}} \int_{\mathbb{R}^d} \left|\frac{\partial^{|\boldsymbol{\beta}-\boldsymbol{\alpha}|}\widehat{\nu}(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}^{\boldsymbol{\beta}-\boldsymbol{\alpha}}}\right| \left(\|\nabla\theta_n(\mathbf{0}) - \nabla\theta(\mathbf{0})\|_\infty + \frac{M(B)\,\|\boldsymbol{\zeta}\|_1}{\sqrt{t_n}}\right) d\boldsymbol{\zeta}\\
&\leq \beta^{(i)} \left(\frac{\|\nabla\theta_n(\mathbf{0}) - \nabla\theta(\mathbf{0})\|_\infty}{\sqrt{t_n}} + \frac{\sqrt{d^3\,\rho(K^{-1})}\,M(B)}{t_n}\right) (d\rho(K^{-1}))^{\frac{d}{2}} \sqrt{(2\pi)^d\,d!}\end{aligned}$$

(2) Finally, if $|\boldsymbol{\alpha}| \geq 2$, then the derivative with multi-index $\boldsymbol{\alpha}$ of δ_n is the same as the derivative of θ_n, so one can use the same bounds as in Lemma 2.16. Thus, the term with multi-index $\boldsymbol{\alpha}$ gives a contribution smaller than

$$M(B)\binom{\boldsymbol{\beta}}{\boldsymbol{\alpha}}\left(\frac{1}{\sqrt{t_n\,d\,\rho(K^{-1})}}\right)^{|\boldsymbol{\alpha}|} (d\,\rho(K^{-1}))^{\frac{d+1}{2}} \sqrt{2\pi\,|\boldsymbol{\beta}-\boldsymbol{\alpha}|!}.$$

The claim follows by summing over the three kind of multi-indices $\boldsymbol{\alpha}$. ☐

Theorem 2.20 (Convex distance to the modified normal approximation)**.** *Let $(\mathbf{X}_n)_{n\in\mathbb{N}}$ be a sequence of random variables that satisfies Hypothesis (BE2). We denote μ_n the law of $\frac{\mathbf{X}_n}{\sqrt{t_n}}$, and ν_n the modified Gaussian distribution whose Fourier transform is*

$$\widehat{\nu}_n(\boldsymbol{\zeta}) = \mathrm{e}^{-\frac{\boldsymbol{\zeta}^t K \boldsymbol{\zeta}}{2}} \left(1 + \frac{\boldsymbol{\zeta}^t\,\nabla\theta(\mathbf{0})}{\sqrt{t_n}}\right).$$

One has

$$d_{\mathrm{convex}}(\mu_n,\nu_n) = o\left(\frac{1}{\sqrt{t_n}}\right).$$

Proof. One can assume without loss of generality that $\det K = 1$. Fix $\eta > 0$, and take $B = \frac{1}{\eta}$, and $\varepsilon = \frac{1}{B\sqrt{t_n}} = \frac{\eta}{\sqrt{t_n}}$. By Corollary 2.14 and Lemma 2.18,

$$d_{\mathrm{convex}}(\mu_n,\nu_n) \leq_{n\to\infty} \frac{2}{1-\frac{4}{9\pi}} \left((d+1)^{\frac{d+1}{2}}\,\Delta_{\frac{1}{B\sqrt{t_n}}}(\widehat{\mu}_n,\widehat{\nu}_n) + 2\sqrt{(d+1)\,\rho(K^{-1})}\,\frac{\eta}{\sqrt{t_n}}\right).$$

By hypothesis, $\nabla(\theta_n - \theta)$ converges locally uniformly to 0. Therefore, in the estimate of $\Delta_{\frac{1}{B\sqrt{t_n}}}(\widehat{\mu}_n,\widehat{\nu}_n)$ given by Lemma 2.19:

- The term that is proportional to $\frac{\|\nabla\theta_n(\mathbf{0})-\nabla\theta(\mathbf{0})\|_\infty}{\sqrt{t_n}}$ gets smaller than $\frac{\eta}{\sqrt{t_n}}$ for t_n large enough.
- This is also trivially true for the term proportional to $\frac{M(B)}{t_n}$, since B is fixed.

Thus, there is a constant $C_3(d,\rho(K^{-1}))$ such that, for n large enough,

$$d_{\mathrm{convex}}(\mu_n,\nu_n) \leq C_3(d,\rho(K^{-1}))\,\frac{\eta}{\sqrt{t_n}}.$$

Since this is true for any $\eta > 0$, the result is proven. ☐

Remark 2.21. There is an important precision that one can make to the statement of Theorem 2.20: if $\varepsilon_n = \sqrt{t_n}\, d_{\mathrm{convex}}(\mu_n, \nu_n)$, then the speed of convergence of $(\varepsilon_n)_{n\in\mathbb{N}}$ to 0 only depends on d, on $\rho(K^{-1})$ and on the speed of convergence to 0 of $\|\nabla(\theta_n - \theta)(\mathbf{0})\|_\infty$. By this, we mean that if $((\mu_n^{(i)}, \nu_n^{(i)})_{n\in\mathbb{N}})_{i\in I}$ are families of sequences of distributions on $\mathbb{R}^d$ that all satisfy the hypotheses above, and with a uniform bound on $\rho((K^{(i)})^{-1})$ and on $\|\nabla(\theta_n^{(i)} - \theta^{(i)})(\mathbf{0})\|_\infty$, then the corresponding sequences $(\varepsilon_n^{(i)})_{i\in I}$ converge uniformly to 0:

$$\lim_{n\to\infty}\left(\sup_{i\in I} \varepsilon_n^{(i)}\right) = 0.$$

We shall use this precision during the proof of Theorem 3.6.

If $\nabla\theta(\mathbf{0}) = 0$, then Theorem 2.20 ensures that the distance computed in Theorem 1.9 is a $o(1/\sqrt{t_n})$ instead of a $O(1/\sqrt{t_n})$. In this situation, one can develop a theory of zones of control, which is analogous to the hypotheses of Theorem 1.10 and, and which gives upper bounds

$$d_{\mathrm{convex}}(\mu_n, \mathcal{N}_{\mathbb{R}^d}(\mathbf{0}, K)) = O\Big((t_n)^{-\frac{1}{2}-\gamma}\Big)\,.$$

As the main examples that we have in mind for this theory all rely on the method of cumulants, we postpone the exposition of these better estimates to Section 4.

3. Large deviation results

In this section, we fix a sequence $(\mathbf{X}_n)_{n\in\mathbb{N}}$ of random vectors in $\mathbb{R}^d$ that is mod-Gaussian convergent in the Laplace sense on a multi-strip $\mathcal{S}_{\mathbf{a},\mathbf{b}}$, with $\mathbf{0} \in \prod_{i=1}^d (a^{(i)}, b^{(i)})$. As before, $(t_n K)_{n\in\mathbb{N}}$ is the sequence of parameters, and the limiting function is denoted by ψ. We also suppose that ψ does not vanish on the real part of the multi-strip $\mathcal{S}_{\mathbf{a},\mathbf{b}}$; this hypothesis will be useful in certain arguments of exponential change of measure. We are interested in the asymptotics of the probabilities

$$\mathbb{P}[\mathbf{X}_n \in t_n B],$$

where $B = [b, +\infty) \times S$, and S is some measurable part of the $(d-1)$-dimensional K-ellipsoid $\mathbb{S}^{d-1}(K) = K^{-1/2}(\mathbb{S}^{d-1}) = \{\mathbf{x} \in \mathbb{R}^d \,|\, \mathbf{x}^t K \mathbf{x} \leq 1\}$. For instance, when $d = 2$, assuming to simplify that $K = I_2$, we want to be able to deal with angular sectors

$$B = \{z = r\mathrm{e}^{\mathrm{i}\theta} \in \mathbb{C} \mid r \geq b,\ \theta \in (\theta_1, \theta_2)\}.$$

This problem is the multi-dimensional generalisation of the discussion of [FMN16, Section 4] and of Theorem 1.8. In Subsection 3.1, we study a toy model in order to make a correct conjecture for the asymptotics of $\mathbb{P}[\mathbf{X}_n \in t_n B]$. In Subsection 3.2, we use Theorem 2.20 and a method of tilting of measures in order to obtain the precise asymptotics of the probabilities of certain cones. Finally, in Subsection 3.3, we detail the approximation of spherical sectors $B = [b, +\infty) \times S$ by unions of cones, thereby getting the asymptotic formula conjectured in Subsection 3.1 (Theorem 3.6). Examples will be examined in Section 4.

3.1. **Sum of a Gaussian noise and an independent random variable.** In order to make a correct conjecture, we consider the trivial example of mod-Gaussian convergence

$$\mathbf{X}_n = \sqrt{t_n}\,\mathbf{G} + \mathbf{Y},$$

where $\mathbf{G}$ is a random vector with law $\mathcal{N}_{\mathbb{R}^d}(0, I_d)$, and $\mathbf{Y}$ is a bounded random vector independent from $\mathbf{G}$. If $\mathbb{E}[\mathrm{e}^{\langle \mathbf{z} \,|\, \mathbf{Y}\rangle}] = \psi(\mathbf{z})$, then $\psi(\mathbf{z})$ is well-defined over $\mathbb{C}^d$ since $\mathbf{Y}$ is bounded, and

$$\psi_n(\mathbf{z}) = \mathbb{E}\left[\mathrm{e}^{\langle \mathbf{z} \,|\, \mathbf{X}_n\rangle}\right]\,\mathrm{e}^{-\frac{t_n \|\mathbf{z}\|^2}{2}} = \psi(\mathbf{z}).$$

Therefore, $(\mathbf{X}_n)_{n\in\mathbb{N}}$ is mod-Gaussian convergent on $\mathbb{C}^d$ with parameters $t_n I_d$ and limit $\psi(\mathbf{z})$. To simplify a bit, we also assume that $\mathbf{Y}$ has a density $\nu(\mathbf{y})\,d\mathbf{y}$ with respect to the Lebesgue measure. Then, a Borel set B being fixed, we have

$$\begin{aligned}\mathbb{P}[\mathbf{X}_n \in t_n B] &= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbf{x}\in\mathbb{R}^d} \int_{\mathbf{y}\in\mathbb{R}^d} \mathrm{e}^{-\frac{\|\mathbf{x}\|^2}{2}}\, 1_{(\sqrt{t_n}\mathbf{x}+\mathbf{y}\,\in\, t_n B)}\, \nu(\mathbf{y})\, d\mathbf{x}\, d\mathbf{y} \\ &= \left(\frac{t_n}{2\pi}\right)^{\frac{d}{2}} \int_{\mathbf{u}\in\mathbb{R}^d} \int_{\mathbf{y}\in\mathbb{R}^d} \mathrm{e}^{-\frac{t_n\|\mathbf{u}\|^2}{2}}\, 1_{(\mathbf{u}+(t_n)^{-1}\mathbf{y}\,\in\, B)}\, \nu(\mathbf{y})\, d\mathbf{u}\, d\mathbf{y} \\ &= \left(\frac{t_n}{2\pi}\right)^{\frac{d}{2}} \int_{\mathbf{b}\in B} \mathrm{e}^{-\frac{t_n\|\mathbf{b}\|^2}{2}}\, d\mathbf{b} \left(\int_{\mathbf{y}\in\mathbb{R}^d} \mathrm{e}^{\langle \mathbf{b}\,|\,\mathbf{y}\rangle}\, \mathrm{e}^{-\frac{\|\mathbf{y}\|^2}{2t_n}}\, \nu(\mathbf{y})\, d\mathbf{y}\right).\end{aligned}$$

By Lebesgue's dominated convergence theorem, the term in parentheses converges to the integral $\int_{\mathbb{R}^d} \mathrm{e}^{\langle \mathbf{b}\,|\,\mathbf{y}\rangle}\, \nu(\mathbf{y})\, d\mathbf{y} = \psi(\mathbf{b})$. This convergence is locally uniform in the parameter $\mathbf{b}$. Consequently, on any bounded subset of $\mathbb{R}^d$, one can write

$$\int_{\mathbf{y}\in\mathbb{R}^d} \mathrm{e}^{\langle \mathbf{b}\,|\,\mathbf{y}\rangle}\, \mathrm{e}^{-\frac{\|\mathbf{y}\|^2}{2t_n}}\, \nu(\mathbf{y})\, d\mathbf{y} = \psi(\mathbf{b})\,(1+o(1))$$

with a uniform $o(1)$. We now suppose that $B = [b,+\infty) \times S$, where $b > 0$ and S is a measurable part of the sphere $\mathbb{S}^{d-1}$. We split B in two parts

$$B_{<b+\varepsilon} = [b, b+\varepsilon) \times S \qquad ; \qquad B_{\geq b+\varepsilon} = [b+\varepsilon, +\infty) \times S,$$

where ε is a small parameter to be chosen later. We denote $I_{<b+\varepsilon}$ and $I_{\geq b+\varepsilon}$ the corresponding parts of the integral form of $\mathbb{P}[\mathbf{X}_n \in t_n B]$ computed above. For the integral $I_{<b+\varepsilon}$, recall that for any non-negative measurable function f over $\mathbb{R}^d$, one can make the polar change of coordinates

$$\int_{\mathbb{R}^d} f(\mathbf{x})\, d\mathbf{x} = \int_{r=0}^{\infty} \left(\int_{\mathbb{S}^{d-1}} f(r\mathbf{s})\ \mu_{\mathbb{S}^{d-1}}(d\mathbf{s})\right) r^{d-1}\, dr,$$

where $\mu_{\mathbb{S}^{d-1}}$ is the unique $\mathrm{SO}(d)$-invariant measure on the sphere with total mass $\frac{2\,\pi^{d/2}}{\Gamma(d/2)}$. Therefore,

$$\begin{aligned}(1+o(1))\, I_{<b+\varepsilon} &= \left(\frac{t_n}{2\pi}\right)^{\frac{d}{2}} \int_{\mathbf{b}\in B_{<b+\varepsilon}} \mathrm{e}^{-\frac{t_n\|\mathbf{b}\|^2}{2}}\, \psi(\mathbf{b})\, d\mathbf{b} \\ &= \left(\frac{t_n}{2\pi}\right)^{\frac{d}{2}} \int_{r=b}^{b+\varepsilon} \left(\int_S \psi(r\mathbf{s})\, \mu_{\mathbb{S}^{d-1}}(d\mathbf{s})\right) r^{d-1}\, \mathrm{e}^{-\frac{t_n r^2}{2}}\, dr.\end{aligned}$$

In the formula above, the multiplicative factor $(1+o(1))$ corresponds to the replacement of the integral $\int_{\mathbf{y}\in\mathbb{R}^d} \mathrm{e}^{\langle \mathbf{b}\,|\,\mathbf{y}\rangle}\, \mathrm{e}^{-\|\mathbf{y}\|^2/2t_n}\, \nu(\mathbf{y})\, d\mathbf{y}$ by $\psi(\mathbf{b})$. Since ψ is a continuous function, one can write uniformly on the sphere $\psi(r\mathbf{s}) = \psi(b\mathbf{s})\,(1+o(1))$ for any $r \in [b, b+\varepsilon)$ if $\varepsilon = o(1)$. So,

$$\begin{aligned}(1+o(1))\, I_{<b+\varepsilon} &= \left(\frac{t_n}{2\pi}\right)^{\frac{d}{2}} \left(\int_{r=b}^{b+\varepsilon} r^{d-1}\, \mathrm{e}^{-\frac{t_n r^2}{2}}\, dr\right) \left(\int_S \psi(b\mathbf{s})\, \mu_{\mathbb{S}^{d-1}}(d\mathbf{s})\right) \\ &= \left(\frac{t_n}{2\pi}\right)^{\frac{d}{2}} \frac{b^{d-2}}{t_n} \left(\mathrm{e}^{-\frac{t_n b^2}{2}} - \mathrm{e}^{-\frac{t_n(b+\varepsilon)^2}{2}}\right) \left(\int_S \psi(b\mathbf{s})\, \mu_{\mathbb{S}^{d-1}}(d\mathbf{s})\right) \\ &= \left(\frac{t_n}{2\pi}\right)^{\frac{d}{2}} \frac{b^{d-2}}{t_n}\, \mathrm{e}^{-\frac{t_n b^2}{2}} \left(1 - \mathrm{e}^{-\frac{t_n\varepsilon^2}{2} - t_n b\varepsilon}\right) \left(\int_S \psi(b\mathbf{s})\, \mu_{\mathbb{S}^{d-1}}(d\mathbf{s})\right).\end{aligned}$$

Consider now the integral $I_{\geq b+\varepsilon}$. Since $B_{\geq b+\varepsilon} \subset [b+\varepsilon,+\infty) \times \mathbb{S}^{d-1}$ and $\|\mathbf{Y}\|$ is bounded almost surely by some constant C, we have

$$\begin{aligned}
I_{\geq b+\varepsilon} &\leq \left(\frac{t_n}{2\pi}\right)^{\frac{d}{2}} \int_{\|\mathbf{b}\|\geq b+\varepsilon} \mathrm{e}^{-\frac{t_n\|\mathbf{b}\|^2}{2}}\, \mathrm{e}^{C\|\mathbf{b}\|}\, d\mathbf{b} \\
&\leq \frac{2}{\Gamma(\frac{d}{2})} \left(\frac{t_n}{2}\right)^{\frac{d}{2}} \int_{b+\varepsilon}^{\infty} w^{d-1}\, \mathrm{e}^{-\frac{t_n r^2}{2}}\, \mathrm{e}^{Cr}\, dr \\
&\leq \frac{2\,\mathrm{e}^{\frac{C^2}{2t_n}}}{\Gamma(\frac{d}{2})} \left(\frac{t_n}{2}\right)^{\frac{d}{2}} \int_{b+\varepsilon-\frac{C}{t_n}}^{\infty} \left(x+\frac{C}{t_n}\right)^{d-1} \mathrm{e}^{-\frac{t_n x^2}{2}}\, dx \\
&\leq \frac{2\,\mathrm{e}^{\frac{C^2}{2t_n}}}{\Gamma(\frac{d}{2})} \left(\frac{t_n}{2}\right)^{\frac{d}{2}} \left(1+\frac{C}{t_n b}\right)^{d-1} \int_{b+\varepsilon-\frac{C}{t_n}}^{\infty} x^{d-1}\, \mathrm{e}^{-\frac{t_n x^2}{2}}\, dx \\
&\leq \frac{1}{\Gamma(\frac{d}{2})} \left(\frac{t_n}{2}\right)^{\frac{d}{2}-1} \left(1+\frac{C}{t_n b}\right)^{d-1} \left(1-\frac{d-2}{t_n b^2}\right)^{-1} \mathrm{e}^{-\frac{t_n(b+\varepsilon)^2}{2}}\, \mathrm{e}^{C(b+\varepsilon)}\, (b+\varepsilon)^{d-2}
\end{aligned}$$

assuming $\frac{C}{t_n} \leq \varepsilon$, and using integration by parts at the end to estimate the Gaussian integral. For t_n large enough, we have $t_n b \geq C$ and $t_n b^2 \geq 2(d-2)$, which leads to

$$I_{\geq b+\varepsilon} \leq M(d)\,(t_n)^{\frac{d}{2}-1}\, \mathrm{e}^{-\frac{t_n(b+\varepsilon)^2}{2}}\, \mathrm{e}^{C(b+\varepsilon)}\, (b+\varepsilon)^{d-2},$$

where $M(d)$ is a certain constant that depends only on d. We take $\varepsilon = (t_n)^{-1/3}$. For n large enough, we have indeed $\frac{C}{t_n} \leq \varepsilon$, and from the previous estimates:

$$\begin{aligned}
I_{\geq b+\varepsilon} &\leq M(d)\,(t_n)^{\frac{d}{2}}\, \frac{b^{d-2}}{t_n}\, \mathrm{e}^{Cb}\, \mathrm{e}^{-\frac{t_n b^2}{2}}\, \mathrm{e}^{-\frac{(t_n)^{1/3}}{2}}; \\
(1+o(1))\, I_{<b+\varepsilon} &= \left(\frac{t_n}{2\pi}\right)^{\frac{d}{2}} \frac{b^{d-2}}{t_n}\, \mathrm{e}^{-\frac{t_n b^2}{2}} \left(\int_S \psi(b\mathbf{s})\, \mu_{\mathbb{S}^{d-1}}(d\mathbf{s})\right).
\end{aligned}$$

Therefore, assuming that S has a non-zero surface measure in $\mathbb{S}^{d-1}$, $I_{\geq b+\varepsilon}$ is negligible in comparison to the other integral $I_{<b+\varepsilon}$, and we obtain the asymptotic formula

$$\mathbb{P}[\mathbf{X}_n \in t_n B] \simeq \left(\frac{t_n}{2\pi}\right)^{\frac{d}{2}} \frac{b^{d-2}}{t_n}\, \mathrm{e}^{-\frac{t_n b^2}{2}} \left(\int_S \psi(b\mathbf{s})\, \mu_{\mathbb{S}^{d-1}}(d\mathbf{s})\right).$$

It is convenient to rewrite the spherical integral as an integral over the sphere of radius b, which we denote $\mathbb{S}^{d-1}(b)$. We also denote μ_{surface} the Lebesgue surface measure on it, which has total mass

$$\mu_{\text{surface}}(\mathbb{S}^{d-1}(b)) = \frac{2\,\pi^{\frac{d}{2}}\, b^{d-1}}{\Gamma(\frac{d}{2})}.$$

Then, the previous formula rewrites as

$$\mathbb{P}[\mathbf{X}_n \in t_n B] \simeq \left(\frac{t_n}{2\pi}\right)^{\frac{d}{2}} \frac{1}{t_n b}\, \mathrm{e}^{-\frac{t_n b^2}{2}} \left(\int_{S_b} \psi(\mathbf{s})\, \mu_{\text{surface}}(d\mathbf{s})\right),$$

where $S_b = S \times b = \{b\mathbf{s},\ \mathbf{s} \in S\}$ is the base of the spherical sector $B = S \times [b,+\infty)$ under consideration. Our goal will then be to show that this asymptotic formula actually holds in the general setting of multi-dimensional mod-Gaussian convergence.

3.2. **Asymptotics of the probabilities of cones.** Until further notice, we assume that $K = I_d$; the general case will be treated by using the arguments of the remark at the end of Subsection 1.2. In this paragraph, we shall compute the asymptotics of the probabilities $\mathbb{P}[\mathbf{X}_n \in t_n C]$, where

$$C = [1, +\infty) \times D = \{r\mathbf{d},\ r \geq 1 \text{ and } \mathbf{d} \in D\}$$

and D is a convex domain of an hyperplane H that does not contain the origin $\mathbf{0}$ of $\mathbb{R}^d$; see Figure 2.

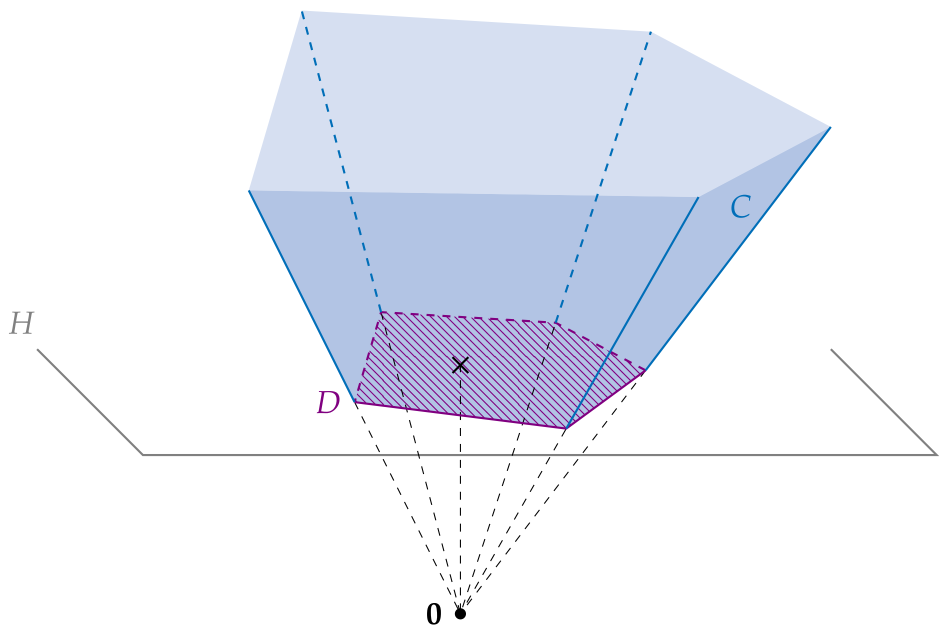

FIGURE 2. Cone C based on a convex domain D of an hyperplane H.

We denote $\mathbf{h}$ the orthogonal projection of the origin $\mathbf{0}$ on the affine hyperplane H, and we assume that it is contained in the hypercube $\prod_{i=1}^d (a^{(i)}, b^{(i)})$, where $\mathcal{S}_{\mathbf{a},\mathbf{b}}$ is the multi-strip of mod-Gaussian convergence of the sequence $(\mathbf{X}_n)_{n\in\mathbb{N}}$. If ρ_n is the law of $\mathbf{X}_n$, we denote $\widetilde{\rho}_n$ the new probability measure

$$\widetilde{\rho}_n(d\mathbf{x}) = \frac{\mathrm{e}^{\langle \mathbf{h} \,|\, \mathbf{x}\rangle}}{\mathbb{E}[\mathrm{e}^{\langle \mathbf{h} \,|\, \mathbf{X}_n\rangle}]}\, \rho_n(d\mathbf{x}).$$

If $\widetilde{\mathbf{X}}_n$ follows the law $\widetilde{\rho}_n$, then the new sequence $(\widetilde{\mathbf{X}}_n - t_n\mathbf{h})_{n\in\mathbb{N}}$ is again mod-Gaussian convergent in the Laplace sense on the multi-strip $\mathcal{S}_{\mathbf{a}-\mathbf{h},\mathbf{b}-\mathbf{h}}$:

$$\mathbb{E}\left[\mathrm{e}^{\langle \mathbf{z} \,|\, \widetilde{\mathbf{X}}_n\rangle}\right] = \frac{\mathbb{E}[\mathrm{e}^{\langle \mathbf{z}+\mathbf{h} \,|\, \mathbf{X}_n\rangle}]}{\mathbb{E}[\mathrm{e}^{\langle \mathbf{h} \,|\, \mathbf{X}_n\rangle}]} = \mathrm{e}^{\frac{t_n\|\mathbf{z}\|^2}{2}}\, \mathrm{e}^{t_n\langle \mathbf{z} \,|\, \mathbf{h}\rangle}\, \frac{\psi_n(\mathbf{z}+\mathbf{h})}{\psi_n(\mathbf{h})};$$

$$\mathrm{e}^{-\frac{t_n\|\mathbf{z}\|^2}{2}}\, \mathbb{E}\left[\mathrm{e}^{\langle \mathbf{z} \,|\, \widetilde{\mathbf{X}}_n - t_n\mathbf{h}\rangle}\right] = \frac{\psi_n(\mathbf{z}+\mathbf{h})}{\psi_n(\mathbf{h})} \to_{n\to\infty} \frac{\psi(\mathbf{z}+\mathbf{h})}{\psi(\mathbf{h})}.$$

Then,

$$\begin{aligned}\mathbb{P}[\mathbf{X}_n \in t_n C] &= \int_{\mathbb{R}^d} 1_{(\mathbf{x}\in t_n C)}\, \rho_n(d\mathbf{x}) = \mathbb{E}\left[\mathrm{e}^{\langle \mathbf{h} \,|\, \mathbf{X}_n\rangle}\right] \int_{\mathbb{R}^d} 1_{(\mathbf{x}\in t_n C)}\, \mathrm{e}^{-\langle \mathbf{h} \,|\, \mathbf{x}\rangle}\, \widetilde{\rho}_n(d\mathbf{x}) \\ &= \mathrm{e}^{-\frac{t_n\|\mathbf{h}\|^2}{2}}\, \psi_n(\mathbf{h}) \int_{\mathbb{R}^d} 1_{(\mathbf{y}\in\sqrt{t_n}C_0)}\, \mathrm{e}^{-\sqrt{t_n}\langle \mathbf{h} \,|\, \mathbf{y}\rangle}\, \mu_n(d\mathbf{y})\end{aligned}$$

where μ_n denotes the law of $(\widetilde{\mathbf{X}}_n - t_n\mathbf{h})/\sqrt{t_n}$, and $C_0 = C - \mathbf{h} = \{\mathbf{y} \in \mathbb{R}^d \,|\, \mathbf{y} + \mathbf{h} \in C\}$. To compute the integral I, notice that if $D_0 = D - \mathbf{h}$, then every element of D_0 is orthogonal to $\mathbf{h}$, and on the other hand,

$$C_0 = \{\mathbf{y} = s\mathbf{h} + (s+1)\mathbf{d} \,|\, s \in \mathbb{R}_+ \text{ and } \mathbf{d} \in D_0\}.$$

We set $C_{(0,s)} = \{\mathbf{y} = t\mathbf{h} + (t+1)\mathbf{d} \,|\, t \in [0,s] \text{ and } \mathbf{d} \in D_0\}$; by definition, $C_0 = C_{(0,+\infty)}$. On the other hand,

$$\begin{aligned} I &= \int_{s=0}^{\infty} \frac{d(\mu_n(\sqrt{t_n}\,C_{(0,s)}))}{ds}\, \mathrm{e}^{-t_n s \|\mathbf{h}\|^2}\, ds \\ &= t_n\|\mathbf{h}\|^2 \int_{s=0}^{\infty} \mu_n(\sqrt{t_n}\,C_{(0,s)})\, \mathrm{e}^{-t_n s\|\mathbf{h}\|^2}\, ds \end{aligned}$$

by using on the first line the identity $\mathrm{e}^{-\sqrt{t_n}\langle \mathbf{h} \,|\, \mathbf{y}\rangle} = \mathrm{e}^{-t_n s\|\mathbf{h}\|^2}$ if $\mathbf{y} = \sqrt{t_n}(s\mathbf{h} + (s+1)\mathbf{d})$, and on the second line an integration by parts. We can now introduce the modified Gaussian measure

$$\nu_n(d\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}}\, \mathrm{e}^{-\frac{\|\mathbf{x}\|^2}{2}} \left(1 + \frac{\langle \nabla\psi(\mathbf{h}) \mid \mathbf{x}\rangle}{\psi(\mathbf{h})\sqrt{t_n}}\right) d\mathbf{x}.$$

By Theorem 2.20, for every Borel convex subset of $\mathbb{R}^d$, $\mu_n(C) = \nu_n(C) + o(\frac{1}{\sqrt{t_n}})$, with a $o(\cdot)$ that is uniform over the class of all Borel convex subsets. Therefore,

$$I = \int_{s=0}^{\infty} \nu_n(\sqrt{t_n}\,C_{(0,s)})\, t_n\|\mathbf{h}\|^2\, \mathrm{e}^{-t_n s\|\mathbf{h}\|^2}\, ds + o\left(\frac{1}{\sqrt{t_n}}\right)$$

with a $o(\cdot)$ that does not depend on the cone C. Denote J the last integral; we have

$$\begin{aligned} &\frac{J}{(t_n)^{\frac{d}{2}}\|\mathbf{h}\|} \\ &= \int_{\substack{t\geq 0\\ \mathbf{d}\in D_0}} \frac{(t+1)^{d-1}}{(2\pi)^{\frac{d}{2}}} \left(1 + \frac{\langle \nabla\psi(\mathbf{h}) \mid t\mathbf{h} + (t+1)\mathbf{d}\rangle}{\psi(\mathbf{h})}\right) \mathrm{e}^{-t_n\left((t+\frac{t^2}{2})(\|\mathbf{h}\|^2+\|\mathbf{d}\|^2)+\frac{\|\mathbf{d}\|^2}{2}\right)}\, dt\, d\mathbf{d}. \end{aligned}$$

We apply the Laplace method to

$$\begin{aligned} &\int_{t=0}^{\infty} \frac{(t+1)^{d-1}}{(2\pi)^{\frac{d}{2}}} \left(1 + \frac{\langle \nabla\psi(\mathbf{h}) \mid t\mathbf{h} + (t+1)\mathbf{d}\rangle}{\psi(\mathbf{h})}\right) \mathrm{e}^{-t_n\left((t+\frac{t^2}{2})(\|\mathbf{h}\|^2+\|\mathbf{d}\|^2)+\frac{\|\mathbf{d}\|^2}{2}\right)}\, dt \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}} \left(1 + \frac{\langle \nabla\psi(\mathbf{h}) \mid \mathbf{d}\rangle}{\psi(\mathbf{h})} + O\left(\frac{1}{t_n}\right)\right) \frac{\mathrm{e}^{-\frac{t_n\|\mathbf{d}\|^2}{2}}}{t_n(\|\mathbf{h}\|^2 + \|\mathbf{d}\|^2)}, \end{aligned}$$

see [Zor04, §19.2]. In this approximation, it is easily seen that if D_0 is a bounded domain and if $\mathbf{h}$ also stays bounded, then the constant in the $O(\cdot)$ of the remainder can be taken uniformly. We conclude:

Proposition 3.1 (Asymptotics of the probabilities of cones)**.** *Consider a mod-Gaussian convergent sequence $(\mathbf{X}_n)_{n\in\mathbb{N}}$ with parameters $t_n I_d$. We fix a bound $M > 0$, and we suppose that:*

- *The vector $\mathbf{h}$ belongs to $\prod_{i=1}^d (a^{(i)}, b^{(i)})$, and $\|\mathbf{h}\| \leq M$,*
- *The domain D_0 is a convex part of $\mathbf{h}^\perp$ such that $\sup\{\|\mathbf{d}\| \in D_0\} \leq M$.*

If C is the cone based on the domain $D = D_0 + \mathbf{h}$, then

$$\begin{aligned} &(2\pi)^{\frac{d}{2}}\, \mathrm{e}^{\frac{t_n\|\mathbf{h}\|^2}{2}}\, \frac{1}{\psi_n(\mathbf{h})}\, \mathbb{P}[\mathbf{X}_n \in t_n C] \\ &= (t_n)^{\frac{d}{2}-1} \int_{\mathbf{d}\in D_0} \left(1 + \frac{\langle \nabla\psi(\mathbf{h}) \mid \mathbf{d}\rangle}{\psi(\mathbf{h})} + O\left(\frac{1}{t_n}\right)\right) \frac{\|\mathbf{h}\|\, \mathrm{e}^{-\frac{t_n\|\mathbf{d}\|^2}{2}}}{\|\mathbf{h}\|^2 + \|\mathbf{d}\|^2}\, d\mathbf{d} + o\left(\frac{1}{\sqrt{t_n}}\right) \end{aligned}$$

with constants in the $O(\frac{1}{t_n})$ *and the* $o(\frac{1}{\sqrt{t_n}})$ *that only depend on the bound* M *(and on the sequence* $(\mathbf{X}_n)_{n\in\mathbb{N}}$*).*

Proof. The only non-trivial fact that remains to be proven after the previous discussion is that the $o(1/\sqrt{t_n})$ can be taken uniformly if $\mathbf{h}$ and D_0 stay bounded by M. This is a consequence of the remark following Theorem 2.20. □

3.3. **Approximation of spherical sectors by unions of cones.** We still assume until further notice that $K = I_d$. Recall that our final goal is to compute the asymptotics of the probabilities $\mathbb{P}[\mathbf{X}_n \in t_n B]$, where B is a *spherical sector*, that is to say that it can be written as $B = [b, +\infty) \times S$ where S is some measurable part of the sphere $\mathbb{S}^{d-1}$. After Proposition 3.1, a natural method consists in approximating the spherical sector B by a disjoint union of cones with small bases D_0 and vectors $\mathbf{h}$ placed on the sphere of radius b; see Figure 3. This is only possible if S is a sufficiently regular subset of $\mathbb{S}^{d-1}$. We develop hereafter an *ad hoc* notion of regularity, which is akin to *Jordan mesurability* in $\mathbb{R}^d$, but with respect to the sphere. We were not able to find an adequate reference for the notion of Jordan mesurability on manifolds; for the Euclidean case, we refer to [Tao11, Section 1.1.2] and [Spi65, Chapter 3].

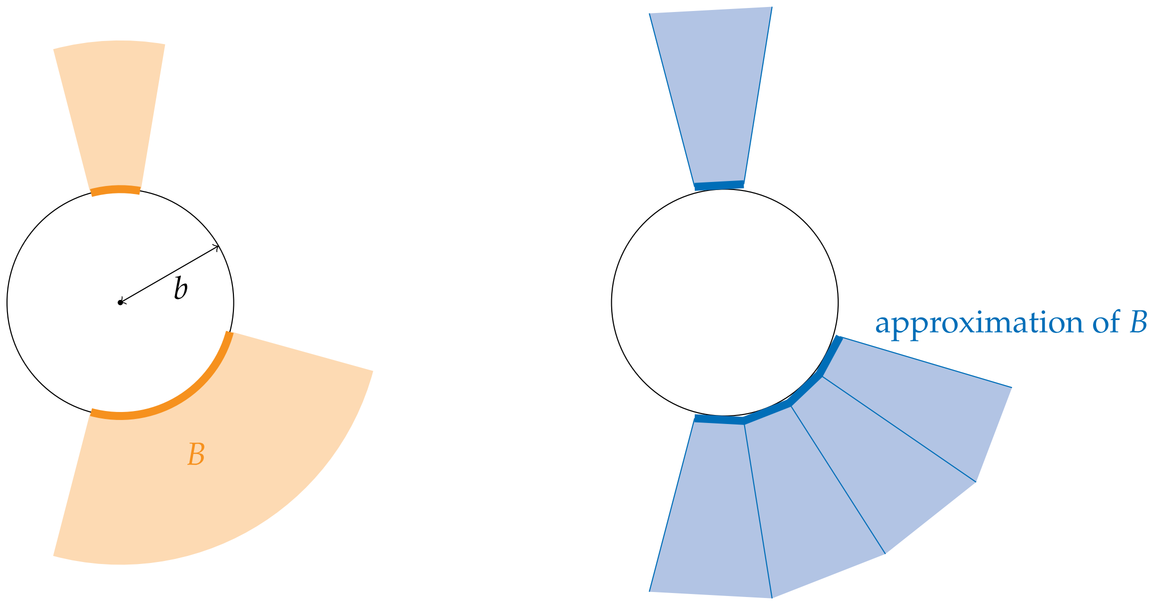

FIGURE 3. Approximation of a spherical sector $B = [b + \infty) \times S$ by a disjoint union of cones.

Let E be the hypercube $[-1, 1]^d$. If $A \subset \partial E$ is a rectangular part on one of the face of ∂E, we call *hypercubic facet* associated to A on the sphere of radius b the set of vectors

$$A_b = \{\mathbf{v} \in \mathbb{R}^d \mid \|\mathbf{v}\| = b \text{ and } \mathbf{v} \text{ is colinear to a vector in } A\},$$

see Figure 4 for an example in dimension $d = 3$.

The notion of hypercubic facet allows one to transform the rectangular partitions of the boundary ∂E of the hypercube into partitions of the sphere of radius b. Since rectangles can be used to define Jordan mesurability on $\mathbb{R}^d$ and thereby on ∂E, by projection, we are led to the analogous notion for subsets of the sphere:

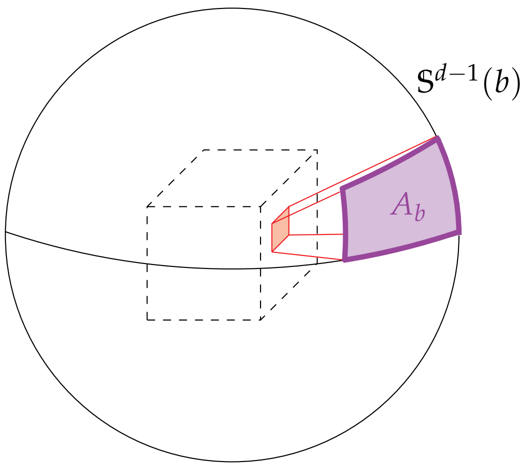

FIGURE 4. An hypercubic facet on a 2-dimensional sphere.

Definition 3.2 (Measurability by hypercubic facets)**.** *Let S be a measurable part of $\mathbb{S}^{d-1}(b)$. One says that S is measurable by hypercubic facets if, for every $\varepsilon > 0$, there exists two subsets T_1 and T_2 of the sphere such that:*

(1) T_1 and T_2 are finite unions of hypercubic facets on $\mathbb{S}^{d-1}(b)$;

(2) $T_1 \subset S \subset T_2$;

(3) $\mu_{\mathrm{surface}}(T_2 \setminus T_1) \leq \varepsilon$.

Proposition 3.3 (Topological criterion for measurability)**.** *A subset S of the sphere of radius $b > 0$ is mesurable by hypercubic facets if and only if its topological boundary ∂S has zero Lebesgue measure: $\mu_{\mathrm{surface}}(\partial S) = 0$.*

Proof. The map ψ which projects the hypercube ∂E onto the sphere $\mathbb{S}^{d-1}(b)$ is an homeomorphism which sends sets of Lebesgue measure zero to sets of Lebesgue surface measure zero, and conversely. On the other hand, $S \subset \mathbb{S}^{d-1}(b)$ is measurable by hypercubic facets if and only if $\psi^{-1}(S)$ is a subset of ∂E measurable by rectangles, that is to say Jordan measurable in the usual sense. However, it is well known that a bounded set of $\mathbb{R}^d$ is Jordan measurable if and only if its boundary has Lebesgue measure equal to zero; see [Spi65, p. 56]. So, S is measurable by hypercubic facets if and only if $\partial(\psi^{-1}(S)) = \psi^{-1}(\partial S)$ has zero Lebesgue measure, and by the remark at the beginning of the proof, this is equivalent to the fact that ∂S has zero Lebesgue surface measure. $\square$

Let us now relate the notion of mesurability by hypercubic facets to the problem of approximation of spherical sectors by unions of cones. To begin with, notice that if S_b is an hypercubic facet on $\mathbb{S}^{d-1}(b)$, then the corresponding spherical sector is naturally approximated by two cones with convex bases. Indeed, the boundary of the hypercubic facet S_b is the convex domain in the sense of geodesics of $\mathbb{S}^{d-1}$ that is delimited by 2^{d-1} points $\mathbf{v}_1, \ldots, \mathbf{v}_{2^{d-1}}$, which are the projections of the corners of the rectangle A included in ∂E that corresponds to $S_b = A_b$. Let $\mathbf{h}$ be the barycenter of these 2^{d-1} points (in the Euclidean sense); it is easily seen by recurrence on the dimension that $\|\mathbf{h}\| \leq b$, the inequality being strict as soon as the hypercubic facet is non-degenerate (*i.e.* with positive surface measure). For every positive real number $r > 0$, we can consider the domain $D(r)$ that is defined as the part of the affine hyperplane

$$H(r) = \left\{ r\frac{\mathbf{h}}{\|\mathbf{h}\|} + \mathbf{d},\ \mathbf{d} \in \mathbf{h}^{\perp} \right\}$$

which is the convex hull of the 2^{d-1} points of

$$H(r) \cap \left(\bigcup_{i=1}^{2^{d-1}} \mathbb{R}_+ \mathbf{v}_i \right).$$

For r small enough, the cone $C(r)$ based on the domain $D(r)$ contains S_b and therefore the whole spherical sector $S_b \times [1, +\infty)$. We call *negative approximation* of the spherical sector $B = S_b \times [1, +\infty)$ the cone $C(r)$ based on the domain $D(r)$, with $r \in (0, b]$ maximal such that

$$S_b \times [1, +\infty) \subset C(r).$$

We then denote $C(r) = C_-(B)$ this negative approximation; it satisfies $B \subset C_-(B)$, and in some sense it is the smallest possible cone with this property. On the other hand, note that the domain $D(b)$ lies on the outside of the sphere $\mathbb{S}^{d-1}(b)$, because it is a part of an hyperplane tangent to the sphere. Therefore, the spherical sector $B = S_b \times [1, +\infty)$ contains the cone $C(b)$, which we call the *positive approximation* of the spherical sector, and denote $C(b) = C_+(B)$. Thus, we have

$$C_+(B) \subset B \subset C_-(B).$$

We refer to Figure 5 for drawings in dimension $d = 2$ that clarify the discussion above (in dimension $d = 2$, hypercubic facets are simply arcs of circles). The terminology negative/positive recalls that the cone $C_-(B)$ is based on a domain $D(r)$ with $r \leq b$ (beware that the negative approximation is the largest cone).

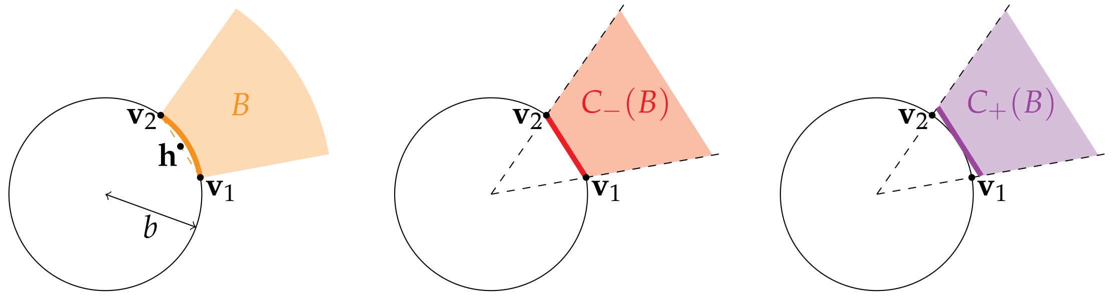

FIGURE 5. Negative and positive approximations of a spherical sector based on an hypercubic facet.

Consider an hypercubic facet S_b, with $C_-(B)$ and $C_+(B)$ negative and positive approximations of the spherical sector $B = S_b \times [1, +\infty)$. We say that the pair $(C_-(B), C_+(B))$ is an approximation of *level* $\varepsilon > 0$ of B if

$$\min\{\|\mathbf{v}\|,\ \mathbf{v} \in D_-(B)\} \geq b - \varepsilon;$$
$$\max\{\|\mathbf{v}\|,\ \mathbf{v} \in D_+(B)\} \leq b + \varepsilon,$$

where $D_-(B)$ and $D_+(B)$ are the two convex bases on which the cones $C_-(B)$ and $C_+(B)$ are based. The radius b being fixed, if the diameter (for instance in the geodesic sense) of an hypercubic facet S_b is smaller than ε, then the corresponding pair of approximations is of level $O(\varepsilon^2)$, where the constants in the $O(\cdot)$ depend only on b and the dimension d.

We are now ready to deal with approximations of general spherical sectors (not necessarily based on hypercubic facets). Let S be a part of $\mathbb{S}^{d-1}$ that is measurable by hypercubic facets, and $B = S \times [b, +\infty)$. A negative approximation of the spherical sector B is given by:

(1) a finite union T_2 of hypercubic facets $A_b^{(1)}, \dots, A_b^{(m)}$ of $\mathbb{S}^{d-1}(b)$, such that

$$T_2 = \bigcup_{i=1}^{m} A_b^{(i)} \supset S_b;$$

(2) and a radius $b_- \in (0, b)$, such that for every $i \in [\![1, m]\!]$,

$$b_- \leq r^{(i)}.$$

Here, $r^{(i)}$ is the radius such that $C^{(i)}(r^{(i)}) = C_-(B^{(i)})$; $B^{(i)}$ (respectively, $(C^{(i)}(r))_{r \in \mathbb{R}_+}$) is the spherical sector (resp., the decreasing family of cones) based on the hypercubic facet $A_b^{(i)}$.

Combining the two hypotheses, we see that if $((A_b^{(1)}, \dots, A_b^{(m)}), b_-)$ is a negative approximation of the spherical sector B, then one has

$$B = S \times [b, +\infty) \subset \bigcup_{i=1}^{m} \left(A_b^{(i)} \times [1, +\infty)\right) \subset \bigcup_{i=1}^{m} C_-(B^{(i)}) \subset \bigcup_{i=1}^{m} C^{(i)}(b_-).$$

On the other hand, given a negative approximation, we can assume without loss of generality that the interiors of the hypercubic facets $A_b^{(i)}$ are all disjoint. Similarly, a positive approximation of a general spherical sector $B = S \times [b, \infty)$ with S measurable by hypercubic facets is given by a finite union T_1 of hypercubic facets $A_b^{(1)}, \dots, A_b^{(l)}$ of $\mathbb{S}^{d-1}(b)$, such that

$$T_1 = \bigcup_{i=1}^{l} A_b^{(i)} \subset S_b.$$

Then,

$$\bigcup_{i=1}^{l} C_+(B^{(i)}) = \bigcup_{i=1}^{l} C^{(i)}(b) \subset \bigcup_{i=1}^{l} \left(A_b^{(i)} \times [1, +\infty)\right) \subset S \times [b, +\infty) = B.$$

For instance, Figure 3 represents a positive approximation of a spherical sector B.

Proposition 3.4 (Existence of negative and positive approximations). *Fix $\eta > 0$. For every part $S \subset \mathbb{S}^{d-1}$ that is measurable by hypercubic facets, and for every $\varepsilon > 0$ small enough (this condition depending on S and η), there exists a positive approximation T_1 and a negative approximation (T_2, b_-) of the spherical sector $B = S \times [b, +\infty)$, with:*

(1) $\mu_{\mathrm{surface}}(T_2 \setminus T_1) \leq \eta$;

(2) the interiors of the hypercubic facets in T_1 (respectively, in T_2) are disjoint;

(3) the number of hypercubic facets in T_1 and T_2 is a $O(\frac{1}{\varepsilon^{d-1}})$;

(4) these hypercubic facets are all of diameter smaller than $O(\varepsilon)$, and $b - b_- = O(\varepsilon^2)$, where the constants in the $O(\cdot)$'s only depend on b and d.

Proof. Fix $\eta > 0$. In the definition of Jordan mesurability of a part of the hypercube ∂E, we can take approximations of sets by unions of cells that have disjoint interiors, and are all taken from a regular grid traced on ∂E. However, the notion of measurability by hypercubic facets (Definition 3.2) derives from the notion of Jordan mesurability by the projection $\psi : \partial E \to \mathbb{S}^{d-1}(b)$. Consequently, if $S \subset \mathbb{S}^{d-1}(b)$ is measurable by hypercubic facets, then for every $\eta > 0$, there exists a pair (T_1, T_2) of unions of hypercubic facets such that $T_1 \subset S_b \subset T_2$ and $\mu_{\mathrm{surface}}(T_2 \setminus T_1) \leq \eta$, and moreover, one can assume that the hypercubic facets in T_1 and T_2 have disjoint interiors, and that they all come from the projection on $\mathbb{S}^{d-1}(b)$ of a regular grid on ∂E. Now, given a regular grid on ∂E, one can split every cell of the grid in 2^{d-1} subcells with half the size of the original cell. Thus, given $\varepsilon > 0$ smaller than the size of the cells of the original grid, one can assume without loss of generality that the hypercubic facets of T_1 and T_2 all come from a regular grid on ∂E of mesh smaller than ε. The number of cells in such a grid is $O(\frac{1}{\varepsilon^{d-1}})$, so T_1 and T_2 have less than $O(\frac{1}{\varepsilon^{d-1}})$ hypercubic facets. Finally, each hypercubic facet A_b of T_1 or T_2 has diameter smaller than $O(\varepsilon)$, and then, the pair of approximations $(C_-(B), C_+(B))$ of the spherical sector based on such an hypercubic facet has level $O(\varepsilon^2)$ by a previous remark, so in particular one can choose $b_- \geq b - O(\varepsilon^2)$. ☐

We can finally prove our large deviation results. We start with the case when $K = I_d$:

Proposition 3.5 (Large deviations in the isotropic case $K = I_d$)**.** *Let $(\mathbf{X}_n)_{n\in\mathbb{N}}$ be a sequence of random vectors in $\mathbb{R}^d$, that converges mod-Gaussian in the Laplace sense on the multi-strip $\mathcal{S}_{\mathbf{a},\mathbf{b}}$, with parameters $t_n I_d$ and limit ψ. We assume that $\mathbb{S}^{d-1}(b) \subset \prod_{i=1}^d (a^{(i)}, b^{(i)})$, and we consider a Borel subset $S \subset \mathbb{S}^{d-1}$ that is measurable by hypercubic facets, and of non-zero surface measure. Then, if B is the spherical sector $S \times [b, +\infty)$, we have*

$$\mathbb{P}[\mathbf{X}_n \in t_n B] = \left(\frac{t_n}{2\pi}\right)^{\frac{d}{2}} \frac{1}{t_n b}\, \mathrm{e}^{-\frac{t_n b^2}{2}} \left(\int_{S_b} \psi(\mathbf{s})\, \mu_{\text{surface}}(d\mathbf{s})\right) (1 + o(1)),$$

where $S_b = \{b\mathbf{s},\ \mathbf{s} \in S\}$.

Proof. Fix $\eta > 0$. For every cone C based on a domain D bounded by $M > 0$, Proposition 3.1 enables us to write

$$\begin{aligned}
&(2\pi)^{\frac{d}{2}}\, \mathrm{e}^{\frac{t_n \|\mathbf{h}\|^2}{2}}\, \mathbb{P}[\mathbf{X}_n \in t_n C] \\
&= \psi_n(\mathbf{h}) \left((t_n)^{\frac{d}{2}-1} \int_{\mathbf{d}\in D_0} \left(1 + \frac{\langle \nabla\psi(\mathbf{h}) \mid \mathbf{d}\rangle}{\psi(\mathbf{h})} + O\left(\frac{1}{t_n}\right)\right) \frac{\|\mathbf{h}\|\, \mathrm{e}^{-\frac{t_n \|\mathbf{d}\|^2}{2}}}{\|\mathbf{h}\|^2 + \|\mathbf{d}\|^2}\, d\mathbf{d} + o\left(\frac{1}{\sqrt{t_n}}\right)\right),
\end{aligned}$$

where $\mathbf{h}$ is the projection of the origin $\mathbf{0}$ on the affine hyperplane containing D, and where the $o(\frac{1}{\sqrt{t_n}})$ writes more precisely as

$$\frac{\delta_n(D)}{\sqrt{t_n}}$$

with $|\delta_n(D)| \leq \delta_n$, $(\delta_n)_{n\in\mathbb{N}}$ being a positive sequence that converges to 0, and that depends only on the choice of the bound M. In the following, we fix a positive sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ such that

$$\lim_{n\to\infty} \varepsilon_n \sqrt{t_n} = 0 \qquad ; \qquad \lim_{n\to\infty} \frac{\varepsilon_n \sqrt{t_n}}{(\delta_n)^{\frac{1}{d-1}}} = +\infty.$$

For n large enough, $\varepsilon = \varepsilon_n$ is small enough and one can choose corresponding positive and negative approximations T_1 and (T_2, b_-) of the spherical sector B, as in Proposition 3.4. We have obviously $\mathbb{P}[\mathbf{X}_n \in t_n T_1] \leq \mathbb{P}[\mathbf{X}_n \in t_n B] \leq \mathbb{P}[\mathbf{X}_n \in t_n T_2]$, and on the other hand, we can evaluate the probabilities related to T_1 and T_2 by using the estimates on cones. We write

$$T_1 = \bigcup_{i=1}^{l} A_b^{(i)} \qquad ; \qquad T_2 = \bigcup_{j=1}^{m} \widetilde{A}_b^{(j)},$$

where the $A_b^{(i)}$'s (resp., the $\widetilde{A}_b^{(j)}$'s) are hypercubic facets of $\mathbb{S}^{d-1}(b)$ with disjoint interiors. We denote $D^{(i)}$ the convex domain on which is based the cone $C^{(i)}(b)$ corresponding to the hypercubic facet $A_b^{(i)}$, and $\widetilde{D}^{(j)}$ the convex domain on which is based the cone $\widetilde{C}^{(j)}(b_-)$ corresponding to the hypercubic facet $\widetilde{A}_b^{(j)}$. We also denote $\mathbf{h}_i$ (resp., $\widetilde{\mathbf{h}}_j$) the orthogonal projection of $\mathbf{0}$ on the affine hyperplane that contains $D^{(i)}$ (respectively, $\widetilde{D}^{(j)}$). Notice that

$\|\mathbf{h}_i\| = b$, whereas $\|\widetilde{\mathbf{h}}_j\| = b_-$. Moreover, the vectors in the domains $D_0^{(i)} = D^{(i)} - \mathbf{h}_i$ and $\widetilde{D}_0^{(j)} = \widetilde{D}^{(j)} - \widetilde{\mathbf{h}}_j$ have norm bounded by $O(\varepsilon)$. By the discussion at the beginning of the proof,

$$\begin{aligned}
&(2\pi)^{\frac{d}{2}}\,\mathrm{e}^{\frac{t_n b^2}{2}}\,\mathbb{P}[\mathbf{X}_n \in t_n T_1] \\
&= \sum_{i=1}^{l} (2\pi)^{\frac{d}{2}}\,\mathrm{e}^{\frac{t_n b^2}{2}}\,\mathbb{P}[\mathbf{X}_n \in t_n C^{(i)}(b)] \\
&= \sum_{i=1}^{l} \psi_n(\mathbf{h}_i) \left((t_n)^{\frac{d}{2}-1} \int_{\mathbf{d}\in D_0^{(i)}} \left(1 + \frac{\langle \nabla\psi(\mathbf{h}_i) \mid \mathbf{d}\rangle}{\psi(\mathbf{h}_i)} + O\left(\frac{1}{t_n}\right)\right) \frac{b\,\mathrm{e}^{-\frac{t_n\|\mathbf{d}\|^2}{2}}}{b^2+\|\mathbf{d}\|^2}\,d\mathbf{d} + \frac{\delta_n(D^{(i)})}{\sqrt{t_n}} \right) \\
&= \sum_{i=1}^{l} \frac{\psi_n(\mathbf{h}_i)}{b} \left((t_n)^{\frac{d}{2}-1} \int_{\mathbf{d}\in D_0^{(i)}} \left(1 + O(\varepsilon) + O\left(\frac{1}{t_n}\right)\right) \frac{1+O(t_n\varepsilon^2)}{1+O(\varepsilon^2)}\,d\mathbf{d} \right) \\
&\quad + \sum_{i=1}^{l} \psi_n(\mathbf{h}_i)\,\frac{\delta_n(D^{(i)})}{\sqrt{t_n}}
\end{aligned}$$

where the $O(\cdot)$'s on the last line are uniform with respect to the index l. The second part on the last line is smaller than a constant times

$$\frac{1}{(\varepsilon_n)^{d-1}}\,\frac{\delta_n}{\sqrt{t_n}} = \frac{(t_n)^{\frac{d}{2}-1}}{\left(\frac{\varepsilon_n\sqrt{t_n}}{(\delta_n)^{\frac{1}{d-1}}}\right)^{d-1}} = o\left((t_n)^{\frac{d}{2}-1}\right)$$

by the hypotheses made on $(\varepsilon_n)_{n\in\mathbb{N}}$. On the other hand, the first part is

$$\begin{aligned}
&\sum_{i=1}^{l} \frac{\psi_n(\mathbf{h}_i)\,(t_n)^{\frac{d}{2}}}{t_n b} \left(\int_{\mathbf{d}\in D_0^{(i)}} (1+o(1))\,d\mathbf{d} \right) \\
&= \frac{(t_n)^{\frac{d}{2}}}{t_n b} \left(\int_{T_1} \psi_n(\mathbf{s})\,\mu_{\mathrm{surface}}(d\mathbf{s}) \right) + o\left((t_n)^{\frac{d}{2}-1}\right).
\end{aligned}$$

Similarly, for the upper approximation T_2, we get

$$(2\pi)^{\frac{d}{2}}\,\mathrm{e}^{\frac{t_n(b_-)^2}{2}}\,\mathbb{P}[\mathbf{X}_n \in t_n T_1] = \frac{(t_n)^{\frac{d}{2}}}{t_n b_-} \left(\int_{T_2} \psi_n(\mathbf{s})\,\mu_{\mathrm{surface}}(d\mathbf{s}) \right) + o\left((t_n)^{\frac{d}{2}-1}\right).$$

Since $b - b_- = O(\varepsilon^2) = o(\frac{1}{t_n})$, and since ψ_n converges locally uniformly towards ψ, we conclude that

$$\begin{aligned}
\int_{T_1} \psi(\mathbf{s})\,\mu_{\mathrm{surface}}(d\mathbf{s}) + o(1) &\leq \left(\frac{2\pi}{t_n}\right)^{\frac{d}{2}} t_n b\,\mathrm{e}^{\frac{t_n b^2}{2}}\,\mathbb{P}[\mathbf{X}_n \in t_n B] \\
&\leq \int_{T_2} \psi(\mathbf{s})\,\mu_{\mathrm{surface}}(d\mathbf{s}) + o(1).
\end{aligned}$$

As $T_1 \subset S_b \subset T_2$ and $\mu_{\mathrm{surface}}(T_2 \setminus T_1) \leq \eta$ can be taken as small as wanted, we obtain finally

$$\left(\frac{2\pi}{t_n}\right)^{\frac{d}{2}} t_n b\,\mathrm{e}^{\frac{t_n b^2}{2}}\,\mathbb{P}[\mathbf{X}_n \in t_n B] = \int_{S_b} \psi(\mathbf{s})\,\mu_{\mathrm{surface}}(d\mathbf{s}) + o(1),$$

which ends the proof of the estimate of large deviations. □

Theorem 3.6 (Large deviations)**.** *Let $(\mathbf{X}_n)_{n\in\mathbb{N}}$ be a sequence of random vectors that is mod-Gaussian convergent in the Laplace sense, with parameters t_nK and limit $\psi(\mathbf{z})$. We consider an ellipsoidal sector $B = S \times [b,+\infty)$, where S is a Borel subset of $K^{1/2}(\mathbb{S}^{d-1})$ that is the image by $K^{1/2}$ of a subset of $\mathbb{S}^{d-1}$ that is measurable by hypercubic facets, and with non-zero surface measure. Then,*

$$\mathbb{P}[\mathbf{X}_n \in t_nB] = \left(\frac{t_n}{2\pi}\right)^{\frac{d}{2}} \frac{1}{t_nb}\,\mathrm{e}^{-\frac{t_nb^2}{2}} \left(\int_{(K^{-1/2}(S))_b} \psi(K^{-1/2}\mathbf{s})\,\mu_{\mathrm{surface}}(d\mathbf{s})\right)(1+o(1)).$$

Proof. Consider $\mathbf{Y}_n = K^{-1/2}\mathbf{X}_n$; it is mod-Gaussian convergent in the Laplace sense, with parameters t_nI_d and limit $\psi(K^{-1/2}\mathbf{z})$. By Proposition 3.5, if

$$K^{-1/2}(B) = B' = S' \times [b,+\infty)$$

with $S' = K^{-1/2}(S) \subset \mathbb{S}^{d-1}$, then

$$\mathbb{P}[\mathbf{Y}_n \in t_nB'] = \left(\frac{t_n}{2\pi}\right)^{\frac{d}{2}} \frac{1}{t_nb}\,\mathrm{e}^{-\frac{t_nb^2}{2}} \left(\int_{(S')_b} \psi(K^{-1/2}\mathbf{s})\,\mu_{\mathrm{surface}}(d\mathbf{s})\right)(1+o(1)).$$

Finally, $\mathbb{P}[\mathbf{Y}_n \in t_nB'] = \mathbb{P}[K^{-1/2}\mathbf{X}_n \in t_nK^{-1/2}(B)] = \mathbb{P}[\mathbf{X}_n \in t_nB]$. □

Remark 3.7. The case $d = 1$ of the asymptotic formula of Theorem 3.6 allows one to recover Theorem 1.8 if one agrees that the surface measure on the zero-dimensional sphere $\mathbb{S}^0 = \{+1,-1\}$ is the counting measure.

4. Examples of multi-dimensional convergent sequences

In this section, we give applications of the theory developed in Sections 2 and 3.

4.1. **First examples.**

We start by looking at the two examples of multi-dimensional mod-Gaussian convergence proposed in the introductory Section 1.2.

Example 4.1 (Sums of i.i.d. random vectors). Let $\mathbf{A}$ be a random variable in $\mathbb{R}^d$ with entire Laplace transform $\mathbb{E}[\mathrm{e}^{\langle \mathbf{z}\,|\,\mathbf{A}\rangle}]$, and $(\mathbf{A}_n)_{n\in\mathbb{N}}$ be a sequence of independent copies of $\mathbf{A}$. We assume that $\mathbf{A}$ is centered and that $\mathrm{cov}(\mathbf{A}) = I_d$; up to a linear change of coordinates and a possible reduction of the dimension, these assumptions do not restrict the generality. The rescaled sum

$$\mathbf{X}_n = \frac{\mathbf{S}_n}{n^{1/3}} = \frac{1}{n^{1/3}} \sum_{k=1}^{n} \mathbf{A}_k$$

is mod-Gaussian convergent in the Laplace sense on $\mathbb{C}^d$, with parameters $n^{1/3}I_d$ and limit

$$\psi(\mathbf{z}) = \exp\left(\frac{1}{6}\sum_{i,j,k=1}^{d} \mathbb{E}[A^{(i)}A^{(j)}A^{(k)}]\,z^{(i)}z^{(j)}z^{(k)}\right).$$

By Theorem 3.6, for any spherical sector $B = S \times [b,+\infty)$ with S part of $\mathbb{S}^{d-1}$ measurable by hypercubic facets,

$$\begin{aligned}&\mathbb{P}[\mathbf{S}_n \in n^{2/3}B] \\ &\simeq \frac{n^{\frac{d-2}{6}}\,\mathrm{e}^{-\frac{n^{1/3}b^2}{2}}}{(2\pi)^{\frac{d}{2}}\,b} \left(\int_{S_b} \exp\left(\frac{1}{6}\sum_{i,j,k=1}^{d} \mathbb{E}[A^{(i)}A^{(j)}A^{(k)}]\,x^{(i)}x^{(j)}x^{(k)}\right)\mu_{\mathrm{surface}}(d\mathbf{x})\right).\end{aligned}$$

Hence, up to the scale $n^{2/3}$, the sum $\mathbf{S}_n$ of i.i.d. vectors is described by the Gaussian approximation, and at the scale $n^{2/3}$, this normal approximation is corrected by the exponential $\exp(\frac{1}{6}\sum_{i,j,k=1}^{d}\mathbb{E}[A^{(i)}A^{(j)}A^{(k)}]\,x^{(i)}x^{(j)}x^{(k)})$.

If $\mathbf{A}$ is symmetric in law ($\mathbf{A}$ and $-\mathbf{A}$ have the same law), then all the third moments $\mathbb{E}[A^{(i)}A^{(j)}A^{(k)}]$ vanish, and one has to look at another renormalisation of the sum $\mathbf{S}_n$. Hence, for sums of independent symmetric random variables, the new rescaled sum

$$\mathbf{Y}_n = \frac{\mathbf{S}_n}{n^{1/4}} = \frac{1}{n^{1/4}}\sum_{k=1}^{n}\mathbf{A}_k$$

is mod-Gaussian convergent in the Laplace sense, with parameters $n^{1/2}I_d$ and limit

$$\psi(\mathbf{z}) = \exp\left(\frac{1}{24}\sum_{i,j,k,l=1}^{d}\kappa(A^{(i)},A^{(j)},A^{(k)},A^{(l)})\,z^{(i)}z^{(j)}z^{(k)}z^{(l)}\right),$$

where

$$\begin{aligned}\kappa(A^{(i)},A^{(j)},A^{(k)},A^{(l)}) &= \mathbb{E}[A^{(i)}A^{(j)}A^{(k)}A^{(l)}] - \mathbb{E}[A^{(i)}A^{(j)}]\,\mathbb{E}[A^{(k)}A^{(l)}]\\ &\quad - \mathbb{E}[A^{(i)}A^{(k)}]\,\mathbb{E}[A^{(j)}A^{(l)}] - \mathbb{E}[A^{(i)}A^{(l)}]\,\mathbb{E}[A^{(j)}A^{(k)}].\end{aligned}$$

The quantities $\kappa(A^{(i)},A^{(j)},A^{(k)},A^{(l)})$ are the joint cumulants of order 4 of the coordinates of $\mathbf{A}$; we shall detail this theory in Section 4.2. In this setting, by Theorem 3.6, for any spherical sector $B = S\times[b,+\infty)$,

$$\begin{aligned}&\mathbb{P}[\mathbf{S}_n\in n^{3/4}B]\\ &\simeq \frac{n^{\frac{d-2}{4}}\mathrm{e}^{-\frac{n^{1/2}b^2}{2}}}{(2\pi)^{\frac{d}{2}}b}\left(\int_{S_b}\exp\left(\frac{1}{24}\sum_{i,j,k,l=1}^{d}\kappa(A^{(i)},A^{(j)},A^{(k)},A^{(l)})\,x^{(i)}x^{(j)}x^{(k)}x^{(l)}\right)\mu_{\mathrm{surface}}(d\mathbf{x})\right).\end{aligned}$$

A simple consequence of these multi-dimensional results is the loss of symmetry of the random walks on $\mathbb{Z}^d$ conditioned to be far away from the origin; this loss of symmetry has also been brought out in dimension 2 in [Ben19]. Thus, consider the simple 2-dimensional random walk $\mathbf{S}_n = \sum_{k=1}^{n}\mathbf{A}_k$, where $\mathbf{A}_k = (\pm 1, 0)$ or $(0,\pm 1)$ with probability $\frac{1}{4}$ for each direction. The non-zero fourth cumulants of $\mathbf{A}$ are

$$\kappa((\mathrm{Re}\,\mathbf{A})^{\otimes 4}) = \kappa((\mathrm{Im}\,\mathbf{A})^{\otimes 4}) = \kappa((\mathrm{Re}\,\mathbf{A})^{\otimes 2},(\mathrm{Im}\,\mathbf{A})^{\otimes 2}) = -\frac{1}{4},$$

so one has mod-Gaussian convergence of $n^{-1/4}\,\mathbf{S}_n$ with parameters $\frac{n^{1/2}}{2}\,I_2$ and limiting function

$$\psi(\mathbf{z}) = \exp\left(-\frac{(z^{(1)})^4+(z^{(2)})^4+6(z^{(1)}z^{(2)})^2}{96}\right).$$

Therefore, for every cylindric sector $C(r,\theta_1,\theta_2) = \{R\mathrm{e}^{\mathrm{i}\theta}\in\mathbb{C}\mid R\geq r,\ \theta\in(\theta_1,\theta_2)\}$, Theorem 3.6 gives the estimate:

$$\mathbb{P}\Big[\mathbf{S}_n\in n^{3/4}\,C(r,\theta_1,\theta_2)\Big] = \mathrm{e}^{-n^{1/2}r^2}\left(\int_{\theta_1}^{\theta_2}\psi(2r\mathrm{e}^{\mathrm{i}\theta})\,\frac{d\theta}{2\pi}\right)(1+o(1)).$$

This leads to the following limiting result: if $\mathbf{S}_n = R_n\,\mathrm{e}^{\mathrm{i}\theta_n}$ with $\theta_n\in[0,2\pi)$, then

$$\lim_{n\to\infty}\mathbb{P}\Big[\theta_n\in(\theta_1,\theta_2)\,\Big|\,R_n\geq rn^{3/4}\Big] = \frac{\int_{\theta_1}^{\theta_2}\psi(2r\mathrm{e}^{\mathrm{i}\theta})\,d\theta}{\int_0^{2\pi}\psi(2r\mathrm{e}^{\mathrm{i}\theta})\,d\theta} = \int_{\theta_1}^{\theta_2}F(r,\theta)\,d\theta$$

with $F(r,\theta) = \dfrac{\exp\left(-\frac{r^4\,(\sin 2\theta)^2}{6}\right)}{\int_0^{2\pi}\exp\left(-\frac{r^4\,(\sin 2\theta)^2}{6}\right)d\theta}$ drawn in Figure 6.

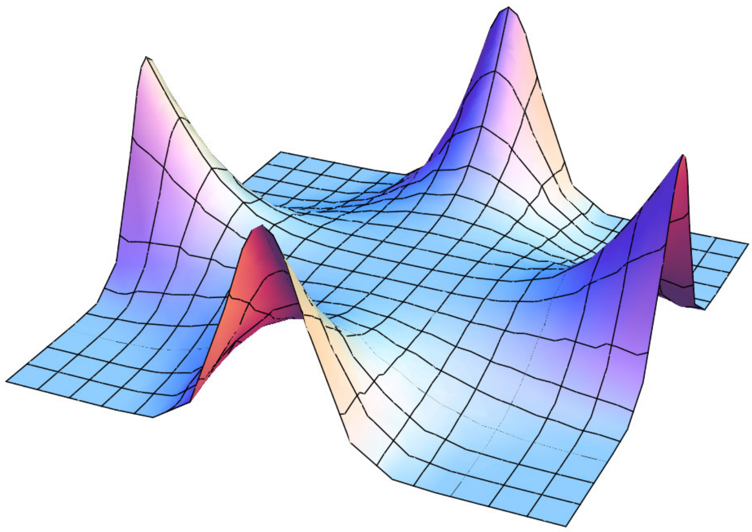

FIGURE 6. The function $F(r,\theta)$ measuring the loss of symmetry of the 2-dimensional random walk $\mathbf{S}_n$ conditioned to have large radius $\|\mathbf{S}_n\| \geq rn^{3/4}$ (using `Mathematica`®).

This function gets concentrated around the two axes of $\mathbb{R}^2$ when $r \to \infty$, whence a loss of symmetry in comparison to the behavior of the 2-dimensional Brownian motion (the scaling limit of the random walk). In dimension $d \geq 3$, one obtains the similar result

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{\mathbf{S}_n}{\|\mathbf{S}_n\|} \in S \,\middle|\, \|\mathbf{S}_n\| \geq rn^{3/4}\right] = K(r) \int_S \exp\left(-\frac{r^4}{12\,d} \sum_{1\leq i<j\leq d} (x^{(i)}x^{(j)})^2\right) \mu(d\mathbf{x})$$

for any set $S \subset \mathbb{S}^{d-1}$ that is measurable by hypercubic facets. The conditional probability is therefore concentrated around the axes of $\mathbb{R}^d$.

On the other hand, Theorem 2.17 shows that if $\mathbf{S}_n = \sum_{k=1}^n \mathbf{A}_k$ and $\mathbf{A}$ is only assumed to be centered, non-degenerate ($\mathrm{cov}(\mathbf{A}) \in \mathrm{S}_+(d,\mathbb{R})$) and with a third moment, then

$$d_{\mathrm{convex}}\left(\frac{\mathbf{S}_n}{n^{1/2}}, \mathcal{N}_{\mathbb{R}^d}(\mathbf{0},\mathrm{cov}(\mathbf{A}))\right) = O\left(\frac{1}{n^{1/6}}\right).$$

This is not optimal, and we shall see in the next section that with the same hypotheses, one can prove a bound $O(\frac{1}{n^{1/2}})$ on the convex distance.

Example 4.2 (Characteristic polynomials of random unitary matrices). Let U_n be a random unitary matrix in $\mathrm{U}(n)$ taken according to the Haar measure. We saw in Section 1.2 that the sequence $(\mathbf{X}_n = \log\det(I_n - U_n))_{n\in\mathbb{N}}$ is mod-Gaussian convergent on the strip $\mathcal{S}_{(-1,+\infty)} \times \mathbb{C}$, with parameters $\frac{\log n}{2} I_2$ and limit

$$\psi(\mathbf{z}) = \frac{G\left(1+\frac{z^{(1)}+\mathrm{i}z^{(2)}}{2}\right)\,G\left(1+\frac{z^{(1)}-\mathrm{i}z^{(2)}}{2}\right)}{G(1+z^{(1)})}.$$

Therefore, for any $r<1$ and any circular sector $C(r,\theta_1,\theta_2)$,

$$\mathbb{P}\left[\mathbf{X}_n \in \frac{\log n}{2}\, C(r,\theta_1,\theta_2)\right] = n^{-\frac{r^2}{4}} \left(\int_{\theta=\theta_1}^{\theta_2} \frac{G(1+\frac{r}{2}\mathrm{e}^{\mathrm{i}\theta})\;G(1+\frac{r}{2}\mathrm{e}^{-\mathrm{i}\theta})}{G(1+r\cos\theta)}\,\frac{d\theta}{2\pi}\right)(1+o(1)).$$

The function $H(r,\theta) = \frac{G(1+\frac{r}{2}\mathrm{e}^{\mathrm{i}\theta})\,G(1+\frac{r}{2}\mathrm{e}^{-\mathrm{i}\theta})}{G(1+r\cos\theta)}$ takes higher values for θ close to π, hence a loss of symmetry of $\mathbf{X}_n$ at scale $\log n$.

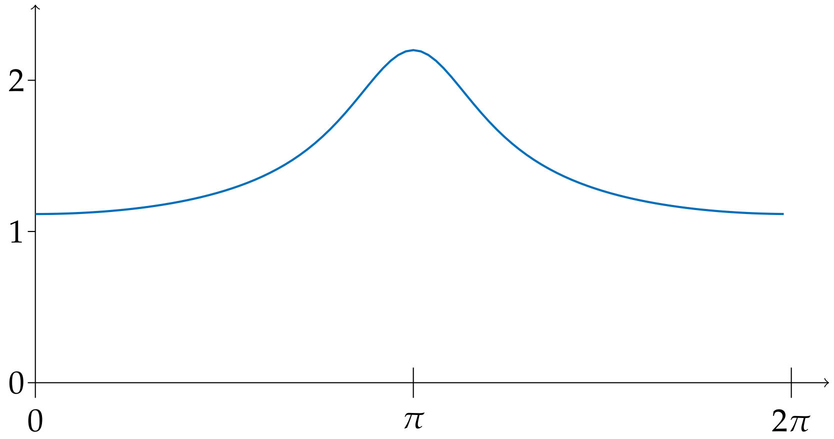

FIGURE 7. The function $H(r,\theta)$ when $r = 0.7$.

On the other hand, the gradient of ψ at $\mathbf{0}$ is equal to $\mathbf{0}$, so by Theorem 2.20, we have

$$d_{\mathrm{convex}}\left(\sqrt{\frac{2}{\log n}}\,\mathbf{X}_n\,,\,\mathcal{N}_{\mathbb{R}^2}(\mathbf{0},I_2)\right) = o\left(\frac{1}{\sqrt{\log n}}\right).$$

A more precise analysis shows that this convex distance is actually a $O((\log n)^{-3/2})$; see the arguments of the next paragraph, and [MN22, Section 3] for the analogous result in dimension 1 with the real part of the log-characteristic polynomial.

4.2. **The method of cumulants.** In dimension $d = 1$, assuming that the Laplace transform $\mathbb{E}[\mathrm{e}^{zX_n}]$ is convergent on a disk around $z = 0$, it is possible to reformulate the mod-Gaussian convergence of $(X_n)_{n\in\mathbb{N}}$ in terms of the behavior of the coefficients of the power series

$$\log \mathbb{E}[\mathrm{e}^{zX_n}] = \sum_{r=1}^{\infty} \frac{\kappa^{(r)}(X_n)}{r!}\,z^r,$$

called the *cumulants* of $(X_n)_{n\in\mathbb{N}}$. We refer to [FMN16, Chapters 5 and 9] and [FMN19, Section 4] for developments around this notion. Now, there is a notion of *joint cumulants* (*cf.* [LS59]) that allows one to generalise this method of cumulants, and to give a numerical criterion of mod-Gaussian convergence in arbitrary dimension $d \geq 1$.

Definition 4.3 (Joint cumulant). *Given random variables $Y_1,\ldots,Y_r$ with convergent generating series $\mathbb{E}[\mathrm{e}^{z_1Y_1+\cdots+z_rY_r}]$, their joint cumulant is*

$$\kappa(Y_1,\ldots,Y_r) = \left.\frac{\partial^r}{\partial z_1\cdots\partial z_r}\right|_{z_1=\cdots=z_r=0} \log\left(\mathbb{E}[\mathrm{e}^{z_1Y_1+\cdots+z_rY_r}]\right).$$

The joint cumulant is an homogeneous polynomial of degree r in the joint moments of the variables $Y_1,\ldots,Y_r$. More precisely, if $\mathfrak{Q}_r$ is the set of set partitions of $[\![1,r]\!]$, then

$$\kappa(Y_1,\ldots,Y_r) = \sum_{\pi\in\mathfrak{Q}_r} \mu(\pi)\left(\prod_{j=1}^{\ell(\pi)} \mathbb{E}\left[\prod_{i\in\pi_j} Y_j\right]\right),$$

where $\mu(\pi) = (-1)^{\ell-1}\,(\ell-1)!$ if $\pi = \pi_1 \sqcup \pi_2 \sqcup \cdots \sqcup \pi_\ell$ is a set partition with ℓ parts. For instance, $\kappa(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$ is the covariance, and

$$\kappa(X,Y,Z) = \mathbb{E}[XYZ] - \mathbb{E}[XY]\,\mathbb{E}[Z] - \mathbb{E}[XZ]\,\mathbb{E}[Y] - \mathbb{E}[YZ]\,\mathbb{E}[Z] + 2\,\mathbb{E}[X]\,\mathbb{E}[Y]\,\mathbb{E}[Z].$$

Let us recall the main properties of the joint cumulants, which follow readily from their definition:

Proposition 4.4 (Properties of joint cumulants). *Let $Y_1, \ldots, Y_r$ be random variables with a convergent Laplace transform.*

(1) The joint cumulants are multilinear and invariant by permutation of the random variables.

(2) If the variables $Y_1, \ldots, Y_r$ can be separated into two blocks of independent random variables, then $\kappa(Y_1, \ldots, Y_r) = 0$.

(3) For any random variable Y, the classical r-th cumulant $\kappa^{(r)}(Y)$ is equal to the joint cumulant $\kappa(Y, \ldots, Y)$ (with r occurrences of Y).

(4) For any random vector $\mathbf{X}$ in $\mathbb{R}^d$, assuming that the Laplace transform $\mathbb{E}[\mathrm{e}^{\langle \mathbf{z} \,|\, \mathbf{X}\rangle}]$ is convergent, one has the expansion

$$\begin{aligned}\log \mathbb{E}[\mathrm{e}^{\langle \mathbf{z} \,|\, \mathbf{X}\rangle}] &= \sum_{r\geq 1} \frac{1}{r!}\, \kappa^{(r)}(z^{(1)}X^{(1)} + \cdots + z^{(d)}X^{(d)}) \\ &= \sum_{r\geq 1} \sum_{(i_1,\ldots,i_r)\in [\![1,d]\!]^r} \frac{\kappa(X^{(i_1)}, \ldots, X^{(i_r)})}{r!}\, z^{(i_1)}z^{(i_2)}\cdots z^{(i_r)}.\end{aligned}$$

Let $(\mathbf{S}_n)_{n\in\mathbb{N}}$ be a sequence of random vectors in $\mathbb{R}^d$, and $K \in \mathrm{S}_+(d,\mathbb{R})$. The following definition is a multi-dimensional analogue of [FMN19, Definition 28]:

Definition 4.5 (Method of cumulants). *One says that $(\mathbf{S}_n)_{n\in\mathbb{N}}$ satisfies the hypotheses of the multi-dimensional method of cumulants with covariance matrix K, index $v \geq 3$ and positive parameters (D_n, N_n, A) if, for any choice of coordinates:*

(MC1) The random vectors $\mathbf{S}_n$ are centered: $\mathbb{E}[\mathbf{S}_n] = \mathbf{0}$.

(MC2) The covariance matrix of $\mathbf{S}_n$ is given by

$$\kappa(S_n^{(i)}, S_n^{(j)}) = N_n D_n\, K_{ij}\, \left(1 + o\left(\left(\frac{D_n}{N_n}\right)^{1-\frac{2}{v}}\right)\right)$$

with $\lim_{n\to\infty} \frac{D_n}{N_n} = 0$.

(MC3) For any $r \geq 3$,

$$\left|\kappa(S_n^{(i_1)}, S_n^{(i_2)}, \ldots, S_n^{(i_r)})\right| \leq N_n\, (2D_n)^{r-1}\, A^r\, r^{r-2}.$$

(MC4) For any $r \in [\![3, v-1]\!]$, $\kappa(S_n^{(i_1)}, S_n^{(i_2)}, \ldots, S_n^{(i_r)}) = 0$.

(MC5) There exist limits

$$L_{i_1,i_2,\ldots,i_v} = \lim_{n\to\infty} \frac{\kappa(S_n^{(i_1)}, S_n^{(i_2)}, \ldots, S_n^{(i_v)})}{N_n\, (D_n)^{v-1}}.$$

Remark 4.6. The one-dimensional version of Definition 4.5 was given in [FMN16, Section 5.1] and [FMN19, Section 4.1]. The main difference is that we do not ask the covariance matrix of $\frac{\mathbf{S}_n}{N_n D_n}$ to be exactly equal to K. Indeed, though the coefficients of $\mathrm{cov}(\mathbf{S}_n)$ will have the same order of magnitude $N_n D_n$ in our examples, in general we shall not be able to write $\mathrm{cov}(\mathbf{S}_n) = N_n D_n\, K$ with K constant matrix.

In the last two paragraphs 4.3 and 4.4, we shall see that many random models yield random vectors that satisfy the hypotheses of Definition 4.5, and in particular the bound on cumulants (MC3). The purpose of this section is to give the theoretical consequences of the multi-dimensional method of cumulants. We start with the mod-Gaussian convergence and the implied large deviation results:

Theorem 4.7 (Method of cumulants and mod-Gaussian convergence)**.** *Let $(\mathbf{S}_n)_{n\in\mathbb{N}}$ be a sequence of random vectors that satisfies the hypotheses of the multi-dimensional method of cumulants. We set*

$$\mathbf{X}_n = \frac{1}{(N_n)^{1/v}\,(D_n)^{1-1/v}}\,\mathbf{S}_n \qquad ; \qquad \mathbf{Y}_n = \frac{1}{(N_nD_n)^{1/2}}\,\mathbf{S}_n.$$

(1) The sequence $(\mathbf{X}_n)_{n\in\mathbb{N}}$ is mod-Gaussian convergent in the Laplace sense, with parameters $(\frac{N_n}{D_n})^{1-\frac{2}{v}}K$ and limit

$$\psi(\mathbf{z}) = \exp\left(\frac{1}{v!}\sum_{i_1,\ldots,i_v=1}^{d} L_{i_1,\ldots,i_v}\, z^{(i_1)}z^{(i_2)}\cdots z^{(i_v)}\right).$$

(2) Therefore, we have the convergence in law $\mathbf{Y}_n \rightharpoonup \mathcal{N}_{\mathbb{R}^d}(\mathbf{0},K)$.

(3) Consider an ellipsoidal sector $B = S\times[b,+\infty)$, where S is a subset of $K^{1/2}(\mathbb{S}^{d-1})$ that is measurable by hypercubic facets, and with non-zero surface measure. Then,

$$\mathbb{P}[\mathbf{Y}_n \in u_nB] = \frac{(u_n)^{d-2}}{(2\pi)^{\frac{d}{2}}\,b}\,\mathrm{e}^{-\frac{(u_nb)^2}{2}}\left(\int_{(K^{-1/2}(S))_b}\psi(K^{-1/2}\mathbf{s})\,\mu_{\mathrm{surface}}(d\mathbf{s})\right)(1+o(1))$$

with $u_n = \sqrt{t_n} = \left(\frac{N_n}{D_n}\right)^{\frac{1}{2}-\frac{1}{v}}$.

Proof. The mod-Gaussian convergence follows readily from the hypotheses in Definition 4.5, and from the fourth item of Proposition 4.4, which relates the joint cumulants of the coordinates of $\mathbf{S}_n$ to the Taylor expansion of the log-Laplace transform of $\mathbf{S}_n$. Indeed,

$$\begin{aligned}\log\mathbb{E}[\mathrm{e}^{\langle\mathbf{z}\,|\,\mathbf{X}_n\rangle}] &= \sum_{r\geq 1}\sum_{i_1,\ldots,i_r=1}^{d}\frac{\kappa(S_n^{(i_1)},\ldots,S_n^{(i_r)})}{r!\,N_n(D_n)^{r-1}}\left(\frac{D_n}{N_n}\right)^{\frac{r}{v}-1} z^{(i_1)}z^{(i_2)}\cdots z^{(i_r)}\\ &= \frac{t_n}{2}\sum_{i,j=1}^{d}\frac{\mathrm{cov}(S_n^{(i)},S_n^{(j)})}{N_nD_n}\,z^{(i)}z^{(j)} + \frac{1}{v!}\sum_{i_1,\ldots,i_v=1}^{d}\frac{\kappa(S_n^{(i_1)},\ldots,S_n^{(i_v)})}{N_n(D_n)^{v-1}}\,z^{(i_1)}\cdots z^{(i_v)}\\ &\quad + \text{remainder}\end{aligned}$$

with a remainder smaller than

$$\begin{aligned}&\sum_{r\geq v+1}\sum_{i_1,\ldots,i_r=1}^{d}\frac{|\kappa(S_n^{(i_1)},\ldots,S_n^{(i_r)})|}{r!\,N_n(D_n)^{r-1}}\left(\frac{D_n}{N_n}\right)^{\frac{r}{v}-1}|z^{(i_1)}z^{(i_2)}\cdots z^{(i_r)}|\\ &\leq \frac{N_n}{D_n}\sum_{r\geq v+1}\frac{2^{r-1}\,A^r\,r^{r-2}}{r!}\left(\left(\frac{D_n}{N_n}\right)^{\frac{1}{v}}\|\mathbf{z}\|_1\right)^r.\end{aligned}$$

The power series is convergent and a $O((D_n/N_n)^{1+\frac{1}{v}}(\|\mathbf{z}\|_1)^{v+1})$; hence, the remainder goes to 0. Now, the term with covariances is equivalent to $t_n\,\frac{\mathbf{z}^tK\mathbf{z}}{2}$ by Hypothesis (MC2), and the term of order v is equivalent to

$$\frac{1}{v!}\sum_{i_1,\ldots,i_v=1}^{d} L_{i_1,\ldots,i_v}\,z^{(i_1)}\cdots z^{(i_v)}$$

by Hypothesis (MC5). This ends the proof of the mod-Gaussian convergence, and the two other points are then immediate consequences of Proposition 1.5 and Theorem 3.6. ∎

We now focus on the speed of convergence of $\mathbf{Y}_n$ to $\mathcal{N}_{\mathbb{R}^d}(\mathbf{0},K)$. By Theorems 2.17 and 2.20, under the assumptions (MC1)-(MC5),

$$d_{\mathrm{convex}}\left(\mathbf{Y}_n\,,\,\mathcal{N}_{\mathbb{R}^d}(\mathbf{0},K)\right)=o\left(\left(\frac{D_n}{N_n}\right)^{\frac{1}{2}-\frac{1}{v}}\right),$$

because $\nabla\psi(\mathbf{0})=\mathbf{0}$. Under a slightly stronger hypothesis, one can show a better bound:

Theorem 4.8 (Method of cumulants and speed of convergence)**.** *Let $(\mathbf{S}_n)_{n\in\mathbb{N}}$ be a sequence of random vectors that satisfies (MC1),*

(MC2′) The covariance matrix of $\mathbf{S}_n$ is given by

$$\kappa(S_n^{(i)},S_n^{(j)})=N_nD_n\,K_{ij}\,\left(1+O\left(\left(\frac{D_n}{N_n}\right)^{\frac{1}{2}}\right)\right).$$

and (MC3). There exists a constant $C=C(d,K,A,B)$ that depends only on d, K, the constant A of (MC3) and the constant B in the $O(\cdot)$ of (MC2′), such that for n large enough,

$$d_{\mathrm{convex}}(\mathbf{Y}_n\,,\,\mathcal{N}_{\mathbb{R}^d}(\mathbf{0},K))\leq C\sqrt{\frac{D_n}{N_n}}.$$

The strategy of proof of Theorem 4.8 is the following:

- We first deal with the fact that $K_n=\mathrm{cov}(\mathbf{S}_n)/(N_nD_n)$ is not exactly equal to K, by computing the convex distance between two Gaussian distributions (see Lemma 4.9 below).
- Then, we prove a bound on $\Delta_\varepsilon(\widehat{\mu}_n,\mathcal{N}_{\mathbb{R}^d}(\mathbf{0},K_n))$ with $\varepsilon=O(\sqrt{D_n/N_n})$ (Lemma 4.10), by using the bounds on cumulants and the fast decay of $\mathrm{e}^{-\zeta^tK_n\zeta/2}$.

Lemma 4.9. *Let K_1 and K_2 be two positive-definite symmetric matrices; $\nu_1=\mathcal{N}_{\mathbb{R}^d}(\mathbf{0},K_1)$ and $\nu_2=\mathcal{N}_{\mathbb{R}^d}(\mathbf{0},K_2)$. If K_1 is fixed, then for any K_2 such that K_2-K_1 is sufficiently small, one has*

$$\Delta_\varepsilon(\widehat{\nu}_1,\widehat{\nu}_2)\leq 8\,\sqrt{(2\pi)^d(d+1)!}\,\frac{(\max(1,d\rho(K_1^{-1})))^{\frac{d+3}{2}}}{\sqrt{\det K_1}}\,\rho(K_2-K_1)$$

for any $\varepsilon>0$.

Proof. Since $(K_1)^{-1/2}K_2(K_1)^{-1/2}$ is a positive-definite symmetric matrix, there exists positive eigenvalues $\lambda^{(1)},\ldots,\lambda^{(d)}$ and an orthogonal matrix $R\in\mathrm{SO}(\mathbb{R}^d)$, such that

$$R^t(K_1)^{-1/2}K_2(K_1)^{-1/2}R=D^2=\mathrm{diag}(\lambda^{(1)},\ldots,\lambda^{(d)}).$$

Set $M=(K_1)^{-1/2}R$. If $\boldsymbol{\beta}$ is a multi-index of total weight $|\boldsymbol{\beta}|\leq d+1$, then

$$\begin{aligned}
&\int_{D_\varepsilon^d}\left|\frac{\partial^{|\boldsymbol{\beta}|}(\widehat{\nu}_1-\widehat{\nu}_2)(\boldsymbol{\zeta})}{\partial\boldsymbol{\zeta}^{\boldsymbol{\beta}}}\right|\,d\boldsymbol{\zeta}\\
&\leq\int_{\mathbb{R}^d}\left|\prod_{i=1}^d\left(\sum_{j=1}^d M_{ij}\frac{\partial}{\partial\xi^{(j)}}\right)^{\beta^{(i)}}\left(\mathrm{e}^{-\frac{\xi^tK_1\xi}{2}}-\mathrm{e}^{-\frac{\xi^tK_2\xi}{2}}\right)\right|\,d\boldsymbol{\zeta}\\
&\leq\frac{1}{\sqrt{\det K_1}}\left(\max_{i\in[\![1,d]\!]}\sum_{j=1}^d|M_{ij}|\right)^{|\boldsymbol{\beta}|}\max_{|\boldsymbol{\alpha}|=|\boldsymbol{\beta}|}\int_{\mathbb{R}^d}\left|\frac{\partial^{|\boldsymbol{\alpha}|}}{\partial\boldsymbol{\xi}^{\boldsymbol{\alpha}}}\left(\mathrm{e}^{-\frac{\|\xi\|^2}{2}}-\mathrm{e}^{-\frac{\|D\xi\|^2}{2}}\right)\right|\,d\boldsymbol{\xi}.
\end{aligned}$$

We evaluate separately each term. Since R is an orthogonal matrix and $(K_1)^{-1/2}$ is a symmetric matrix,

$$\max_{i\in[\![1,d]\!]}\sum_{j=1}^{d}|M_{ij}| = \max_{\|\mathbf{v}\|_\infty\leq 1}\|M\mathbf{v}\|_\infty$$
$$\leq \max_{\|\mathbf{v}\|_2\leq\sqrt{d}}\|M\mathbf{v}\|_2 = \max_{\|\mathbf{v}\|_2\leq\sqrt{d}}\|(K_1)^{-1/2}\,\mathbf{v}\|_2 \leq \sqrt{d\,\rho(K_1^{-1})}.$$

On the other hand,

$$I_{\boldsymbol{\alpha}} = \int_{\mathbb{R}^d}\left|\frac{\partial^{|\boldsymbol{\alpha}|}}{\partial\boldsymbol{\xi}^{\boldsymbol{\alpha}}}\left(\mathrm{e}^{-\frac{\|\boldsymbol{\xi}\|^2}{2}}-\mathrm{e}^{-\frac{\|D\boldsymbol{\xi}\|^2}{2}}\right)\right|d\boldsymbol{\xi}$$
$$= \int_{\mathbb{R}^d}\left|\left(\prod_{i=1}^{d}H_{\alpha^{(i)}}(\xi^{(i)})\,\mathrm{e}^{-\frac{(\xi^{(i)})^2}{2}}\right)-\left(\prod_{i=1}^{d}(\lambda^{(i)})^{\frac{\alpha^{(i)}}{2}}\,H_{\alpha^{(i)}}((\lambda^{(i)})^{\frac12}\xi^{(i)})\,\mathrm{e}^{-\frac{\lambda^{(i)}(\xi^{(i)})^2}{2}}\right)\right|d\boldsymbol{\xi}.$$

The multi-index $\boldsymbol{\alpha}=(\alpha^{(1)},\ldots,\alpha^{(d)})$ being fixed, we can assume without loss of generality that the eigenvalues $(\lambda^{(i)})_{i\in[\![1,d]\!]}$ are ordered as follows:

(0) For $i\in[\![1,d_0]\!]$, $\alpha^{(i)}=0$. Moreover, $\lambda^{(1)}\leq\lambda^{(2)}\leq\cdots\leq\lambda^{(d_0)}$, and we denote $e_0\in[\![0,d_0]\!]$ the largest index such that $\lambda^{(e_0)}\leq 1$.

(1) For $i\in[\![d_0+1,d_1]\!]$, $\alpha^{(i)}=1$.

(2) For $i\in[\![d_1+1,d]\!]$, $\alpha^{(i)}\geq 2$. Moreover, $\lambda^{(d_1+1)}\leq\lambda^{(d_1+2)}\leq\cdots\leq\lambda^{(d)}$, and we denote $e_2\in[\![d_1,d]\!]$ the largest index such that $\lambda^{(e_2)}\leq 1$.

In the following, we shall use several times the inequality

$$\int_{\mathbb{R}}|H_\alpha(\xi)\,H_\beta(\xi)|\,\mathrm{e}^{-\frac{\xi^2}{2}}\,d\xi\leq\sqrt{2\pi\,\alpha!\,\beta!},$$

and evaluate the terms of the sum in the right-hand side of

$$I_{\boldsymbol{\alpha}}\leq\sum_{j=1}^{d}\int_{\mathbb{R}^d}\left|\prod_{i=1}^{j-1}(\lambda^{(i)})^{\frac{\alpha^{(i)}}{2}}\,H_{\alpha^{(i)}}((\lambda^{(i)})^{\frac12}\xi^{(i)})\,\mathrm{e}^{-\frac{\lambda^{(i)}(\xi^{(i)})^2}{2}}\prod_{i=j+1}^{d}H_{\alpha^{(i)}}(\xi^{(i)})\,\mathrm{e}^{-\frac{(\xi^{(i)})^2}{2}}\right|$$
$$\times\,\Delta(\xi^{(j)},\lambda^{(j)},\alpha^{(j)})\,d\boldsymbol{\xi}$$
$$\leq\sum_{j=1}^{d}\left(\prod_{i\neq j}\sqrt{2\pi(\alpha^{(i)})!}\right)\left(\prod_{i<j}(\lambda^{(i)})^{\frac{\alpha^{(i)}-1}{2}}\right)\int_{\mathbb{R}}\Delta(\xi^{(j)},\lambda^{(j)},\alpha^{(j)})\,d\xi^{(j)},$$

where $\Delta(\xi^{(j)},\lambda^{(j)},\alpha^{(j)})=\left|H_{\alpha^{(j)}}(\xi^{(j)})\,\mathrm{e}^{-\frac{(\xi^{(j)})^2}{2}}-(\lambda^{(j)})^{\frac{\alpha^{(j)}}{2}}\,H_{\alpha^{(j)}}((\lambda^{(j)})^{\frac12}\xi^{(j)})\,\mathrm{e}^{-\frac{\lambda^{(j)}(\xi^{(j)})^2}{2}}\right|$.

We abbreviate $(\lambda^{(j)})^{\frac{\alpha^{(j)}-1}{2}}=\rho^{(j)}$.

(0) Terms with $j\in[\![1,d_0]\!]$. We then have $\int_{\mathbb{R}}\Delta(\xi^{(j)},\lambda^{(j)},0)\,d\xi^{(j)}=\sqrt{2\pi}\,|\rho^{(j)}-1|$, so, if $I_{\boldsymbol{\alpha},0}$ is the part of the sum bounding I_α with indices $j\in[\![1,d_0]\!]$, then

$$I_{\boldsymbol{\alpha},0}=\left(\prod_{i=1}^{d}\sqrt{2\pi(\alpha^{(i)})!}\right)\left(\sum_{j=1}^{e_0}\left(\rho^{(j)}-1\right)\prod_{i<j}\rho^{(i)}+\sum_{j=e_0+1}^{d_0}\left(1-\rho^{(j)}\right)\prod_{i<j}\rho^{(i)}\right)$$
$$=\sqrt{(2\pi)^d\,\boldsymbol{\alpha}!}\,\left(2\,\rho^{(1)}\rho^{(2)}\cdots\rho^{(e_0)}-\rho^{(1)}\rho^{(2)}\cdots\rho^{(d_0)}-1\right).$$

(1) Terms with $j \in [\![d_0+1, d_1]\!]$. Notice that $\prod_{i<j} \rho^{(i)} = \rho^{(1)}\rho^{(2)}\cdots\rho^{(d_0)}$ for all these indices. On the other hand,

$$\int_{\mathbb{R}} \Delta(\xi^{(j)}, \lambda^{(j)}, 1)\, d\xi^{(j)} = 4\,\mathrm{e}^{-\frac{\log \lambda^{(j)}}{\lambda^{(j)}-1}} \left|(\lambda^{(j)})^{-1} - 1\right| \leq 4\,|(\lambda^{(j)})^{-1} - 1|.$$

Therefore, if $I_{\boldsymbol{\alpha},1}$ is the part of the sum bounding $I_{\boldsymbol{\alpha}}$ with indices $j \in [\![d_0+1, d_1]\!]$, then

$$I_{\boldsymbol{\alpha},1} \leq \sqrt{(2\pi)^d\,\boldsymbol{\alpha}!}\left(\frac{4}{\sqrt{2\pi}}\,\rho^{(1)}\rho^{(2)}\cdots\rho^{(d_0)}\right)\sum_{j=d_0+1}^{d_1}\left|(\lambda^{(j)})^{-1} - 1\right|.$$

(2) Terms with $j \in [\![d_1+1, d]\!]$. Denote $f_{\xi,\alpha}(\lambda) = \lambda^{\frac{\alpha}{2}} H_\alpha(\lambda^{\frac{1}{2}}\xi)\,\mathrm{e}^{-\frac{\lambda\xi^2}{2}}$; by using the relations $H_n'(x) = nH_{n-1}(x) = x\,H_n(x) - H_{n+1}(x)$, one computes its derivative

$$\begin{aligned} f_{\xi,\alpha}'(\lambda) &= \frac{\lambda^{\frac{\alpha-2}{2}}}{2}\left((\alpha - x^2)\,H_\alpha(x) + \alpha\,x\,H_{\alpha-1}(x)\right)\mathrm{e}^{-\frac{x^2}{2}} \\ &= \frac{\lambda^{\frac{\alpha-2}{2}}}{2}\left((\alpha-1)\,H_\alpha(x) - H_2(x)\,H_\alpha(x) + \alpha\,H_1(x)\,H_{\alpha-1}(x)\right)\mathrm{e}^{-\frac{x^2}{2}}, \end{aligned}$$

with $x = \lambda^{\frac{1}{2}}\xi$. As a consequence, assuming for instance $\lambda^{(j)} \geq 1$, one obtains

$$\begin{aligned} &\int_{\mathbb{R}} |\Delta(\xi^{(j)}, \lambda^{(j)}, \alpha^{(j)})|\,d\xi^{(j)} \\ &\leq \frac{1}{2}\int_1^{\lambda^{(j)}} \lambda^{\frac{\alpha^{(j)}-3}{2}} \int_{\mathbb{R}} \begin{pmatrix} (\alpha^{(j)}-1)\,|H_{\alpha^{(j)}}(x)| + |H_2(x)\,H_{\alpha^{(j)}}(x)| \\ +\alpha^{(j)}\,|H_1(x)\,H_{\alpha^{(j)}-1}(x)| \end{pmatrix} \mathrm{e}^{-\frac{x^2}{2}}\,dx\,d\lambda \\ &\leq \frac{1}{2}\sqrt{2\pi(\alpha^{(j)})!}\left((\alpha^{(j)}-1) + \sqrt{2} + \sqrt{\alpha^{(j)}}\right)\int_1^{\lambda^{(j)}} \lambda^{\frac{\alpha^{(j)}-3}{2}}\,d\lambda \\ &\leq \left(1+2\sqrt{2}\right)\sqrt{2\pi(\alpha^{(j)})!}\left|(\lambda^{(j)})^{\frac{\alpha^{(j)}-1}{2}} - 1\right|. \end{aligned}$$

The same bound holds when $\lambda^{(j)} \leq 1$. Therefore, if $I_{\boldsymbol{\alpha},2}$ is the part of the sum bounding $I_{\boldsymbol{\alpha}}$ with indices $j \in [\![d_1+1, d]\!]$, then

$$I_{\boldsymbol{\alpha},2} \leq \sqrt{(2\pi)^d\,\boldsymbol{\alpha}!}\,(1+2\sqrt{2})\left(\rho^{(1)}\rho^{(2)}\cdots\rho^{(d_0)} + \rho^{(1)}\rho^{(2)}\cdots\rho^{(d)} - 2\,\rho^{(1)}\rho^{(2)}\cdots\rho^{(e_2)}\right).$$

To conclude, suppose that $\{\lambda^{(1)}, \ldots, \lambda^{(d)}\} \subset [\lambda_- = 1-\rho,\ \lambda_+ = 1+\rho]$. Then,

$$\begin{aligned} \frac{I_{\boldsymbol{\alpha},0}}{\prod_{i=1}^d \sqrt{2\pi(\alpha_i)!}} &\leq (\lambda_-)^{-d/2}\,(2 - (\lambda_+)^{-d/2} - (\lambda_-)^{d/2}) \leq_{\rho\to 0} d\rho \\ \frac{I_{\boldsymbol{\alpha},1}}{\prod_{i=1}^d \sqrt{2\pi(\alpha_i)!}} &\leq \frac{4d}{\sqrt{2\pi}}\,(\lambda_-)^{-d/2}\,\max(1-\lambda_+^{-1}, \lambda_-^{-1}-1) \leq_{\rho\to 0} \frac{4}{\sqrt{2\pi}}\,d\rho \\ \frac{I_{\boldsymbol{\alpha},2}}{\prod_{i=1}^d \sqrt{2\pi(\alpha_i)!}} &\leq (1+2\sqrt{2})\,(\lambda_-)^{-d/2}((\lambda_+)^d - (\lambda_-)^d) \leq_{\rho\to 0} (2+4\sqrt{2})\,d\rho. \end{aligned}$$

Hence, if the spectral radius $\rho((K_1)^{-1/2}K_2(K_1)^{-1/2} - I_d) = \rho$ is small enough, then

$$I_{\boldsymbol{\alpha}} \leq \sqrt{(2\pi)^d\,\boldsymbol{\alpha}!}\;8d\,\rho((K_1)^{-1/2}K_2(K_1)^{-1/2} - I_d).$$

The inequality follows by noticing that $\sqrt{\boldsymbol{\alpha}!} \leq \sqrt{(d+1)!}$ if $|\boldsymbol{\alpha}| \leq d+1$. $\square$

Lemma 4.10. *Fix a sequence of random vectors* $(\mathbf{S}_n)_{n\in\mathbb{N}}$ *as in Theorem 4.8, and denote* μ_n *the law of* $\mathbf{Y}_n = \mathbf{S}_n/(\sqrt{N_n D_n})$*, and* $\nu = \mathcal{N}_{\mathbb{R}^d}(\mathbf{0},K)$*. There exist positive constants* $C_1(d,K,A,B)$ *and* $C_2(d,K,A,B)$ *such that, if n is large enough, then*

$$\left(\varepsilon \geq C_1(d,K,A,B)\sqrt{\frac{D_n}{N_n}}\right) \quad\Rightarrow\quad \left(\Delta_\varepsilon(\widehat{\mu}_n,\widehat{\nu}) \leq C_2(d,K,A,B)\sqrt{\frac{D_n}{N_n}}\right).$$

Proof. Let $K_n = \mathrm{cov}(\mathbf{S}_n)/(N_nD_n)$, and $\nu_n = \mathcal{N}_{\mathbb{R}^d}(\mathbf{0},K_n)$. By Lemma 4.9, for n large enough,

$$\Delta_\varepsilon(\widehat{\nu}_n,\widehat{\nu}) \leq C_{2,\nu}(d,K,A,B)\,\sqrt{\frac{D_n}{N_n}},$$

because $\rho(K_n-K) = O(\sqrt{D_n/N_n})$ by Hypothesis (MC2′). Therefore, it suffices to find an upper bound for $\Delta_\varepsilon(\widehat{\mu}_n,\widehat{\nu}_n)$. We set $M = (K_n)^{-1/2}$, and $\boldsymbol{\zeta} = M\boldsymbol{\xi}$. We have

$$\begin{aligned}
&\widehat{\mu}_n(\boldsymbol{\zeta}) - \widehat{\nu}_n(\boldsymbol{\zeta})\\
&= \mathrm{e}^{-\frac{1}{2}\sum_{i,j=1}^d (K_n)_{ij}\,\zeta^{(i)}\zeta^{(j)}}\left(\exp\left(\sum_{r\geq 3}\frac{1}{r!}\sum_{i_1,\ldots,i_r=1}^d \frac{(\mathrm{i}^r)\,\kappa(S_n^{(i_1)},\ldots,S_n^{(i_r)})}{(N_nD_n)^{\frac{r}{2}}}\,\zeta^{(i_1)}\cdots\zeta^{(i_r)}\right)-1\right)\\
&= \mathrm{e}^{-\frac{\|\boldsymbol{\xi}\|^2}{2}}\left(\exp\left(\sum_{r\geq 3}\frac{1}{r!}\left(\frac{N_n}{D_n}\right)^{1-\frac{r}{2}}\sum_{j_1,\ldots,j_r=1}^d \delta(S_n^{(j_1)},\ldots,S_n^{(j_r)})\,\xi^{(j_1)}\cdots\xi^{(j_r)}\right)-1\right)\\
&= f(\boldsymbol{\xi}),
\end{aligned}$$

where $\delta(S_n^{(j_1)},\ldots,S_n^{(j_r)}) = \mathrm{i}^r\sum_{i_1,\ldots,i_r=1}^d M_{i_1j_1}\cdots M_{i_rj_r}\,\frac{\kappa(S_n^{(i_1)},\ldots,S_n^{(i_r)})}{N_n\,(D_n)^{r-1}}$. We can then write

$$\begin{aligned}
\Delta_\varepsilon(\widehat{\mu}_n,\widehat{\nu}_n) &= \max_{|\boldsymbol{\beta}|\in[\![0,d+1]\!]}\int_{D_\varepsilon^d}\left|\frac{\partial^{|\boldsymbol{\beta}|}(\widehat{\mu}_n-\widehat{\nu}_n)(\boldsymbol{\zeta})}{\partial\boldsymbol{\zeta}^{\boldsymbol{\beta}}}\right|d\boldsymbol{\zeta}\\
&\leq \max_{|\boldsymbol{\alpha}|\in[\![0,d+1]\!]}\left(\max_{i\in[\![1,d]\!]}\sum_{j=1}^d|M_{ij}|\right)^{|\boldsymbol{\alpha}|}\int_{M^{-1}(D_\varepsilon^d)}\left|\frac{\partial^{|\boldsymbol{\alpha}|}f(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}^{\boldsymbol{\alpha}}}\right|d\boldsymbol{\xi}\\
&\leq \max_{|\boldsymbol{\alpha}|\in[\![0,d+1]\!]}\left(d\rho((K_n)^{-1})\right)^{\frac{|\boldsymbol{\alpha}|}{2}}\int_{M^{-1}(D_\varepsilon^d)}\left|\frac{\partial^{|\boldsymbol{\alpha}|}f(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}^{\boldsymbol{\alpha}}}\right|d\boldsymbol{\xi}.
\end{aligned}$$

Until the end of the proof, the multi-index $\boldsymbol{\alpha}$ is fixed, and we need to compute bounds on the derivatives of $f(\boldsymbol{\xi})$. We have

$$\left|\frac{\partial^{|\boldsymbol{\alpha}|}f(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}^{\boldsymbol{\alpha}}}\right| \leq \sum_{\boldsymbol{\gamma}\leq\boldsymbol{\alpha}}\binom{\boldsymbol{\alpha}}{\boldsymbol{\gamma}}\,|H_{\boldsymbol{\alpha}-\boldsymbol{\gamma}}(\boldsymbol{\xi})|\,\mathrm{e}^{-\frac{\|\boldsymbol{\xi}\|^2}{2}}\left|\frac{\partial^{|\boldsymbol{\gamma}|}g(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}^{\boldsymbol{\gamma}}}\right|,$$

where

$$g(\boldsymbol{\xi}) = \exp\left(\sum_{r\geq 3}\frac{1}{r!}\left(\frac{N_n}{D_n}\right)^{1-\frac{r}{2}}\sum_{j_1,\ldots,j_r=1}^d\delta(S_n^{(j_1)},\ldots,S_n^{(j_r)})\,\xi^{(j_1)}\cdots\xi^{(j_r)}\right)-1.$$

Suppose first that $\boldsymbol{\gamma} = \mathbf{0}$. Then, $|g(\boldsymbol{\xi})| \leq |h(\boldsymbol{\xi})|\,\exp|h(\boldsymbol{\xi})|$, where

$$\begin{aligned}
|h(\boldsymbol{\xi})| &\leq \sum_{r=3}^\infty\frac{2^{r-1}r^{r-2}}{r!}\left(\frac{N_n}{D_n}\right)^{1-\frac{r}{2}}\sum_{i_1,\ldots,i_r=1}^d A^r\left|\zeta^{(i_1)}\cdots\zeta^{(i_r)}\right|\\
&\leq \frac{N_n}{4\mathrm{e}^3D_n}\sum_{r=3}^\infty\left(2A\mathrm{e}\|\boldsymbol{\zeta}\|_1\sqrt{\frac{D_n}{N_n}}\right)^r \leq 2A^3\sqrt{\frac{D_n}{N_n}}\frac{(\|\boldsymbol{\zeta}\|_1)^3}{1-2A\mathrm{e}\|\boldsymbol{\zeta}\|_1\sqrt{\frac{D_n}{N_n}}}.
\end{aligned}$$

Suppose that

$$\varepsilon \geq \varepsilon_1 = 2A\mathrm{e}\,(2d)(2d+2)^{3/2}\sqrt{\frac{D_n}{N_n}}.$$

Then, for any $\boldsymbol{\xi} \in M^{-1}(D^d_\varepsilon)$,

$$2A\mathrm{e}\|\boldsymbol{\zeta}\|_1\sqrt{\frac{D_n}{N_n}} \leq 2d\;A\mathrm{e}\,\|\boldsymbol{\zeta}\|_\infty\sqrt{\frac{D_n}{N_n}} \leq \frac{A\mathrm{e}(2d)(2d+2)^{3/2}}{\varepsilon}\sqrt{\frac{D_n}{N_n}} \leq \frac{1}{2};$$

$$|X| \leq 4A^3\sqrt{\frac{D_n}{N_n}}\,(\|\boldsymbol{\zeta}\|_1)^3 \leq \frac{(C\|\boldsymbol{\xi}\|^3)}{2}\sqrt{\frac{D_n}{N_n}},$$

where $C = 2A\sqrt{d\rho(K^{-1})}$. More generally, denote $*$ any quantity with modulus smaller than eC. Notice that

$$\begin{aligned}|\delta(S_n^{(j_1)},\ldots,S_n^{(j_r)})| = \|\delta\|_\infty \leq \|\delta\|_2 &\leq \rho(M^{\otimes r})\left\|\frac{\kappa(S_n^{(i_1)},\ldots,S_n^{(i_r)})}{N_n\,(D_n)^{r-1}}\right\|_2 \\ &\leq \frac{1}{2r^2}\left(2A\sqrt{d\rho(K^{-1})}\,r\right)^r,\end{aligned}$$

so each coefficient $\delta(S_n^{(j_1)},\ldots,S_n^{(j_r)})$ is equal to $\frac{1}{2r^2}\,(\frac{*r}{\mathrm{e}})^r$. Therefore, for $\gamma > \mathbf{0}$, $\frac{\partial^{|\gamma|}g(\xi)}{\partial\xi^\gamma} = \frac{\partial^{|\gamma|}(\exp h(\xi))}{\partial\xi^\gamma}$, with

$$h(\boldsymbol{\xi}) = \frac{N_n}{2D_n}\sum_{r=3}^{\infty}\frac{r^{r-2}}{\mathrm{e}^r\,r!}\sum_{j_1,\ldots,j_r=1}^{d}(*)^r\,\frac{\xi^{(j_1)}\cdots\xi^{(j_r)}}{(N_n/D_n)^{r/2}};$$

$$|h(\boldsymbol{\xi})| \leq \frac{(C\|\boldsymbol{\xi}\|_1)^3}{4}\sqrt{\frac{D_n}{N_n}}\,\frac{1}{1-\mathrm{e}C\|\boldsymbol{\xi}\|_1\sqrt{\frac{D_n}{N_n}}}.$$

Here and in the sequel, we assume that $\mathrm{e}C\|\boldsymbol{\xi}\|_1\sqrt{D_n/N_n} < 1$; we shall give in a moment a sufficient condition for this inequality. By the multi-dimensional version of Faà di Bruno's formula,

$$\frac{\partial^c(\exp h(\boldsymbol{\xi}))}{\partial\xi^{(i_1)}\cdots\partial\xi^{(i_c)}} = \left(\sum_{\pi\in\mathfrak{P}(c)}\prod_{j=1}^{\ell(\pi)}\frac{\partial^{|\pi_j|}h(\boldsymbol{\xi})}{\prod_{k\in\pi_j}\partial\xi^{(i_k)}}\right)\exp(h(\boldsymbol{\xi})),$$

where $\mathfrak{P}(c)$ is the set of set partitions $\pi = \pi_1 \sqcup \pi_2 \sqcup \cdots \sqcup \pi_\ell$ of $[\![1,c]\!]$. In this formula, writing $r^{\downarrow p} = r(r-1)\cdots(r-p+1)$,

$$\frac{\partial^p h(\boldsymbol{\xi})}{\prod_{k\in\pi_j}\partial\xi^{(i_k)}} = \frac{N_n}{2D_n}\sum_{r=\max(3,p)}^{\infty}\frac{r^{r-2}}{\mathrm{e}^r\,r!}\sum_{j_1,\ldots,j_{r-p}=1}^{d}\frac{(*)^r\,r^{\downarrow p}}{(N_n/D_n)^{r/2}}\,\xi^{(j_1)}\cdots\xi^{(j_{r-p})};$$

$$\begin{aligned}\left|\frac{\partial^p h(\boldsymbol{\xi})}{\prod_{k\in\pi_j}\partial\xi^{(i_k)}}\right| &\leq \frac{N_n}{4\mathrm{e}^3 D_n}\sum_{r=\max(3,p)}^{\infty}\frac{(\mathrm{e}C)^r\,(\|\boldsymbol{\xi}\|_1)^{r-p}\,r^{\downarrow p}}{(N_n/D_n)^{r/2}} \\ &\leq \frac{(\mathrm{e}C)^{\max(3,p)}\,(\max(3,p))!\,(\|\boldsymbol{\xi}\|_1)^{\max(3-p,0)}}{4\mathrm{e}^3\,(N_n/D_n)^{\frac{\max(3,p)}{2}-1}\left(1-\mathrm{e}C\|\boldsymbol{\xi}\|_1\sqrt{\frac{D_n}{N_n}}\right)^{p+1}},\end{aligned}$$

where for the last inequality one has to treat separately the cases $p \geq 3$ and $p = 1$ or 2.

Suppose that $\boldsymbol{\zeta} \in C^d_{(\mathbf{0},(2d+2)^{3/2}/\varepsilon)} = D^d_\varepsilon$. Then,

$$\|\boldsymbol{\xi}\|_1 \leq \sqrt{d}\,\|\boldsymbol{\xi}\|_2 \leq \sqrt{d\rho(K)}\,\|\boldsymbol{\zeta}\|_2 \leq d\sqrt{\rho(K)}\,\|\boldsymbol{\zeta}\|_\infty \leq \frac{(2d)(2d+2)^{3/2}\sqrt{\rho(K)}}{2\varepsilon};$$

$$\mathrm{e}C\|\boldsymbol{\xi}\|_1\sqrt{D_n/N_n} \leq \frac{\varepsilon_1}{\varepsilon}\sqrt{d\,\rho(K)\rho(K^{-1})} = \frac{\varepsilon_1}{\varepsilon}\sqrt{d\,\tau(K)}.$$

Set

$$\varepsilon_2 = 2\sqrt{d\,\tau(K)}\,\varepsilon_1 = 16A\mathrm{e}\sqrt{2(d(d+1))^3\,\tau(K)}\sqrt{\frac{D_n}{N_n}}.$$

Then, if $\varepsilon \geq \varepsilon_2$ (this condition is stronger than $\varepsilon \geq \varepsilon_1$) and $\boldsymbol{\zeta} \in D^d_\varepsilon$, we have for any $p \geq 1$:

$$\left|\frac{\partial^p h(\boldsymbol{\xi})}{\prod_{k\in\pi_j}\partial\xi^{(i_k)}}\right| \leq \frac{(\mathrm{e}C)^{\max(3,p)}\,2^{p+1}\,(\max(3,p))!\,(\|\boldsymbol{\xi}\|_1)^{\max(3-p,0)}}{4\mathrm{e}^3\,(N_n/D_n)^{\frac{\max(3,p)}{2}-1}}.$$

As a consequence,

$$J_{\boldsymbol{\alpha}} = \int_{M^{-1}(D^d_\varepsilon)}\left|\frac{\partial^{|\boldsymbol{\alpha}|}f(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}^{\boldsymbol{\alpha}}}\right|d\boldsymbol{\xi} \leq \sum_{\gamma\leq\boldsymbol{\alpha}}\sum_{\pi\in\mathfrak{P}(|\gamma|)}\int_{M^{-1}(D^d_\varepsilon)}\mathrm{e}^{-\frac{\|\boldsymbol{\xi}\|^2}{2}+\frac{(C\|\boldsymbol{\xi}\|)^3}{2}\sqrt{\frac{D_n}{N_n}}}\,\frac{P_{\gamma,\boldsymbol{\alpha},\pi}(\boldsymbol{\xi})}{(N_n/D_n)^{e(\pi)}}\,d\boldsymbol{\xi},$$

where each $P_{\gamma,\boldsymbol{\alpha},\pi}(\boldsymbol{\xi})$ is a quantity that is bounded by a polynomial in $\|\boldsymbol{\xi}\|_1$, and with

$$e(\pi) = \sum_{i=1}^{\ell(\pi)}\left(\frac{\max(3,|\pi_i|)}{2}-1\right).$$

Note that if $\|\boldsymbol{\xi}\|\sqrt{\frac{D_n}{N_n}}$ is small enough, then

$$-\frac{\|\boldsymbol{\xi}\|^2}{2}+\frac{(C\|\boldsymbol{\xi}\|)^3}{2}\sqrt{\frac{D_n}{N_n}} \leq -\frac{\|\boldsymbol{\xi}\|^2}{4}.$$

Therefore, one can find a positive constant $C_1(d,K,A,B) \geq \varepsilon_2\sqrt{N_n/D_n}$ such that, if $\varepsilon \geq C_1(d,K,A,B)\,\sqrt{D_n/N_n}$, then

$$J_{\boldsymbol{\alpha}} \leq \sum_{\gamma\leq\boldsymbol{\alpha}}\sum_{\pi\in\mathfrak{P}(|\gamma|)}\frac{1}{(N_n/D_n)^{e(\pi)}}\int_{\mathbb{R}^d}\mathrm{e}^{-\frac{\|\boldsymbol{\xi}\|^2}{4}}\,P_{\gamma,\boldsymbol{\alpha},\pi}(\boldsymbol{\xi})\,d\boldsymbol{\xi}.$$

All the integrals appearing in the right-hand side of this inequality are convergent, and the main contribution to the sum is provided by set partitions such that $e(\pi) = \frac{1}{2}$ is minimal. This only happens when $\ell(\pi) = 1$ and $|\gamma| \leq 3$. Therefore, keeping only these terms, we obtain

$$\begin{aligned}J_{\boldsymbol{\alpha}} &\leq \sum_{\substack{\gamma\leq\boldsymbol{\alpha}\\|\gamma|\leq 3}}\binom{\boldsymbol{\alpha}}{\gamma}\int_{\mathbb{R}^d}\mathrm{e}^{-\frac{\|\boldsymbol{\xi}\|^2}{2}}\,|H_{\boldsymbol{\alpha}-\gamma}(\boldsymbol{\xi})|\,\frac{\partial^{|\gamma|}h(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}^\gamma}\,g(\boldsymbol{\xi})\,d\boldsymbol{\xi}+O\left(\frac{D_n}{N_n}\right),\\ &\leq \sum_{\substack{\gamma\leq\boldsymbol{\alpha}\\|\gamma|\leq 3}}\binom{\boldsymbol{\alpha}}{\gamma}\int_{\mathbb{R}^d}\mathrm{e}^{-\frac{\|\boldsymbol{\xi}\|^2}{4}}\,|H_{\boldsymbol{\alpha}-\gamma}(\boldsymbol{\xi})|\,\frac{\partial^{|\gamma|}h(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}^\gamma}\,d\boldsymbol{\xi}+O\left(\frac{D_n}{N_n}\right)\end{aligned}$$

where the constant in the O only depends on d,K,A,B. Using the previous estimates on the derivatives of $h(\boldsymbol{\xi})$, we conclude that

$$J_{\boldsymbol{\alpha}} = O\left(\sqrt{\frac{D_n}{N_n}}\sum_{\substack{\gamma\leq\boldsymbol{\alpha}\\|\gamma|\leq 3}}\binom{\boldsymbol{\alpha}}{\gamma}\int_{\mathbb{R}^d}\mathrm{e}^{-\frac{\|\boldsymbol{\xi}\|^2}{4}}\,|H_{\boldsymbol{\alpha}-\gamma}(\boldsymbol{\xi})|\,(\|\boldsymbol{\xi}\|_1)^{3-|\gamma|}\,d\boldsymbol{\xi}\right)+O\left(\frac{D_n}{N_n}\right).$$

Therefore, there exists a positive constant $C_{2,\mu}(d,K,A,B)$ such that, if $\varepsilon \geq C_1(d,K,A,B)$, then

$$\Delta_\varepsilon(\widehat{\mu}_n,\widehat{\nu}_n) \leq C_{2,\mu}(d,K,A,B)\,\sqrt{\frac{D_n}{N_n}}.$$

This ends the proof of the lemma, taking $C_2 = C_{2,\mu} + C_{2,\nu}$. □

Proof of Theorem 4.8. We apply Corollary 2.14 to the inequality of Lemma 4.10, with $\varepsilon = C_1(d,K,A,B)\sqrt{D_n/N_n}$ and the regularity constant $R = 2\sqrt{(d+1)\rho(K^{-1})}$. □

4.3. **Sparse dependency graphs.** A large set of applications of Theorems 4.7 and 4.8 is provided by the theory of *dependency graphs*, which has already been used successfully in dimension $d = 1$ in [FMN16; FMN19].

Definition 4.11 (Dependency graph)**.** *Let $(\mathbf{A}_v)_{v\in V_{\mathbb{R}^d}}$ be a family of random vectors in $\mathbb{R}^d$. A simple undirected graph $G_{\mathbb{R}^d} = (V_{\mathbb{R}^d}, E_{\mathbb{R}^d})$ is called a dependency graph for this family if, for any disjoint subsets V_1 and V_2 of $V_{\mathbb{R}^d}$, if there is no edge $e = (v_1,v_2)$ in $E_{\mathbb{R}^d}$ that connects $v_1 \in V_1$ and $v_2 \in V_2$, then $(\mathbf{A}_v)_{v\in V_1}$ and $(\mathbf{A}_v)_{v\in V_2}$ are independent families of random vectors of $\mathbb{R}^d$.*

The parameters of a dependency graph $G_{\mathbb{R}^d}$ are:

(1) its *size* $N = \operatorname{card} V_{\mathbb{R}^d}$.

(2) its *maximal degree* D: every vertex $v \in V_{\mathbb{R}^d}$ has at most $D-1$ neighbors w such that $\{v,w\} \in E_{\mathbb{R}^d}$.

In the sequel, we always put an index $\mathbb{R}^d$ on the dependency graph of a family of d-dimensional random vectors. The reason for this is the following. Fix a family $(\mathbf{A}_v)_{v\in V_{\mathbb{R}^d}}$ of random vectors in $\mathbb{R}^d$, a dependency graph $G_{\mathbb{R}^d} = (V_{\mathbb{R}^d}, E_{\mathbb{R}^d})$ for this family, and denote $W = V_{\mathbb{R}^d} \times [\![1,d]\!]$. To any pair $w = (v,i) \in W$, we associate the real random variable

$$A_w = A_v^{(i)}.$$

We then endow the vertices of W with the following structure of graph $G = (W,E)$:

$$(v_1,i_1) \sim_G (v_2,i_2) \quad\iff\quad (v_1 = v_2 \text{ and } i_1 \neq i_2) \text{ or } (v_1 \sim_{G_{\mathbb{R}^d}} v_2).$$

It is then easily seen that $(A_w)_{w\in W}$ is a family of real random variables with dependency graph $G = (W,E)$. Moreover, if N and D are the parameters of $G_{\mathbb{R}^d}$, then dN and dD are the parameters of G.

Theorem 4.12 (Upper bound on cumulants)**.** *Let*

$$\mathbf{S} = \sum_{v\in V_{\mathbb{R}^d}} \mathbf{A}_v$$

be a sum of random vectors in $\mathbb{R}^d$, such that $G_{\mathbb{R}^d} = (V_{\mathbb{R}^d}, E_{\mathbb{R}^d})$ is a dependency graph for the family $(\mathbf{A}_v)_{v\in V_{\mathbb{R}^d}}$. We also suppose that $\|\mathbf{A}_v\|_\infty \leq A$ for any $v \in V_{\mathbb{R}^d}$. For any choice of indices $i_1,\ldots,i_r \in [\![1,d]\!]$,

$$|\kappa(S^{(i_1)},\ldots,S^{(i_r)})| \leq N\,(2D)^{r-1}\,A^r\,r^{r-2},$$

where N and D are the parameters of the graph $G_{\mathbb{R}^d}$.

Proof. Let $G = (W, E)$ be the dependency graph on one-dimensional random variables $A_v^{(i)}$ that is obtained from $G_{\mathbb{R}^d}$ by the aforementioned construction. By [FMN16, Equation 9.9], for any choice of vertices $v_1, \ldots, v_r$,

$$\left|\kappa(A_{v_1}^{(i_1)}, \ldots, A_{v_r}^{(i_r)})\right| \leq 2^{r-1}\, A^r\, \mathrm{ST}_H,$$

where $H = G[(v_1, i_1), \ldots, (v_r, i_r)]$ is the (multi)graph induced by G on the set of vertices $\{(v_j, i_j) \,|\, j \in [\![1, r]\!]\}$ (see [FMN16, Section 9.3.2]), and ST_H is the number of spanning trees of H (hence, 0 if H is not connected). Therefore, if we extend by multilinearity the joint cumulant of the coordinates of the sum S, then we obtain

$$\begin{aligned}\left|\kappa(S^{(i_1)}, \ldots, S^{(i_r)})\right| &\leq \sum_{v_1, \ldots, v_r \in V_{\mathbb{R}^d}} \left|\kappa(A_{v_1}^{(i_1)}, \ldots, A_{v_r}^{(i_r)})\right| \\ &\leq 2^{r-1} A^r \sum_{v_1, \ldots, v_r \in V_{\mathbb{R}^d}} \mathrm{ST}_{G[(v_1,i_1), \ldots, (v_r, i_r)]}.\end{aligned}$$

We now use an argument similar to the one of [FMN16, Lemma 9.3.5]. Consider a pair $(T, (v_1, \ldots, v_r))$ such that $v_1, \ldots, v_r \in V_{\mathbb{R}^d}$, and T is a spanning tree on $[\![1, r]\!]$ included in the induced subgraph $G[(v_1, i_1), \ldots, (v_r, i_r)]$. By Cayley's formula for the number of spanning trees on r vertices, there are r^{r-2} possible choices for T. Then, to choose $(v_1, \ldots, v_r)$ such that $T \subset G[(v_1, i_1), \ldots, (v_r, i_r)]$, we proceed as follows. There are N possible choices for the vector v_1. Then, for each j such that $1 \sim j$ in T, there are at most D possible choices for v_j, such that either $(v_1, i_1) = (v_j, i_j)$, or $(v_1, i_1) \sim_G (v_j, i_j)$. Indeed, one can only choose v_j among the vertices v_j with $v_1 \sim_{G_{\mathbb{R}^d}} v_j$, plus v_1. We pursue this reasoning with the neighbors of the neighbors j of 1 in T, performing a breadth-first search of T. We conclude that T being fixed, the number of compatible choices for $v_1, \ldots, v_r$ is smaller than $N\, D^{r-1}$. This ends the proof of the theorem. $\square$

By combining Theorems 4.7, 4.8 and 4.12, we obtain the following general result on sums of random vectors with sparse dependency graphs:

Theorem 4.13 (Sums of random vectors with sparse dependency graphs). *Consider a sequence of sums $(\mathbf{S}_n = \sum_{i=1}^{N_n} \mathbf{A}_{i,n})_{n \in \mathbb{N}}$ of centered random vectors in $\mathbb{R}^d$. We endow each family $(\mathbf{A}_{i,n})_{i \in [\![1, N_n]\!]}$ with a dependency graph of parameters N_n and D_n, and we make the following assumptions:*

1. *All the random vectors $\mathbf{A}_{i,n}$ are bounded by A in norm $\|\cdot\|_\infty$.*
2. *The sequences of parameters $(N_n)_{n\in\mathbb{N}}$ and $(D_n)_{n\in\mathbb{N}}$ satisfy*
$$\lim_{n\to\infty} N_n = +\infty \qquad ; \qquad \lim_{n\to\infty} \frac{D_n}{N_n} = 0.$$
3. *There exists a matrix $K \in \mathrm{S}_+(d, \mathbb{R})$, and a family of real numbers $(L_{i,j,k})_{1 \leq i,j,k \leq d}$, such that*
$$\frac{\mathrm{cov}(\mathbf{S}_n)}{N_n D_n} = K + J_n \quad \text{with } \rho(J_n) \leq B\sqrt{\frac{D_n}{N_n}}.$$

The sums $\mathbf{S}_n$ satisfy:

- *Central limit theorem: if $\mathbf{Y}_n = \frac{\mathbf{S}_n}{\sqrt{N_n D_n}}$, then $\mathbf{Y}_n \rightharpoonup \mathcal{N}_{\mathbb{R}^d}(\mathbf{0}, K)$.*
- *Speed of convergence: there exists a constant $C(d, K, A, B)$ such that*
$$d_{\mathrm{convex}}(\mathbf{Y}_n\,,\, \mathcal{N}_{\mathbb{R}^d}(\mathbf{0}, K)) \leq C(d, K, A, B)\, \sqrt{\frac{D_n}{N_n}}.$$

If one adds the assumptions (MC4) and (MC5) on $\mathbf{S}_n$, then one also gets large deviation estimates as in Theorem 4.7.

We conclude this section by examining three examples:

Example 4.14 (Random walks with dependent steps, $d = 2$). One can construct an example in dimension $d = 2$ with almost arbitrary parameters $N_n \to +\infty$ and $D_n = o(N_n)$. If λ is a positive parameter, denote $\mathcal{L}(\lambda)$ the law of the Brownian motion on the unit circle $\mathbb{S}^1 = \{z \in \mathbb{C} \,|\, |z| = 1\}$, taken at time $t = \lambda$. This law is given by the density

$$\rho_\lambda(\theta) = \frac{1}{\sqrt{2\pi\lambda}} \sum_{k\in\mathbb{Z}} \mathrm{e}^{-\frac{(\theta-2k\pi)^2}{2\lambda}}$$

for $\theta \in (-\pi, \pi)$. Thus, if $\mathbf{X} = \mathrm{e}^{\mathrm{i}\theta}$ follows the law $\mathcal{L}(\lambda)$, then

$$\mathbb{P}[\theta \in (a,b)] = \frac{1}{\sqrt{2\pi\lambda}} \sum_{k\in\mathbb{Z}} \int_a^b \mathrm{e}^{\frac{(\theta-2k\pi)^2}{2\lambda}}\,d\theta$$

for any a, b such that $-\pi < a < b < \pi$. The density of the angle θ can be rewritten in Fourier series:

$$\rho_\lambda(\theta)\,d\theta = \left(\sum_{l\in\mathbb{Z}} \mathrm{e}^{-\frac{\lambda l^2}{2}}\,\mathrm{e}^{\mathrm{i}l\theta}\right)\frac{d\theta}{2\pi}.$$

The family of laws $(\mathcal{L}(\lambda))_{\lambda\in\mathbb{R}_+}$ is a semigroup of probability measures on the multiplicative semigroup $\mathbb{S}^1$. Thus, if $\mathbf{U}_1, \ldots, \mathbf{U}_D$ follow the same law $\mathcal{L}(\frac{\lambda}{D})$ and are independent, then the product $\mathbf{U}_1\mathbf{U}_2\cdots\mathbf{U}_D$ follows the law $\mathcal{L}(\lambda)$.

Consider a family of independent random variables $(\mathbf{U}_i)_{i\in[\![1,N_n]\!]}$, all following the same law $\mathcal{L}(\frac{\lambda}{D_n})$. For $i \in [\![1, N_n]\!]$, we denote $\mathbf{X}_i = \mathbf{U}_i\mathbf{U}_{i+1}\cdots\mathbf{U}_{i+D_n-1}$, where the $\mathbf{U}_i$'s are labelled cyclically, that is to say that $\mathbf{U}_{N_n+k} = \mathbf{U}_k$. Two variables $\mathbf{X}_i$ and $\mathbf{X}_j$ are independent unless they share some common factor $\mathbf{U}_k$, which happens if and only if the distance between i and j in $\mathbb{Z}/N_n\mathbb{Z}$ is smaller than D_n. Thus, $(\mathbf{X}_i)_{i\in[\![1,N_n]\!]}$ is a family of dependent random variables in $\mathbb{R}^2 \supset \mathbb{S}^1$, with a dependency graph of parameters N_n and $2D_n - 1$. By the previous remark, each $\mathbf{X}_i$ follows the law $\mathcal{L}(\lambda)$; in particular,

$$\mathbb{E}[\mathbf{X}_i] = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}\theta}\,\rho_\lambda(\theta)\,d\theta = \sum_{l\in\mathbb{Z}} \mathrm{e}^{-\frac{\lambda l^2}{2}} \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}(l+1)\theta}\,\frac{d\theta}{2\pi} = \mathrm{e}^{-\frac{\lambda}{2}}.$$

On the other hand, if $\mathbf{X}_i$ and $\mathbf{X}_j$ are two variables with $d(i,j) = D < D_n$ in $\mathbb{Z}/N_n\mathbb{Z}$, then one can compute the covariance matrix of these two vectors of $\mathbb{R}^2$:

$$\mathbb{E}[\mathrm{Re}(\mathbf{X}_i)\,\mathrm{Re}(\mathbf{X}_j)] = \frac{\mathbb{E}[\mathbf{X}_i\mathbf{X}_j + \overline{\mathbf{X}_i}\mathbf{X}_j + \mathbf{X}_i\overline{\mathbf{X}_j} + \overline{\mathbf{X}_i\mathbf{X}_j}]}{4} = \frac{\mathrm{e}^{-\frac{D}{D_n}\lambda} + \mathrm{e}^{-(2-\frac{D}{D_n})\lambda}}{2};$$
$$\mathbb{E}[\mathrm{Im}(\mathbf{X}_i)\,\mathrm{Im}(\mathbf{X}_j)] = \frac{\mathbb{E}[-\mathbf{X}_i\mathbf{X}_j + \overline{\mathbf{X}_i}\mathbf{X}_j + \mathbf{X}_i\overline{\mathbf{X}_j} - \overline{\mathbf{X}_i\mathbf{X}_j}]}{4} = \frac{\mathrm{e}^{-\frac{D}{D_n}\lambda} - \mathrm{e}^{-(2-\frac{D}{D_n})\lambda}}{2};$$
$$\mathbb{E}[\mathrm{Re}(\mathbf{X}_i)\,\mathrm{Im}(\mathbf{X}_j)] = \mathbb{E}[\mathrm{Im}(\mathbf{X}_i)\,\mathrm{Re}(\mathbf{X}_j)] = 0.$$

So,

$$\mathrm{cov}(\mathrm{Re}(\mathbf{X}_i), \mathrm{Re}(\mathbf{X}_j)) = \frac{\mathrm{e}^{-\frac{D}{D_n}\lambda} + \mathrm{e}^{-(2-\frac{D}{D_n})\lambda} - 2\mathrm{e}^{-\lambda}}{2};$$
$$\mathrm{cov}(\mathrm{Im}(\mathbf{X}_i), \mathrm{Im}(\mathbf{X}_j)) = \frac{\mathrm{e}^{-\frac{D}{D_n}\lambda} - \mathrm{e}^{-(2-\frac{D}{D_n})\lambda}}{2};$$
$$\mathrm{cov}(\mathrm{Re}(\mathbf{X}_i), \mathrm{Im}(\mathbf{X}_j)) = \mathrm{cov}(\mathrm{Im}(\mathbf{X}_i), \mathrm{Re}(\mathbf{X}_j)) = 0.$$

As a consequence, if $\mathbf{S}_n = \sum_{i=1}^{N_n} \mathbf{X}_i$, then

$$\begin{aligned}
\frac{\mathrm{cov}(\mathbf{S}_n)}{(2D_n-1)\,N_n} &= \frac{1}{2(2D_n-1)} \sum_{D=-(D_n-1)}^{D_n-1} \begin{pmatrix} \mathrm{e}^{-\frac{|D|}{D_n}\lambda}+\mathrm{e}^{-(2-\frac{|D|}{D_n})\lambda}-2\mathrm{e}^{-\lambda} & 0 \\ 0 & \mathrm{e}^{-\frac{|D|}{D_n}\lambda}-\mathrm{e}^{-(2-\frac{|D|}{D_n})\lambda} \end{pmatrix} \\
&= \frac{1}{2D_n-1} \begin{pmatrix} \frac{1-\mathrm{e}^{-\lambda}}{1-\mathrm{e}^{-\lambda/D_n}}+\mathrm{e}^{-2\lambda}\,\frac{\mathrm{e}^{\lambda}-1}{\mathrm{e}^{\lambda/D_n}-1}-1-(2D_n-1)\,\mathrm{e}^{-\lambda} & 0 \\ 0 & \frac{1-\mathrm{e}^{-\lambda}}{1-\mathrm{e}^{-\lambda/D_n}}-\mathrm{e}^{-2\lambda}\,\frac{\mathrm{e}^{\lambda}-1}{\mathrm{e}^{\lambda/D_n}-1} \end{pmatrix} \\
&= K + O\left(\frac{1}{D_n}\right) \qquad \text{with } K = \frac{1}{2\lambda}\begin{pmatrix} 1-2\lambda\mathrm{e}^{-\lambda}-\mathrm{e}^{-2\lambda} & 0 \\ 0 & (1-\mathrm{e}^{-\lambda})^2 \end{pmatrix}.
\end{aligned}$$

Therefore, one can apply Theorem 4.13 if $D_n = o(N_n)$ and $(N_n)^{1/3} = O(D_n)$. In this situation, $\mathbf{Y}_n = \frac{\mathbf{S}_n - N_n\,\mathrm{e}^{-\frac{\lambda}{2}}}{\sqrt{(2D_n-1)\,N_n}}$ converges in law towards $\mathcal{N}_{\mathbb{R}^2}(\mathbf{0}, K)$, with

$$d_{\mathrm{convex}}(\mathbf{Y}_n, \mathcal{N}_{\mathbb{R}^2}(\mathbf{0}, K)) = O\left(\sqrt{\frac{D_n}{N_n}}\right).$$

We have drawn on Figure 8 a random walk with steps $\mathbf{X}_i$, with $\lambda = 1$, $N_n = 1000$ and $D_n = 40$ or $D_n = 1$. The deviation of the whole sum $\mathbf{S}_n$ from its mean $1000\,\mathrm{e}^{-1/2} \simeq 600$ follows approximatively a (non-standard) Gaussian law of size $O(\sqrt{D_n\,N_n}) = O(200)$.

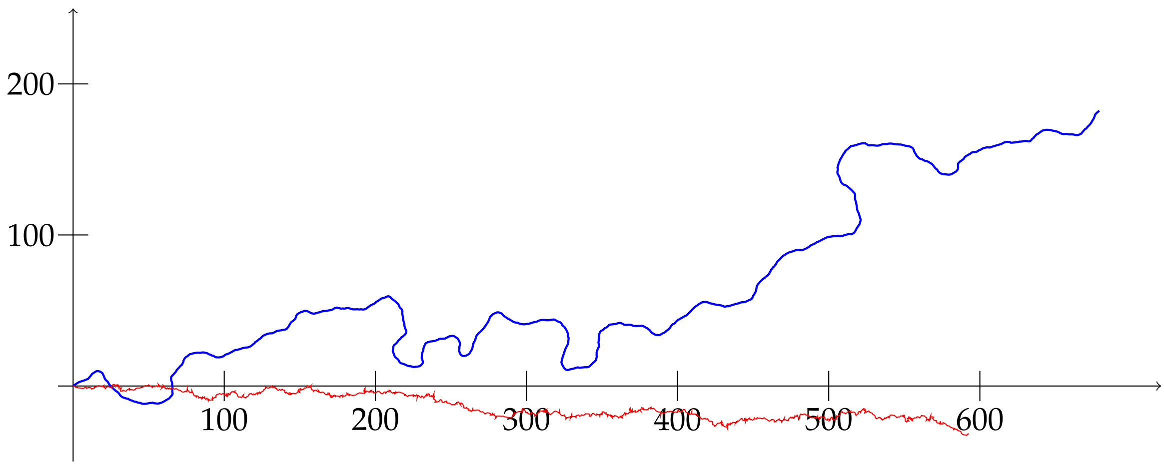

FIGURE 8. Random walks associated to a sum of dependent random vectors in $\mathbb{R}^2$ (thick blue line, $N_n = 1000$ and $D_n = 40$), and to a sum of independent random vectors (thin red line, $N_n = 1000$ and $D_n = 1$).

Example 4.15 (Random walks with dependent steps, $d \geq 2$). Let us describe a generalization of the previous example with an arbitrary dimension $d \geq 2$. Denote $p_{\mathrm{SO}(d),t}(g)$ the density with respect to the Haar measure of a Brownian motion traced on the special orthogonal group $G = \mathrm{SO}(d)$, taken at time $t > 0$. It can be shown that, if $\widehat{G}$ is a set of representatives of the isomorphism classes of finite-dimensional irreducible representations of G, and if χ^λ and d^λ denote respectively the normalised character and the dimension of an irreducible representation $\lambda \in \widehat{G}$, then there exists positive rational coefficients c_λ such that

$$p_{\mathrm{SO}(d),t}(g) = \sum_{\lambda \in \widehat{G}} \mathrm{e}^{-\frac{c_\lambda t}{2}}\,(d^\lambda)^2\,\chi^\lambda(g).$$

We refer to [Mél14, Section 2] for precisions on Brownian motions on compact Lie groups and their symmetric quotients. The coefficients c_λ can be computed by associating to each class of isomorphism $\lambda \in \widehat{G}$ the highest weight of the underlying representation, which sits in a part of the lattice of characters $\widehat{T}$ of a maximal connected torus T in $\mathrm{SO}(d)$. Then, $c_\lambda = \langle \lambda \mid \lambda + 2\rho \rangle$, where 2ρ is the sum of all positive roots of G, and $\langle \cdot \mid \cdot \rangle$ is an appropriate scalar product on $\widehat{T} \otimes_{\mathbb{Z}} \mathbb{R}$.

We denote $\mathcal{L}_{\mathrm{SO}(d)}(t)$ the law on $\mathrm{SO}(d)$ with density $p_{\mathrm{SO}(d),t}(\cdot)$, and we consider a family $(u_i)_{i \in [\![1,N_n]\!]}$ of independent random variables with values in $\mathrm{SO}(d)$, all following the same law $\mathcal{L}_{\mathrm{SO}(d)}(\frac{\lambda}{D_n})$. We then set $g_i = u_i u_{i+1} \cdots u_{i+D_n-1}$, where as in the previous example, the u_i's are labelled cyclically. Each random variable g_i follows the same law $\mathcal{L}_{\mathrm{SO}(d)}(\lambda)$, and g_i and g_j are independent if and only if i and j are at distance larger than D_n in $\mathbb{Z}/N_n\mathbb{Z}$. We finally set $\mathbf{X}_i = g_i(\mathbf{e}_d)$, where $\mathbf{e}_d = (0,\ldots,0,1) \in \mathbb{R}^d$. Each $\mathbf{X}_i$ follows the same law $\mathcal{L}_{\mathbb{S}^{d-1}}(\lambda)$, which is the law of a Brownian motion on the sphere $\mathbb{S}^{d-1}$ starting from the vector $\mathbf{e}_d$ and taken at time $t = \lambda$. The density of this law with respect to the Haar measure on the sphere $\mathbb{S}^{d-1}$ is given by the following formula:

$$p_{\mathbb{S}^{d-1},t}(\mathbf{x}) = \sum_{\lambda \in \widehat{G}^K} \mathrm{e}^{-\frac{c_\lambda t}{2}}\, d^\lambda\, \phi^\lambda(\mathbf{x}).$$

Here, $\widehat{G}^K$ is the subset of $\widehat{G}$ that consists in spherical representations, that is to say, irreducible representations of $G = \mathrm{SO}(d)$ that admits a (normalised) $K = \mathrm{SO}(d-1)$-invariant vector $\mathbf{e}^\lambda$. On the other hand, $\phi^\lambda(\mathbf{x}) = \left\langle \rho^\lambda(g)\,(\mathbf{e}^\lambda) \,\middle|\, \mathbf{e}^\lambda \right\rangle$, where $g \in G$ is any element such that $g(\mathbf{e}_d) = \mathbf{x}$; ρ^λ is the defining morphism of the spherical representation λ; and $\langle \cdot \mid \cdot \rangle$ is a G-invariant scalar product on the representation space, with $\left\langle \mathbf{e}^\lambda \,\middle|\, \mathbf{e}^\lambda \right\rangle = 1$. Standard arguments of harmonic analysis on the real spheres (see [AH10; VV09]) ensure that there is a labeling of the spherical representations in $\widehat{G}^K$ by the set of natural integers $\mathbb{N}$. Indeed, in terms of highest weights, $\widehat{G}^K = \mathbb{N}\omega_0$, where ω_0 is the highest weight associated to the geometric representation of $\mathrm{SO}(d)$ on $\mathbb{C}^d$. Then, $c_{k\omega_0} = k^2 + (d-2)k$, $d^{k\omega_0} = \frac{2k+d-2}{k+d-2}\binom{k+d-2}{d-2}$ by Weyl's formula for dimensions of irreducible representations, and

$$\phi^{k\omega_0}(\mathbf{x}) = P^{d-1,k}(x^{(d)}),$$

where $(P^{d-1,k}(t))_{k \in \mathbb{N}}$ is the set of Legendre polynomials associated to the sphere $\mathbb{S}^{d-1}$: they are the orthogonal polynomials with respect to the weight

$$\frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{1}{2})\,\Gamma(\frac{d-1}{2})}\,(1-t^2)^{\frac{d-3}{2}}\,1_{t \in [-1,1]}\,dt,$$

normalised by the condition $P^{d-1,k}(1) = 1$. So,

$$p_{\mathbb{S}^{d-1},t}(\mathbf{x}) = \sum_{k=0}^{\infty} \mathrm{e}^{-\frac{(k^2+(d-2)k)t}{2}}\, \frac{2k+d-2}{k+d-2}\binom{k+d-2}{d-2}\, P^{d-1,k}(x^{(d)}).$$

The expectation of a variable $\mathbf{X}_i$ following the law $\mathcal{L}_{\mathbb{S}^{d-1}}(\lambda)$ is

$$\sum_{k=0}^{\infty} \mathrm{e}^{-\frac{(k^2+(d-2)k)t}{2}}\, d^{k\omega_0} \int_{\mathbb{S}^{d-1}} \mathbf{x}\,\phi^{k\omega_0}(\mathbf{x})\,\mathrm{Haar}(d\mathbf{x}).$$

For every coordinate $i \in [\![1,d]\!]$, $x^{(i)}$ is a coefficient of the geometric representation of $\mathrm{SO}(d)$. Since coefficients of distinct irreducible representations of $\mathrm{SO}(d)$ are orthogonal by Schur's lemma, the scalar product of $x^{(i)}$ with the spherical function $\phi^{k\omega_0}$ is thus equal to 0, unless $k = 1$. Then, $\phi^{\omega_0}(\mathbf{x}) = x^{(d)}$, and

$$d^{\omega_0} \int_{\mathbb{S}^{d-1}} \mathbf{x}\,x^{(d)}\,\mathrm{Haar}(d\mathbf{x}) = \mathbf{e}_d.$$

Hence, $\mathbb{E}[\mathbf{X}_i] = \mathrm{e}^{-\frac{(d-1)\lambda}{2}}\,\mathbf{e}_d$. Let us now compute the covariance matrix between two variables $\mathbf{X}_i$ and $\mathbf{X}_j$. We fix two coordinates k and l in $[\![1,d]\!]$, and consider $\mathbb{E}[X_i^{(k)}X_j^{(l)}]$. Since $\mathbf{X}_i$ and $\mathbf{X}_j$ are independent if i and j are at distance larger than D_n in $\mathbb{Z}/N_n\mathbb{Z}$, we can exclude this case and assume $d(i,j) = D < D_n$. Then, with $p_{\mathrm{SO}(d),t} = p_t$,

$$\begin{aligned}
&\mathbb{E}\left[X_i^{(k)}X_j^{(l)}\right] \\
&= \int_{G^3} p_{\frac{D}{D_n}\lambda}(f)\,p_{\frac{D_n-D}{D_n}\lambda}(g)\,p_{\frac{D}{D_n}\lambda}(h)\,\langle (fg)(\mathbf{e}_d)\,|\,\mathbf{e}_k\rangle\,\langle (gh)(\mathbf{e}_d)\,|\,\mathbf{e}_l\rangle\,df\,dg\,dh \\
&= \int_{G^3} p_{\frac{D}{D_n}\lambda}(f)\,p_{\frac{D_n-D}{D_n}\lambda}(g)\,p_{\frac{D}{D_n}\lambda}(h)\,\left\langle g(\mathbf{e}_d)\,\middle|\,f^{-1}(\mathbf{e}_k)\right\rangle\,\left\langle h(\mathbf{e}_d)\,\middle|\,g^{-1}(\mathbf{e}_l)\right\rangle\,df\,dg\,dh \\
&= \mathrm{e}^{-\frac{(d-1)D}{D_n}\lambda}\int_G p_{\frac{D_n-D}{D_n}\lambda}(g)\,\langle g(\mathbf{e}_d)\,|\,\mathbf{e}_k\rangle\,\langle g(\mathbf{e}_d)\,|\,\mathbf{e}_l\rangle\,dg
\end{aligned}$$

by using on the second line the invariance of the scalar product under $G = \mathrm{SO}(d)$, and on the third line the computation of the expectation of the law $\mathcal{L}_{\mathbb{S}^{d-1}}(t)$. The remaining integral is the expectation of the product of matrix coefficients $g_{dk}g_{dl}$ under the law $\mathcal{L}_{\mathrm{SO}(d)}(\frac{D_n-D}{D_n}\lambda)$ of the Brownian motion on the group. This computation is performed in [Mél14, Section 4] (beware that the normalisation of time in this article differs by a factor d). Hence, at time t, one has

$$\begin{aligned}
\mathbb{E}[(g_{dd})^2] &= \frac{1}{d} + \left(1-\frac{1}{d}\right)\mathrm{e}^{-dt}; \\
\mathbb{E}[(g_{dk})^2] &= \frac{1}{d}\left(1-\mathrm{e}^{-dt}\right) \quad \forall k \neq d; \\
\mathbb{E}[g_{dk}g_{dl}] &= 0 \quad \forall k \neq l.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{cov}(X_i^{(d)},X_j^{(d)}) &= \frac{\mathrm{e}^{-\frac{(d-1)D}{D_n}\lambda} + (d-1)\mathrm{e}^{-(d-\frac{D}{D_n})\lambda} - d\mathrm{e}^{-(d-1)\lambda}}{d}; \\
\mathrm{cov}(X_i^{(k)},X_j^{(k)}) &= \frac{\mathrm{e}^{-\frac{(d-1)D}{D_n}\lambda} - \mathrm{e}^{-(d-\frac{D}{D_n})\lambda}}{d} \quad \forall k \neq d; \\
\mathrm{cov}(X_i^{(k)},X_j^{(l)}) &= 0 \quad \forall k \neq l.
\end{aligned}$$

These formulæ generalize indeed what has been obtained when $d = 2$. Then, setting $\mathbf{S}_n = \sum_{i=1}^{N_n}\mathbf{X}_i$, we obtain:

$$\frac{\mathrm{cov}(\mathbf{S}_n)}{(2D_n-1)\,N_n} = K + O\!\left(\frac{1}{D_n}\right),$$

where K is the diagonal matrix with

$$\begin{aligned}
K_{dd} &= \frac{1 + d(d-2-(d-1)\lambda)\,\mathrm{e}^{-(d-1)\lambda} - (d-1)^2\,\mathrm{e}^{-d\lambda}}{d(d-1)\lambda}; \\
K_{kk} &= \frac{1 - d\mathrm{e}^{-(d-1)\lambda} + (d-1)\mathrm{e}^{-d\lambda}}{d(d-1)\lambda} \quad \forall k \neq d.
\end{aligned}$$

One can apply Theorem 4.13 to the sum $\mathbf{S}_n = \sum_{i=1}^{N_n}\mathbf{X}_i$ if $D_n = o(N_n)$ and $(N_n)^{1/3} = O(D_n)$; then,

$$\mathbf{Y}_n = \frac{\mathbf{S}_n - N_n\,\mathrm{e}^{-\frac{(d-1)\lambda}{2}}\,\mathbf{e}_d}{\sqrt{(2D_n-1)\,N_n}}$$

converges in law towards $\mathcal{N}_{\mathbb{R}^d}(\mathbf{0}, K)$, with

$$d_{\mathrm{convex}}(\mathbf{Y}_n, \mathcal{N}_{\mathbb{R}^d}(\mathbf{0}, K)) = O\left(\sqrt{\frac{D_n}{N_n}}\right).$$

We have drawn in Figure 9 the projection of the two first coordinates of the random walk associated to the steps $\mathbf{X}_i$ when $d = 3$, in the case of dependent and independent random vectors. As in the previous example, these two random walks do not share the same aspect and fluctuations.

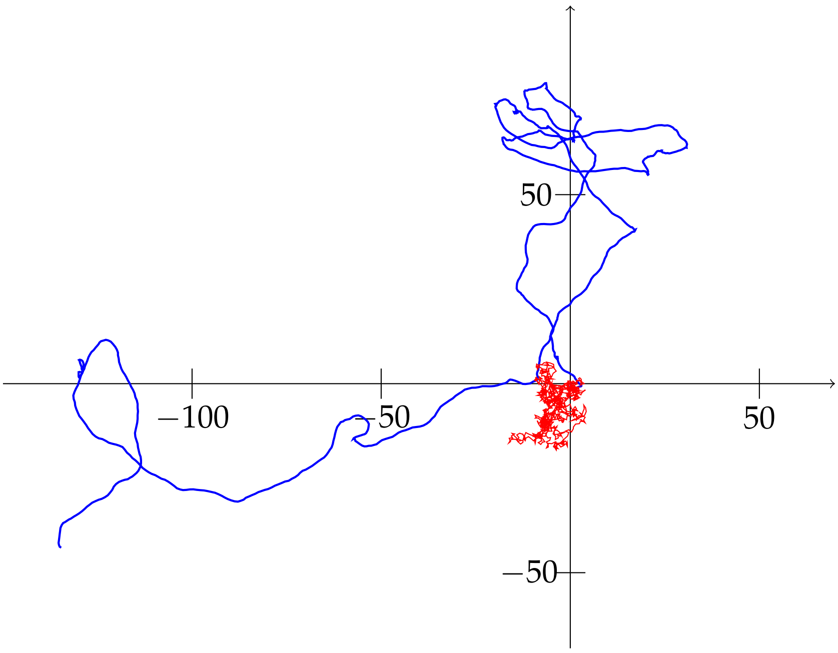

FIGURE 9. Projection of the two first coordinates of random walks associated to a sum of dependent random vectors in $\mathbb{R}^3$ (thick blue line, $N_n = 1000$ and $D_n = 40$), and to a sum of independent random vectors (thin red line, $N_n = 1000$ and $D_n = 1$).

Example 4.16 (Subgraph counts in random Erdös–Rényi graphs). Let us examine an example where the theory of sparse dependency graphs applies to a sum of random vectors, but where the limiting law of the rescaled sum is a degenerate Gaussian distribution. We fix a parameter $p \in (0, 1)$, and we consider the *random Erdös–Rényi graph* $G_n = G(n, p)$, which is the random subgraph of the complete graph on n vertices, such that the family of random variables

$$\left(1_{\{i,j\}\text{ is an edge of the graph}}\right)_{1\leq i<j\leq n}$$

is a family of independent Bernoulli random variables with same parameter p. We have drawn in Figure 10 such a random graph with parameters $n = 30$ and $p = 0.15$. In [FMN16, Section 10], it has been shown that if H is a fixed graph (motive), then the number of such motives H appearing in G_n converges after renormalisation in the mod-Gaussian sense; see also [FMN20] for a generalisation of this result to graph subcounts in graphon models. Theorem 4.12 will enable us to deal with several motives at once, and to understand the fluctuations of a random vector of subgraph counts.

We start by recalling the theory of subgraph counts in random graphs. If G is a (undirected, simple) graph, we denote V_G and E_G its sets of vertices and of edges. If H and G are

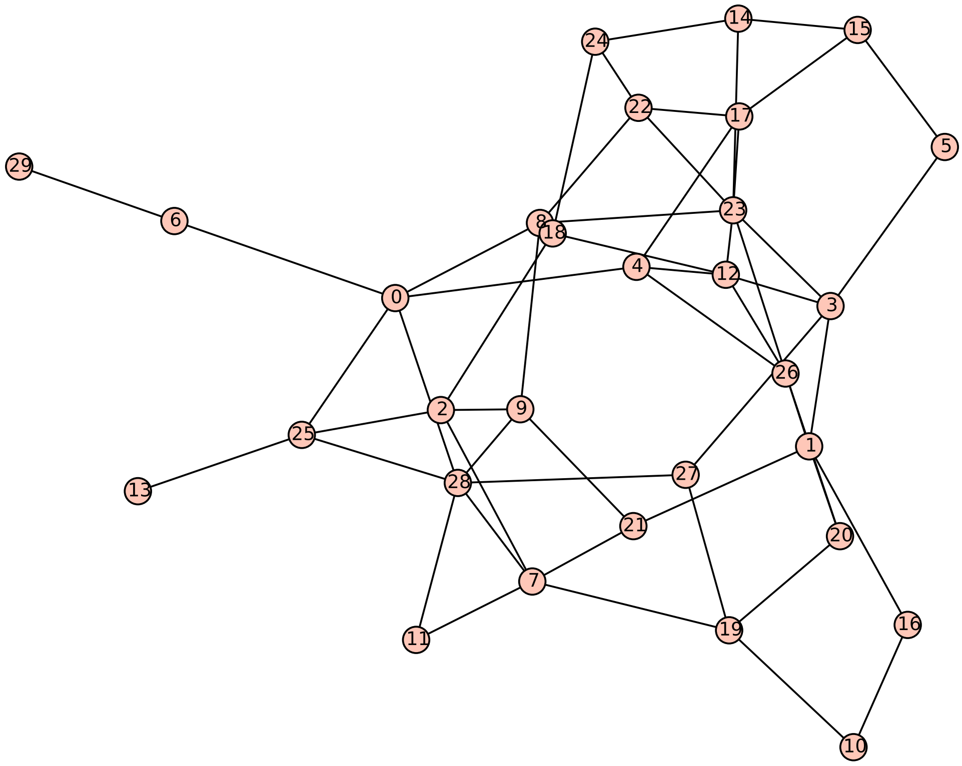

FIGURE 10. Random Erdös-Rényi graph with $n = 30$ vertices and parameter $p = 0.15$.

two graphs, the *number of embeddings* of H in G, denoted $I(H, G)$, is the number of injective maps $i : V_H \to V_G$ such that, if $\{v_1, v_2\} \in E_H$, then $\{i(v_1), i(v_2)\} \in E_G$. Fix a graph H with vertex set $[\![1, k]\!]$. An embedding of H into a graph G_n with vertex set $[\![1, n]\!]$ corresponds a choice of an arrangement $A = (a_1, \ldots, a_k)$ of k distinct elements of $[\![1, n]\!]$. Given such an arrangement A, we set

$$X_A(H, G) = \begin{cases} 1 & \text{if the map } j \mapsto a_j \text{ yields an embedding of } H \text{ into } G, \\ 0 & \text{otherwise.} \end{cases}$$

With the graph of Figure 10, $X_{(3,1,23)}(K_3, G) = 1$, whereas $X_{(13,25,28)}(K_3, G) = 0$. The number of embeddings of H into G is given by the sum

$$I(H, G) = \sum_{A \in \mathfrak{A}(n,k)} X_A(H, G),$$

where $\mathfrak{A}(n, k)$ is the set of all arrangements of size k in $[\![1, n]\!]$.

Suppose now that $G = G_n$ is a random Erdös-Rényi graph of parameters n and p. Then, assuming $k \leq n$, for every arrangement $A \in \mathfrak{A}(n, k)$, $X_A(H, G_n)$ is a random Bernoulli variable of expectation

$$\mathbb{E}[X_A(H, G_n)] = p^h, \quad \text{with } h = \operatorname{card} E_H.$$

Indeed, one can write each $X_A(H, G_n)$ as the product

$$X_A(H, G_n) = \prod_{\{i,j\} \in E_H} 1_{\{a_i, a_j\} \in E_{G_n}},$$

and by assumption on the random Erdös-Rényi graph, the random variables $1_{\{a_i,a_j\} \in E_{G_n}}$ are independent Bernoulli random variables of parameter p. The same decomposition shows that, if A and B are two arrangements that do not share at least 2 different points a and

b, then $X_A(H, G_n)$ and $X_B(H, G_n)$ can be written as products of variables falling into two independent families; therefore, they are independent. As a consequence, a dependency graph for the sum $I(H, G_n)$ is the graph:

- whose vertices are the arrangements $A \in \mathfrak{A}(n, k)$;
- with an edge between A and B if and only if $\mathrm{card}(A \cap B) \geq 2$.

The parameters of this dependency graph are the following. The number of arrangements in $\mathfrak{A}(n, k)$ is $N_n = n^{\downarrow k} = n(n-1)(n-2)\cdots(n-k+1) = n^k\,(1 + O(\frac{1}{n}))$. On the other hand, if an arrangement A is fixed, then there are less than

$$\binom{k}{2}^2 2\,(n-2)(n-3)\cdots(n-k+1)$$

arrangements B that share at least two points with A. So, one can take for parameter

$$D_n = 2\binom{k}{2}^2 (n-2)(n-3)\cdots(n-k+1) = \frac{k^2(k-1)^2}{2}\,n^{k-2}\left(1 + O\left(\frac{1}{n}\right)\right).$$

Fix now several graphs $H^{(1)}, H^{(2)}, \ldots, H^{(d)}$. Notice that when dealing with the random vector $I(\mathbf{H}, G_n) = (I(H^{(1)}, G_n), I(H^{(2)}, G_n), \ldots, I(H^{(d)}, G_n))$, one can always assume without loss of generality that $k = \mathrm{card}\, V_{H^{(1)}} = \mathrm{card}\, V_{H^{(2)}} = \cdots = \mathrm{card}\, V_{H^{(d)}}$. Indeed, suppose that V_H has cardinality $j \leq k$, and denote H^{+k} the graph obtained from H by adding $k - j$ disconnected vertices. Then, for $n \geq k$,

$$I(H^{+k}, G_n) = (n-j)\cdots(n-k+1)\,I(H, G_n),$$

so the study of the fluctuations of $I(H, G_n)$ is equivalent to the study of the fluctuations of $I(H^{+k}, G_n)$. Setting $X_A(\mathbf{H}, G_n) = (X_A(H^{(1)}, G_n), \ldots, X_A(H^{(d)}, G_n))$, one can then write the random vector $I(\mathbf{H}, G_n)$ as a sum of random vectors:

$$I(\mathbf{H}, G_n) = \sum_{A \in \mathfrak{A}(n,k)} X_A(\mathbf{H}, G_n).$$

Moreover, the random vectors $X_A(\mathbf{H}, G_n)$ take their values in $[0,1]^d$, and they have the same dependency graph as described before. Consequently, the conclusions of Theorem 4.12 hold:

$$\left|\kappa(I(H^{(i_1)}, G_n), I(H^{(i_2)}, G_n), \ldots, I(H^{(i_r)}, G_n))\right| \leq r^{r-2}\,k^{4(r-1)}\,n^{k+(r-1)(k-2)}$$

for any choice of indices $i_1, \ldots, i_r$, since $N_n \leq n^k$ and $D_n \leq \frac{k^4}{2}\,n^{k-2}$. Unfortunately, one cannot apply Theorem 4.13 to the recentred sum $\mathbf{S}_n = I(\mathbf{H}, G_n) - \mathbb{E}[I(\mathbf{H}, G_n)]$, as the limiting covariance matrix of $\frac{\mathbf{S}_n}{n^{k-1}}$ is a *degenerate* non-negative symmetric matrix. Indeed, given two arrangements A and B, one has

$$\begin{aligned}&\mathrm{cov}(X_A(H^{(1)}, G_n), X_B(H^{(2)}, G_n))\\ &= \mathbb{E}\left[\prod_{\{i,j\} \in E_{H^{(1)}}} 1_{\{a_i, a_j\} \in E_{G_n}} \prod_{\{i,j\} \in E_{H^{(2)}}} 1_{\{b_i, b_j\} \in E_{G_n}}\right] - p^{h_1 + h_2},\end{aligned}$$

where h_1 and h_2 are the numbers of edges of $H^{(1)}$ and $H^{(2)}$. If A and B do not share at least two points, one obtains 0 as explained before. Suppose now that A and B share exactly two points: there are indices $\{f, g\}$ and $\{i, j\}$ such that $a_f = b_i$ and $a_g = b_j$, and all the other a's are different from the other b's.

(1) If moreover $\{f,g\}$ and $\{i,j\}$ are edges respectively of $H^{(1)}$ and $H^{(2)}$, then the expectation of the product of Bernoulli variables is equal to $p^{h_1+h_2-1}$, since $1_{\{a_f,a_g\}\in E_{G_n}} = 1_{\{b_i,b_j\}\in E_{G_n}}$. Thus,

$$\mathrm{cov}(X_A(H^{(1)},G_n),X_B(H^{(2)},G_n)) = p^{h_1+h_2-1} - p^{h_1+h_2}$$

if $A\cap B = \{a_f,a_g\} = \{b_i,b_j\}$ has cardinality equal to two, and if this intersection comes from an edge of $H^{(1)}$ and an edge of $H^{(2)}$.

(2) On the other hand, if the intersection of cardinality two $A\cap B$ does not come from edges of $H^{(1)}$ and $H^{(2)}$, then the covariance is again equal to 0.

The number of pair of arrangements (A,B) that fulfills the first hypothesis is equal to $2h_1h_2n^{\downarrow 2k-2}$. Thus, the pairs of arrangements (A,B) with an intersection of cardinality 2 yield a contribution

$$2(p^{-1}-1)\,h_1p^{h_1}\,h_2p^{h_2}\,n^{2k-2}\,\left(1+O\left(\frac{1}{n}\right)\right)$$

to $\mathrm{cov}(I(H^{(1)},G_n),I(H^{(2)},G_n))$. The other pairs of arrangements have an intersection of cardinality larger than 3, hence, they yield a contribution which is a $O(n^{2k-3})$. We conclude that

$$\mathrm{cov}(I(H^{(i)},G_n),I(H^{(j)},G_n)) = 2(p^{-1}-1)\,h_ip^{h_i}\,h_jp^{h_j}\,n^{2k-2}\,\left(1+O\left(\frac{1}{n}\right)\right);$$
$$\mathrm{cov}\left(\frac{\mathbf{S}_n}{n^{k-1}}\right) = 2(p^{-1}-1)\mathbf{w}\mathbf{w}^t + O\left(\frac{1}{n}\right),$$

where $\mathbf{w} = (h_1p^{h_1},\ldots,h_dp^{h_d})$. Thus, the limiting covariance is always a rank one symmetric matrix, and one cannot apply Theorem 4.13. These one-dimensional fluctuations are not very surprising: each fluctuation of a random number of embeddings $I(H^{(i)},G_n)$ is driven at first order by the fluctuations of the number of edges in G_n, and truly multi-dimensional fluctuations only occur at higher order. In this setting, one can obtain an estimate of the speed of the convergence by projecting the random variable $\mathbf{S}_n$ to the vector line on which the Gaussian distribution with covariance matrix $\mathbf{w}\mathbf{w}^t$ is supported. Denote $S_n^{(w)} = \frac{\mathbf{w}^t\mathbf{S}_n}{\mathbf{w}^t\mathbf{w}}$, and consider the L^2-norm of $\mathbf{S}_n - S_n^{(w)}\mathbf{w}$:

$$\begin{aligned}\mathbb{E}\left[\left\|\mathbf{S}_n - S_n^{(w)}\mathbf{w}\right\|^2\right] &= \mathbb{E}\left[\|\mathbf{S}_n\|^2 - \frac{(\mathbf{w}^t\mathbf{S}_n)^2}{\|\mathbf{w}\|^2}\right]\\ &= \sum_{i=1}^d \mathbb{E}[(S_n^{(i)})^2] - \frac{1}{\|\mathbf{w}\|^2}\sum_{i=1}^d\sum_{j=1}^d w^{(i)}w^{(j)}\,\mathbb{E}[S_n^{(i)}S_n^{(j)}] = O(n^{2k-3})\end{aligned}$$

by using the estimate $\mathbb{E}[S_n^{(i)}S_n^{(j)}] = 2(p^{-1}-1)\,w^{(i)}w^{(j)}\,n^{2k-2} + O(n^{2k-3})$. Therefore,

$$\mathbb{E}\left[\left\|\frac{\mathbf{S}_n - S_n^{(w)}\mathbf{w}}{n^{k-1}}\right\|^2\right] = O\left(\frac{1}{n}\right),$$

so in particular $\frac{\mathbf{S}_n - S_n^{(w)}\mathbf{w}}{n^{k-1}}$ converges in probability to 0. Then, the projection $S_n^{(w)}$ can be shown to converge in the one-dimensional mod-Gaussian sense:

$$\begin{aligned}
&\log \mathbb{E}\left[\mathrm{e}^{\frac{z\,S_n^{(w)}}{n^{k-1}}}\right] = \log \mathbb{E}\left[\mathrm{e}^{\frac{\langle z\mathbf{w}\,|\,\mathbf{S}_n\rangle}{n^{k-1}\,\langle \mathbf{w}\,|\,\mathbf{w}\rangle}}\right] \\
&= \sum_{r\geq 2} \frac{z^r}{\|w\|^{2r}\, r!\, n^{(k-1)r}} \sum_{i_1,\ldots,i_r=1}^{d} w^{(i_1)}\cdots w^{(i_r)}\,\kappa(S_n^{(i_1)},\ldots,S_n^{(i_r)}) \\
&= z^2(p^{-1}-1) + O\left(\frac{|z|^2}{n}\right) + \sum_{r\geq 3} \frac{z^r}{\|\mathbf{w}\|^{2r}\, r!\, n^{(k-1)r}} \sum_{i_1,\ldots,i_r=1}^{d} w^{(i_1)}\cdots w^{(i_r)}\,\kappa(S_n^{(i_1)},\ldots,S_n^{(i_r)}).
\end{aligned}$$

Using the bounds on joint cumulants previously described, we can write the following upper bound on the remainder of the series $R(z)$:

$$\begin{aligned}
|w^{(i_1)}\cdots w^{(i_r)}\,\kappa(S_n^{(i_1)},\ldots,S_n^{(i_r)})| &\leq |w^{(i_1)}\cdots w^{(i_r)}|\, r^{r-2}\, k^{4(r-1)}\, n^{k+(r-1)(k-2)}; \\
\sum_{i_1,\ldots,i_r=1}^{d} |w^{(i_1)}\cdots w^{(i_r)}\,\kappa(S_n^{(i_1)},\ldots,S_n^{(i_r)})| &\leq (\|\mathbf{w}\|_1)^r\, r^{r-2}\, k^{4(r-1)}\, n^{k+(r-1)(k-2)}; \\
|R(z)| &\leq \frac{n^2}{k^4} \sum_{r\geq 3} \frac{r^{r-2}}{r!} \left(\frac{|z|\,\|\mathbf{w}\|_1\, k^4}{n\,\|\mathbf{w}\|^2}\right)^r,
\end{aligned}$$

and the last term is a $O(|z|^3/n)$ on a zone $z \in [-Dn, Dn]$ with $D > 0$ fixed (depending only on k and on $\|w\|$). Therefore, if μ_n is the law of $S_n^{(w)}/(n^{k-1}\sqrt{2(p^{-1}-1)})$ and ν is the standard one-dimensional Gaussian distribution, then

$$\widehat{\mu_n}(\zeta) - \widehat{\nu}(\zeta) = \mathrm{e}^{-\frac{\zeta^2}{2}} \left(\exp\left(O\left(\frac{|\zeta|^2 + |\zeta|^3}{n}\right)\right) - 1\right)$$

for $|\zeta| \leq Dn$ and a certain constant $D > 0$. The arguments of [FMN19, Section 2], which give estimates of the speed of convergence in the special case $d = 1$, ensure then that

$$d_{\mathrm{Kol}}\left(\frac{S_n^{(w)}}{n^{k-1}\sqrt{2(p^{-1}-1)}}, \mathcal{N}_{\mathbb{R}}(0,1)\right) = O\left(\frac{1}{n}\right).$$

Thus, if $\mathbf{Y}_n = (Y_n^{(1)},\ldots,Y_n^{(d)})$ with $Y_n^{(i)} = \frac{I(H^{(i)},G_n) - n^{\downarrow k} p^{h_i}}{\sqrt{2(p^{-1}-1)}\, n^{k-1}}$, and if $Y_n^{(w)} = \frac{\langle \mathbf{Y}_n\,|\,\mathbf{w}\rangle}{\langle \mathbf{w}\,|\,\mathbf{w}\rangle}$ then:

(1) As n goes to infinity, the difference between $\mathbf{Y}_n$ and its projection $Y_n^{(w)}\mathbf{w}$ converges in probability to 0, and more precisely,

$$\mathbb{E}[\|\mathbf{Y}_n - Y_n^{(w)}\mathbf{w}\|^2] = O\left(\frac{1}{n}\right).$$

(2) As n goes to infinity, the random variable $Y_n^{(w)}$ converges in law towards a standard Gaussian distribution, and more precisely,

$$d_{\mathrm{Kol}}(Y_n^{(w)}, \mathcal{N}_{\mathbb{R}}(0,1)) = O\left(\frac{1}{n}\right).$$

Thus, when the limiting covariance is not a positive-definite matrix, one can usually still apply the theory of mod-Gaussian convergence to some projection of the random vectors, and on the other hand control the difference between the projections and the original random vectors.

4.4. **Empirical measures of Markov chains.** The interest of Definition 4.5 is that one can use it with sums of random vectors $\mathbf{S}_n = \sum_{i=1}^{N_n} \mathbf{A}_{i,n}$ even when there is no sparse dependency graph of parameters N_n and D_n for the random vectors $\mathbf{A}_{i,n}$ of the sums. In [FMN19, Section 5], it was proven that the bounds on cumulants corresponding to Hypothesis (MC3) are also true for:

- certain observables of models from statisticial mechanics;
- the linear functionals of finite ergodic Markov chains.

In this section, we explain how to use the theory of [FMN19, Sections 5.4 and 5.5] to study the fluctuations of the empirical measure of a finite ergodic Markov chain (instead of a linear form of this empirical measure). We fix a state space $\mathfrak{X} = [\![1, d]\!]$, and we consider a transition matrix P on $\mathfrak{X}$, which is assumed ergodic (irreducible and aperiodic), with invariant probability measure $\boldsymbol{\pi} = (\pi(1), \pi(2), \ldots, \pi(d))$. Thus, $\pi P = \pi$, and if $(X_n)_{n\in\mathbb{N}}$ is a Markov chain on $\mathfrak{X}$ with transition matrix P and initial distribution $\mathbb{P}[X_0 = i] = \pi(i)$, then $\mathbb{P}[X_n = i] = \pi(i)$ for any $n \in \mathbb{N}$. We are interested in the *empirical measure* of $(X_n)_{n\in\mathbb{N}}$, which is the random vector in $\mathbb{R}^d$

$$\boldsymbol{\pi}_n = \frac{1}{n}\left(\sum_{t=1}^{n} \delta_{X_t}\right).$$

The ergodic theorem ensures that $\boldsymbol{\pi}_n \to \boldsymbol{\pi}$ almost surely. Set $\mathbf{S}_n = n(\boldsymbol{\pi}_n - \boldsymbol{\pi})$. We define a constant

$$\theta_P = \sqrt{\max\{|z| \,|\, z \neq 1,\ z \text{ eigenvalue of } PD^{-1}P^tD\}},$$

where $D = \mathrm{diag}(\boldsymbol{\pi})$. This constant θ_P is strictly smaller than 1, and when P is reversible it is just the module of the second largest eigenvalue of P. It has been proven [FMN19] that

$$\left|\kappa(1_{X_{t_1}=i_1}, \ldots, 1_{X_{t_r}=i_r})\right| \leq 2^{r-1} \sum_{T \in \mathrm{ST}([\![1,r]\!])} w(T),$$

where the weight of a spanning tree on r vertices is the products of the weights $w(i,j) = (\theta_P)^{|t_j - t_i|}$ of its edges $(i,j) \in E_T$. As a consequence, for any $r \geq 2$ and any choice of indices $i_1, \ldots, i_r$,

$$\left|\kappa(S_n^{(i_1)}, \ldots, S_n^{(i_r)})\right| \leq 2^{r-1} \sum_{T \in \mathrm{ST}([\![1,r]\!])} \sum_{t_1,\ldots,t_r=1}^{n} \prod_{(i,j)\in E_T} (\theta_P)^{|t_j - t_i|} \leq n\, r^{r-2} \left(2\,\frac{1+\theta_P}{1-\theta_P}\right)^{r-1},$$

see [FMN19, Theorem 55]. Therefore, Hypothesis (MC3) is satisfied with $N_n = n$, $D_n = \frac{1+\theta_P}{1-\theta_P}$ and $A = 1$. Let us then check the other hypotheses of Definition 4.5:

(1) By construction, $\mathbb{E}[\mathbf{S}_n] = 0$ (here and in the sequel we assume that $(X_n)_{n\in\mathbb{N}}$ is stationary, hence has initial distribution $\boldsymbol{\pi}$).

(2) We compute

$$\begin{aligned}\mathrm{cov}(S_n^{(i)}, S_n^{(j)}) &= \sum_{a,b=1}^{n} \mathbb{P}[X_a = i \text{ and } X_b = j] - \mathbb{P}[X_a = i]\,\mathbb{P}[X_b = j] \\ &= n\,\delta_{i,j}\,\pi(i)(1-\pi(i)) + 2\sum_{k=1}^{n-1}(n-k)\pi(i)(P^k(i,j) - \pi(j)).\end{aligned}$$

The limit of $\frac{\mathrm{cov}(S_n^{(i)}, S_n^{(j)})}{n}$ is

$$K_{i,j} = \pi(i)(\delta_{i,j} - \pi(j)) + 2\pi(i)\sum_{k=1}^{\infty} R^k(i,j),$$

where $R = P - P^\infty$; $P^\infty(i,j) = \pi(j)$; and R is a matrix with all its eigenvalues strictly smaller than 1. Notice that K has at most rank $(d-1)$: indeed, $\sum_{i=1}^d K_{i,j} = 0$, this being a consequence of

$$\sum_{i=1}^{d} S_n^{(i)} = 0 \quad \text{almost surely.}$$

Therefore, we shall consider the random vectors $\mathbf{S}_n$ as elements of the $(d-1)$-dimensional space

$$H = \left\{ \mathbf{x} \in \mathbb{R}^d \;\middle|\; \sum_{i=1}^{d} x^{(i)} = 0 \right\},$$

and the covariance matrix as a symmetric operator on this space. By the discussion of [FMN19, Proposition 59], if the Markov chain is reversible, then K has indeed rank $(d-1)$ and we shall be able to describe precisely the fluctuations of $(\mathbf{S}_n)_{n\in\mathbb{N}}$ in H.

Notice that the difference between $\frac{\mathrm{cov}(\mathbf{S}_n)}{n}$ and K is a $O(\frac{1}{n})$; therefore, Condition (MC2′) is satisfied.

(3) Finally, Hypotheses (MC4) and (MC5) are discussed at the very end of [FMN19], and they are satisfied with $v = 3$.

We conclude:

Theorem 4.17 (Empirical measures of reversible Markov chains). *Let $(X_n)_{n\in\mathbb{N}}$ be an ergodic and reversible Markov chain on a finite state space $\mathfrak{X} = [\![1,d]\!]$. We denote $\boldsymbol{\pi}$ the stationary measure and $\boldsymbol{\pi}_n$ the empirical measure; their difference is a random vector in the hyperplane H. If*

$$\mathbf{X}_n = \frac{\boldsymbol{\pi}_n - \boldsymbol{\pi}}{n^{1/3}} \qquad ; \qquad \mathbf{Y}_n = \frac{\boldsymbol{\pi}_n - \boldsymbol{\pi}}{\sqrt{n}},$$

and

$$K_{i,j} = \pi(i)(\delta_{i,j} - \pi(j)) + 2\pi(i) \sum_{k=1}^{\infty} (P^k(i,j) - \pi(j)),$$

then:

(1) The sequence $(\mathbf{X}_n)_{n\in\mathbb{N}}$ is mod-Gaussian convergent on H, with parameters $n^{1/3}K$ and limit

$$\psi(\mathbf{z}) = \exp\left(\frac{1}{6} \sum_{i,j,k=1}^{d} \left(\sum_{a,b\in\mathbb{Z}} \kappa(1_{X_0} = i, 1_{X_a} = j, 1_{X_b} = k) \right) z^{(i)} z^{(j)} z^{(k)} \right).$$

(2) We have a convergence in law $\mathbf{Y}_n \rightharpoonup \mathcal{N}_H(\mathbf{0}, K)$, with a convex distance between the two distributions on H that is a $O(\frac{1}{\sqrt{n}})$. The constant in the $O(\cdot)$ only depends on d and on the constant θ_P previously introduced.

REFERENCES

[AH10] K. Atkinson and W. Han. *Spherical Harmonics and Approximations on the Unit Sphere: An Introduction*. Vol. 2044. Lecture Notes in Mathematics. Springer-Verlag, 2010.

[Ben19] C. G. Beneš. "A local central limit theorem and loss of rotational symmetry of planar simple random walk". In: *Trans. Amer. Math. Soc.* 371 (2019), pp. 2553–2573.

[Ber41] A. C. Berry. "The Accuracy of the Gaussian Approximation to the Sum of Independent Variates". In: *Trans. Amer. Math. Soc.* 49.1 (1941), pp. 122–136.

[BR10] R. N. Bhattacharya and R. R. Rao. *Normal Approximation and Asymptotic Expansions*. Vol. 64. Classics in Applied Mathematics. SIAM, 2010.

[Bil99] P. Billingsley. *Convergence of probability measures*. 2nd. Wiley Series in Probability and Statistics. Wiley, 1999.

[BMN19] M. Dal Borgo, P.-L. Méliot, and A. Nikeghbali. "Local limit theorems and mod-ϕ convergence". In: *Latin American Journal of Probability and Mathematical Statistics* 16.1 (2019), pp. 817–853.

[Chh+20] R. Chhaibi, F. Delbaen, P.-L. Méliot, and A. Nikeghbali. "Mod-ϕ convergence: Approximation of discrete measures and harmonic analysis on the torus". In: *Ann. Institut Fourier* 70.3 (2020), pp. 1115–1197.

[Cra38] H. Cramér. "Sur un nouveau théorème-limite de la théorie des probabilités". In: *Actualités Sci. Indust.* 736 (1938), pp. 5–23.

[DKN15] F. Delbaen, E. Kowalski, and A. Nikeghbali. "Mod-ϕ convergence". In: *Int. Math. Res. Notices* 2015 (2015), pp. 3445–3485.

[DZ98] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. 2nd. Vol. 38. Stochastic Modelling and Applied Probability. Springer-Verlag, 1998.

[Ess45] C.-G. Esseen. "Fourier analysis of distribution functions. A mathematical study of the Laplace-Gauss law". In: *Acta Math.* 77.1 (1945), pp. 1–125.

[Fel71] W. Feller. *An Introduction to Probability Theory and Its Applications, Volume II*. 2nd. Wiley Series in Probability and Statistics. Wiley, 1971.

[FMN16] V. Féray, P.-L. Méliot, and A. Nikeghbali. *Mod-ϕ convergence: Normality zones and precise deviations*. SpringerBriefs in Probability and Mathematical Statistics. Springer-Verlag, 2016.

[FMN19] V. Féray, P.-L. Méliot, and A. Nikeghbali. "Mod-ϕ convergence, II: Estimates on the speed of convergence". In: *Séminaire de Probabilités L*. Ed. by C. Donati-Martin, A. Lejay, and A. Rouault. Vol. 2252. Lecture Notes in Mathematics. 2019, pp. 405–478.

[FMN20] V. Féray, P.-L. Méliot, and A. Nikeghbali. "Graphons, permutons and the Thoma simplex: three mod-Gaussian moduli spaces". In: *Proceedings of the London Mathematical Society* 121.4 (2020), pp. 876–926.

[JKN11] J. Jacod, E. Kowalski, and A. Nikeghbali. "Mod-Gaussian convergence: new limit theorems in Probability and Number Theory". In: *Forum Mathematicum* 23.4 (2011), pp. 835–873.

[KS00] J. P. Keating and N. C. Snaith. "Random Matrix Theory and $\zeta(1/2+it)$". In: *Commun. Math. Phys.* 214 (2000), pp. 57–89.

[KN12] E. Kowalski and A. Nikeghbali. "Mod-Gaussian convergence and the value distribution of $\zeta(1/2+it)$ and related quantities". In: *Journal of the London Mathematical Society (2)* 86.1 (2012), pp. 291–319.

[LS59] V. P. Leonov and A. N. Shiryaev. "On a method of calculation of semi-invariants". In: *Theory Prob. Appl.* 4 (1959), pp. 319–329.

[Mél14] P.-L. Méliot. "The cut-off phenomenon for Brownian motions on compact symmetric spaces". In: *Potential Analysis* 40.4 (2014), pp. 427–509.

[MN22] P.-L. Méliot and A. Nikeghbali. *On the precise deviations of the characteristic polynomial of a random matrix*. 2022. URL: https://arxiv.org/abs/2202.08336.

[Sat99] K.-I. Sato. *Lévy processes and infinitely divisible distributions*. Vol. 68. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1999.

[Spi65] M. Spivak. *Calculus on manifolds*. Addison-Wesley, 1965.

[Tao11] T. Tao. *An Introduction to Measure Theory*. Vol. 126. Graduate Studies in Mathematics. American Mathematical Society, 2011.

[Vaa00] A. W. van der Vaart. *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2000.

[VV09] V. V. Volchkov and V. V. Volchkov. *Harmonic Analysis of Mean Periodic Functions on Symmetric Spaces and the Heisenberg Group*. Springer Monographs in Mathematics. Springer-Verlag, 2009.

[Zor04] V. A. Zorich. *Mathematical Analysis II*. Universitext. Springer-Verlag, 2004.