

## Geometric sharp large deviations for random projections of $\ell_p^n$ spheres and balls\*

Yin-Ting Liao<sup>†</sup>      Kavita Ramanan<sup>‡</sup>

### Abstract

Accurate estimation of tail probabilities of projections of high-dimensional probability measures is of relevance in high-dimensional statistics and asymptotic geometric analysis. Whereas large deviation principles identify the asymptotic exponential decay rate of probabilities, sharp large deviation estimates also provide the “prefactor” in front of the exponentially decaying term. For fixed  $p \in (1, \infty)$ , consider independent sequences  $(X^{(n,p)})_{n \in \mathbb{N}}$  and  $(\Theta^n)_{n \in \mathbb{N}}$  of random vectors with  $\Theta^n$  distributed according to the normalized cone measure on the unit  $\ell_2^n$  sphere, and  $X^{(n,p)}$  distributed according to the normalized cone measure on the unit  $\ell_p^n$  sphere. For almost every realization  $(\theta^n)_{n \in \mathbb{N}}$  of  $(\Theta^n)_{n \in \mathbb{N}}$ , (quenched) sharp large deviation estimates are established for suitably normalized (scalar) projections of  $X^{(n,p)}$  onto  $\theta^n$ , that are asymptotically exact (as the dimension  $n$  tends to infinity). Furthermore, the case when  $(X^{(n,p)})_{n \in \mathbb{N}}$  is replaced with  $(\mathcal{X}^{(n,p)})_{n \in \mathbb{N}}$ , where  $\mathcal{X}^{(n,p)}$  is distributed according to the uniform (or normalized volume) measure on the unit  $\ell_p^n$  ball, is also considered. In both cases, in contrast to the (quenched) large deviation rate function, the prefactor exhibits a dependence on the projection directions  $(\theta^n)_{n \in \mathbb{N}}$  that encodes additional geometric information that enables one to distinguish between projections of balls and spheres. Moreover, comparison with numerical estimates obtained by direct computation and importance sampling shows that the obtained analytical expressions for tail probabilities provide good approximations even for moderate values of  $n$ . The results on the one hand provide more accurate quantitative estimates of tail probabilities of random projections of  $\ell_p^n$  spheres than logarithmic asymptotics, and on the other hand, generalize classical sharp large deviation estimates in the spirit of Bahadur and Ranga Rao to a geometric setting. The proofs combine Fourier analytic and probabilistic techniques. Along the way, several results of independent interest are obtained including a simpler representation for the quenched large deviation rate function that shows that it is strictly convex, a central limit theorem for random projections

---

\*The first author was supported by NSF Grant DMS-1954351 and a GSAA fellowship from the Taiwan Government. The second author was supported by the National Science Foundation under grant DMS-1713032 and by the Office of Naval Research under the Vannevar Bush Faculty Fellowship N000142112887.

<sup>†</sup>Brown University. E-mail: [yin-ting\\_liao@alumni.brown.edu](mailto:yin-ting_liao@alumni.brown.edu)

<sup>‡</sup>Brown University. E-mail: [kavita\\_ramanan@brown.edu](mailto:kavita_ramanan@brown.edu)

under a certain family of tilted measures, and multi-dimensional generalized Laplace asymptotics.

**Keywords:** sharp large deviations; random projections; rate function; prefactor; Bahadur-Rao correction;  $\ell_p^n$  spheres; cone measure; spherical caps; central limit theorem for convex sets.

**MSC2020 subject classifications:** 60F10; 52A23; 46B06; 41A60.

Submitted to EJP on August 27, 2021, final version accepted on September 13, 2023.

## 1 Introduction

### 1.1 Motivation and context

The study of high-dimensional norms, the convex bodies that describe their level sets, and other high-dimensional geometric structures are central themes in geometric functional analysis [29], and the burgeoning field of asymptotic geometric analysis [3]. Several results in these fields have shown that the presence of high dimensions often imposes a certain regularity that has a probabilistic flavor. A significant result of this type is the central limit theorem (CLT) for convex sets [26] which, roughly speaking, says that if  $X^n$  is a high-dimensional random vector uniformly distributed on an isotropic convex body (namely, a compact convex set with non-empty interior whose normalized volume measure has zero mean and identity covariance matrix), its one-dimensional scalar projections  $\langle X^n, \theta^n \rangle$  along most directions  $\theta^n$  on the unit  $(n-1)$ -dimensional sphere  $S^{n-1}$  in  $\mathbb{R}^n$  have Gaussian fluctuations. In fact, this result holds for the larger class of isotropic logconcave measures as well as more general high-dimensional measures that satisfy a certain concentration estimates called the thin shell condition (see, e.g. [37, 40, 28]). Of particular interest is the geometry of  $\ell_p^n$  spaces, which has been classically studied using laws of large numbers, CLTs and concentration results [7, 17, 35, 36]. These constitute beautiful universality results that suggest that random projections of the uniform measure on a convex body behave in some aspects like sums of independent random variables. On the other hand, they also imply the somewhat negative conclusion that typical fluctuations of lower-dimensional random projections do not yield much information about high-dimensional measures. It is therefore natural to ask whether such random projections also satisfy other properties exhibited by sums of independent random variables, in particular those that capture non-universal features that would yield useful information about the corresponding high-dimensional measures.

With this objective, large deviation principles (LDP) were established for suitably normalized one-dimensional random projections of  $\ell_p^n$  balls in [15, 16]. These works established both quenched LDPs, conditioned on the sequence of projection directions, as well as annealed LDPs, which average over the randomness of the projection directions. Subsequently, quenched LDPs for multi-dimensional projections were obtained in [22], and annealed large deviation results for norms of  $\ell_p^n$  balls and their multidimensional random projections were established in [1, 20, 19, 23], with [19] also considering moderate deviations (see also [33] for a recent survey). Going beyond the setting of  $\ell_p^n$  balls (and measures with a similar representation), annealed LDPs were obtained for norms of multidimensional projections of more general sequences of high-dimensional random vectors  $(X^n)_{n \in \mathbb{N}}$  that satisfy a so-called asymptotic thin shell condition in [22, 23]. All these LDPs are indeed non-universal, in that the associated speeds (or exponential decay rates) and rate functions (that also captures the exponent) both encode properties of the high-dimensional measures. However, although LDPs (in contrast to concentration results or large deviation upper bounds) identify the precise asymptotic exponential decay rate and allow for the identification of conditional limit laws [24], they have the drawback that in general they only provide approximate estimates of the probabilities, characterizing only the limit of the logarithms of the deviation probabilities, as the

dimension  $n$  goes to infinity. Thus, existing LDPs for random projections cannot be applied directly to provide accurate estimates of tail probabilities or develop efficient algorithms that distinguish between two given high-dimensional measures, tasks that are of importance in statistics, data analysis and computer science [11].

### 1.2 Discussion of results

Our broad goal is to establish sharp (quenched) large deviation results of high-dimensional measures that not only capture the precise asymptotic exponential decay rate of tail probabilities of random projections, but also their “prefactors” (or the terms in front of the exponential), so as to provide more accurate quantitative estimates in finite dimensions, much in the spirit of the local theory of Banach spaces. In addition, we aim to identify additional geometric information that sharp large deviation estimates provide over LDPs. In this article, we focus on one-dimensional projections of  $\ell_p^n$  spheres and balls and obtain estimates of deviation probabilities that are asymptotically exact as the dimension goes to infinity.

It is worthwhile to mention that for the Euclidean norm of a random vector distributed on an isotropic convex body, sharp large deviation upper bounds were obtained in several works (see, for example, [26, 13, 32, 17] and references therein). While these estimates have the very nice feature that they are universal (in that they apply for all isotropic convex bodies or, more generally, logconcave measures), that very feature also makes them not tight for many specific sub-classes of convex bodies. As a consequence, our proof techniques are different from those used in the latter works, and may be of independent interest. In addition, we develop and analyze importance sampling algorithms to compute geometric quantities such as the volume fraction of small  $\ell_p^n$  spherical caps in a certain direction, which would be infeasible to compute with reasonable accuracy using standard Monte Carlo estimation since the quantities are vanishingly small. We expect that such computational approaches based on large deviations may be useful more generally in the study of high-dimensional geometric structures. Indeed, the first version of this article has already spurred further work in this direction. For example, Kaufmann [21] studied annealed (i.e., averaged over the randomness of  $\Theta$ ) sharp large deviation estimates for  $q$ -norms of random vectors uniformly distributed on  $\ell_p^n$  balls, and the paper [25] establishes quenched large deviation estimates for multi-dimensional projections of  $\ell_p^n$  balls and their norms.

We now describe some of the challenges in obtaining such sharp estimates and comment on our proof technique. Our results can be viewed as a geometric generalization of classical sharp large deviation estimates in the spirit of Bahadur and Ranga Rao [4], which we now briefly recall. Given a sequence of independent and identically distributed (i.i.d.) random variables  $(X_i)_{i \in \mathbb{N}}$ , for each  $n \in \mathbb{N}$ , let  $S^n$  denote the corresponding empirical mean:

$$S^n := \frac{1}{n} \sum_{i=1}^n X_i^n = \frac{1}{\sqrt{n}} \langle X^n, \mathcal{J}^n \rangle, \tag{1.1}$$

where  $X^n := (X_1, \dots, X_n)$  and  $\mathcal{J}^n := \frac{1}{\sqrt{n}}(1, 1, \dots, 1) \in \mathbb{S}^{n-1}$ . Under suitable assumptions on the (marginal) distribution of  $X_1$  it was shown in [4] that

$$\mathbb{P}(S^n \geq a) = \frac{e^{-n\mathbb{I}(a)}}{\bar{\sigma}_a \tau_a \sqrt{2\pi n}} (1 + o(1)), \tag{1.2}$$

where  $\mathbb{I}$  is the Legendre transform of  $\Lambda$ , the logarithmic moment generating function of  $X_1$ ,  $\tau_a > 0$  and  $\bar{\sigma}_a > 0$  are suitable constants specified below and  $o(1)$  indicates a term  $\varepsilon_n$  that satisfies  $\varepsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ . Key ingredients of the proof in [4] include first identifying a “tilted” measure under which the rare event on the left-hand side of (1.2) becomes

typical, and second, establishing a quantitative CLT for the sequence  $(S^n)_{n \in \mathbb{N}}$  under the tilted measure. Specifically, this tilted measure is also another product measure of the form  $\otimes \tilde{\mathbb{P}}_a$ , where  $\tilde{\mathbb{P}}_a$  is a measure absolutely continuous with respect to  $\mathbb{P}$ , with Radon-Nikodym derivative given by

$$\frac{d\tilde{\mathbb{P}}_a}{d\mathbb{P}}(x) := e^{\tau_a x - \Lambda(\tau_a)},$$

where  $\tau_a$  is the unique positive constant such that  $X_1$  has mean  $a$  under the marginal  $\tilde{\mathbb{P}}_a$  of the tilted measure. The constant  $\bar{\sigma}_a^2$  in (1.2) is the variance of  $X_1$  under  $\tilde{\mathbb{P}}_a$ . The second step of establishing a quantitative CLT is in this case standard given the product form of the tilted measure, and appeals to well known Edgeworth expansions that also involve the third moment of  $S_n$  under the tilted measure  $\otimes^n \tilde{\mathbb{P}}_a$ .

Fix  $p \in (1, \infty)$ , and let the projection direction  $\Theta^n$  be distributed according to the normalized surface measure on  $\mathbb{S}^{n-1}$ , and let  $X^{(n,p)}$  be a random vector independent of  $\Theta^n$  that is uniformly distributed on the unit  $\ell_p^n$  ball. In this article we obtain estimates of tail probabilities of the scaled random projection

$$W^{(n,p)} := \frac{n^{1/p}}{n^{1/2}} \left\langle X^{(n,p)}, \Theta^n \right\rangle = \frac{1}{n} \sum_{i=1}^n \left( n^{1/p} X_i^{(n,p)} \right) \left( n^{1/2} \Theta_i^n \right), \quad (1.3)$$

conditioned on  $\Theta = (\Theta^n)_{n \in \mathbb{N}} = \theta = (\theta^n)_{n \in \mathbb{N}}$ , for a.e. realization  $\theta$  of  $\Theta$ . Using terminology that originates in statistical physics, due to the fact that we condition on the realization  $\theta$  of  $\Theta$  and obtain results for almost every realization, we refer to these as “quenched” deviation estimates. While (quenched) sharp large deviations of sums of weighted i.i.d. random variables with i.i.d. weights have been considered in more recent work [9], comparing the expressions for  $W^{(n,p)}$  and  $S^n$  in (1.3) and (1.1), respectively, we see that  $W^{(n,p)}$  is a randomly weighted sum of random variables that are not independent, with random weights that are also not independent. Thus, the analysis in this case is significantly more challenging and requires several new ingredients. First, we instead exploit a known probabilistic representation for the cone measure on  $\ell_p^n$  spheres [35] to rewrite the tail event  $\{W^{(n,p)} \geq a\}$  as the probability that a certain two-dimensional random vector lies in a certain domain in  $\mathbb{R}^2$  (see Section 2.4), and then establish sharp large deviation estimates for the latter. This transformation turns out to be useful even though sharp large deviations in multiple dimensions are more involved, and none of the existing results (see, e.g., [2, 5, 18] and references therein) apply to this setting. We use Fourier analysis and a change of measure argument to obtain an asymptotic expansion for the quenched two-dimensional density (see Proposition 5.4 and Section 7) and then integrate this density over the appropriate domain. To identify the appropriate change of measure or “tilted” measure, we first show (in Lemma 2.2) that the quenched large deviation rate function obtained in [16] is strictly convex and has a unique minimizer. Along the way, we also establish several results of possible independent result including quantitative central limit theorems under the change of measure (see Lemma 4.4) and multi-dimensional generalized Laplace asymptotics (see Proposition 5.6).

In addition, we also obtain corresponding results for  $\ell_p^n$  balls, where  $X^{(n,p)}$  is replaced with  $\mathcal{X}^{(n,p)}$ , a random variable independent of  $\Theta^n$  distributed according to the normalized volume measure on a scaled  $\ell_p^n$  ball. Obtaining sharp large deviation estimates for random projections of  $\ell_p^n$  balls is substantially more complex than the  $\ell_p^n$  sphere setting because the probability of interest is now expressed as an integral over a three-dimensional domain whose boundary is non-smooth at the minimizing point of the Laplace-type functional (see Section 2.4). This leads to additional difficulties in the computation of the associated Laplace-type asymptotic integral (see Lemma 5.6). As elaborated in

Remarks 2.12 and 2.8, our analytical sharp large deviation estimates do indeed capture additional geometric information beyond the large deviation rate function, and in fact we show that there is a clear difference between sharp tail probabilities in  $\ell_p^n$  balls and spheres, even though they share the same large deviation rate function. Analogous sharp large deviation asymptotics can also be obtained in the case  $p = \infty$  or, in fact, for more general product measures; the analysis in this case is much easier (see, e.g., [27, Section 4.2]).

In order to provide evidence of the accuracy of our sharp analytical estimates of the deviation probabilities for finite  $n$ , we compare them with numerical approximations. Specifically, we use the tilted measure identified in the sharp large deviations analysis to propose an importance sampling scheme that numerically approximates the deviation probabilities. We then compare the estimates obtained from importance sampling with analytical sharp large deviation estimates for a range of  $n$ .

### 1.3 Outline of the rest of the paper

After a summary of common notation and terminology in Sections 1.4 and 2.1, precise statements of the main results are presented in Sections 2.2 and 2.3. An importance sampling algorithm for calculating tail probabilities and comparisons with resulting simulations and the obtained analytical formulas are presented in Section 3. The main results rely on an asymptotic independence result for the weights induced by the projection direction, which is obtained in Section 4, as well as a reformulation of the rare event of interest as the event that a certain random vector lies in a two-dimensional (or three-dimensional) domain, which is described in Section 2.4. Section 2.4 also contains an outline of the proofs of the main results, with the complete proofs of the refined quenched tail estimates given in Sections 5.5 and 5.6 for projections of  $\ell_p^n$  spheres, and in Section 6 for projections of  $\ell_p^n$  balls. Both proofs proceed by first performing asymptotic expansions for the joint densities of the multi-dimensional random vectors, as formulated in Section 5.2. These expansions are derived from a general result on multi-dimensional generalized Laplace approximations obtained in Section 5.3 (see Propositions 5.6 and 5.7 therein) and estimates obtained in Sections 5.1 and 5.4, which justify the applicability of these approximations in the present context. Proofs of several technical results used in the analysis are deferred to Appendices A–F.

### 1.4 Notation and definitions

We use the notation  $\mathbb{N}$ ,  $\mathbb{R}$  and  $\mathbb{C}$  to denote the set of positive integers, real numbers and complex numbers, respectively. For a complex number  $z \in \mathbb{C}$ , we denote  $\operatorname{Re}\{z\}$  to be the real part of  $z$ . For a set  $A$ , we denote its complement by  $A^c$ . Also, given a  $m \times d$  matrix  $M$ , let  $M^T$  denote its transpose and when  $m = d$ , let  $\det M$  denote its determinant.

Given an extended real-valued function  $f : \mathbb{R}^d \rightarrow [0, \infty]$ , its effective domain is defined as  $\{x \in \mathbb{R}^d : f(x) < \infty\}$ . For a twice differentiable function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  (i.e., for which each partial derivative  $\partial_i \partial_j f$  exists for all  $i, j \in \{1, \dots, d\}$ ), let  $\operatorname{Hess} f(x)$  denote the  $d \times d$  Hessian matrix of  $f$  at  $x$ . For  $q \in \mathbb{N}$ , define the function space  $\mathbb{L}_q(\mathbb{R}^d)$  to be

$$\mathbb{L}_q(\mathbb{R}^d) := \left\{ f : \mathbb{R}^d \rightarrow \mathbb{R} : \int_{\mathbb{R}^d} |f|^q dx < \infty \right\}.$$

For  $p \in (1, \infty)$  and  $n \in \mathbb{N}$ , let  $\|\cdot\|_{n,p}$  denote the  $p$ -norm in  $\mathbb{R}^n$ , that is, for  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ ,

$$\|x\|_{n,p} := (|x_1|^p + \dots + |x_n|^p)^{1/p}.$$

Let  $\mathbb{S}_p^{n-1}$  and  $\mathbb{B}_p^n$  denote the unit  $\ell_p^n$  sphere and ball, respectively:

$$\mathbb{S}_p^{n-1} := \{x \in \mathbb{R}^n : \|x\|_{n,p} = 1\} \quad \text{and} \quad \mathbb{B}_p^n := \{x \in \mathbb{R}^n : \|x\|_{n,p} \leq 1\}. \quad (1.4)$$

Also, define the cone measure on  $\ell_p^n$  as follows: for any Borel measurable set  $A \subset \mathbb{S}_p^{n-1}$ ,

$$\mu_{n,p}(A) := \frac{\text{vol}([0, 1]A)}{\text{vol}(\mathbb{B}_p^n)}, \tag{1.5}$$

where  $[0, 1]A := \{xa \in \mathbb{R}^n : x \in [0, 1], a \in A\}$ , and  $\text{vol}$  denotes Lebesgue measure. Note that when  $p \in \{1, 2, \infty\}$ , the (renormalized) cone measure coincides with the (renormalized) surface measure, and is equal to the unique rotational invariant measure on  $\mathbb{S}^{n-1}$  with total mass 1. For the special case  $p = 2$ , we use just  $\|\cdot\|$  to denote  $\|\cdot\|_{n,2}$ , the Euclidean norm on  $\mathbb{R}^n$ ,  $\mathbb{S}^{n-1}$  to denote  $\mathbb{S}_2^{n-1}$  and  $\sigma_n$  to denote  $\mu_{n,2}$ .

We end this section with the definition of a large deviations principle (LDP); we refer to [10] for general background on large deviations theory. For  $d \in \mathbb{N}$ , let  $\mathcal{P}(\mathbb{R}^d)$  denote the space of probability measures on  $\mathbb{R}^d$ , equipped with the topology of weak convergence, where recall that for  $\eta, \eta_n \in \mathcal{P}(\mathbb{R}^d)$ ,  $n \in \mathbb{N}$ ,  $\eta_n$  is said to converge weakly to  $\eta$  as  $n \rightarrow \infty$ , denoted  $\eta_n \Rightarrow \eta$ , if  $\int_{\mathbb{R}^d} f(x)\eta_n(dx) \rightarrow \int_{\mathbb{R}^d} f(x)\eta(dx)$  as  $n \rightarrow \infty$  for every bounded and continuous function  $f$  on  $\mathbb{R}^d$ .

**Definition 1.1** (Large deviation principle). *The sequence of probability measures  $(\eta_n)_{n \in \mathbb{N}} \subset \mathcal{P}(\mathbb{R}^d)$  is said to satisfy a large deviation principle (in  $\mathbb{R}^d$ ) with (speed  $n$  and) a good rate function  $\mathbb{I} : \mathbb{R}^d \rightarrow [0, \infty]$  if  $\mathbb{I}$  is lower semicontinuous and for any measurable set  $A$ ,*

$$-\inf_{x \in A^\circ} \mathbb{I}(x) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \eta_n(A) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \eta_n(A) \leq -\inf_{x \in \text{cl}(A)} \mathbb{I}(x),$$

where  $A^\circ$  and  $\text{cl}(A)$  denote the interior and closure of  $A$ , respectively. Moreover, we say that  $\mathbb{I}$  is a good rate function if it has compact level sets. A sequence of random variables  $(V_n)_{n \in \mathbb{N}}$  with each  $V_n$  defined on some probability space  $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ , is said to satisfy an LDP if the corresponding sequence of laws  $(\mathbb{P}_n^{-1} \circ V_n)_{n \in \mathbb{N}}$  satisfies an LDP.

## 2 Statement of main results

Fix  $p \in (1, \infty)$ . Consider a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  on which are defined three independent sequences  $X = (X^{(n,p)})_{n \in \mathbb{N}}$  and  $\mathcal{X} = (\mathcal{X}^{(n,p)})_{n \in \mathbb{N}}$ , and  $\Theta = (\Theta^n)_{n \in \mathbb{N}}$ . Each  $X^{(n,p)}$  is distributed according to the cone measure  $\mu_{n,p}$  on the unit  $\ell_p^n$  sphere, as defined in (1.5), and each  $\mathcal{X}^{(n,p)}$  is distributed according to the normalized volume measure on the unit  $\ell_p^n$  ball  $\mathbb{B}_p^n$  defined in (1.4). The random element  $\Theta$  takes values in the sequence space  $\mathbb{S} := \otimes_{n \in \mathbb{N}} \mathbb{S}^{n-1}$ , with  $\Theta^n \in \mathbb{S}^{n-1}$  denoting the  $n$ -th element of that sequence, and is independent of  $X$  (and  $\mathcal{X}$ ) with distribution  $\sigma$ , where  $\sigma$  is any probability measure on  $\mathbb{S}$  whose image under the mapping  $\theta \in \mathbb{S} \mapsto \theta^n \in \mathbb{S}^{n-1}$  coincides with  $\sigma_n$ , the unique rotation invariant measure on  $\mathbb{S}^{n-1}$ . The dependence between the random vectors  $\Theta^n$  for different  $n \in \mathbb{N}$  can be arbitrary. For  $\theta \in \mathbb{S}$ , denote  $\mathbb{P}_\theta$  to be the probability measure  $\mathbb{P}$  conditioned on  $\Theta = \theta$ , and let  $\mathbb{E}$  and  $\mathbb{E}_\theta$  denote expectation with respect to  $\mathbb{P}$  and  $\mathbb{P}_\theta$ , respectively. For  $n \in \mathbb{N}$ , let  $W^{(n,p)}$  be the normalized scalar projection of  $X^{(n,p)}$  along  $\Theta^n$  defined as

$$W^{(n,p)} := \frac{n^{1/p}}{n^{1/2}} \sum_{i=1}^n X_i^{(n,p)} \Theta_i^n, \tag{2.1}$$

and similarly let  $\mathcal{W}^{(n,p)}$  be the normalized scalar projection of  $\mathcal{X}^{(n,p)}$  defined as

$$\mathcal{W}^{(n,p)} := \frac{n^{1/p}}{n^{1/2}} \sum_{i=1}^n \mathcal{X}_i^{(n,p)} \Theta_i^n. \tag{2.2}$$

First, in Section 2.1, we introduce notation that is required to state the quenched sharp large deviation estimates. In Section 2.2 we recall the quenched LDP for  $\ell_p^n$  spheres

and balls established in [16] and obtain an important simplification of the quenched LDP rate function obtained therein, which in particular shows that it is convex and has a unique minimum. The latter property will be crucial for our analysis. We then present our sharp large deviation results for projections of  $\ell_p^n$  spheres. Corresponding results for  $\ell_p^n$  balls are presented in Section 2.3. Finally, in Section 2.4 we provide a brief outline of both proofs, and present a more detailed comparison of our results with classical Bahadur-Ranga Rao bounds.

### 2.1 Preliminary notation

Fix  $p \in (1, \infty)$ . Let  $\gamma_p \in \mathcal{P}(\mathbb{R})$  be the  $p$ -Gaussian distribution with density

$$f_p(y) := \frac{1}{2p^{1/p}\Gamma(1 + \frac{1}{p})} e^{-|y|^p/p}, \quad y \in \mathbb{R}, \tag{2.3}$$

where  $\Gamma$  is the Gamma function. For  $t_1, t_2 \in \mathbb{R}$ , define the extended functions

$$\Lambda_p(t_1, t_2) := \log \left( \int_{\mathbb{R}} e^{t_1 y + t_2 |y|^p} \gamma_p(dy) \right), \tag{2.4}$$

and

$$\Psi_p(t_1, t_2) := \int_{\mathbb{R}} \Lambda_p(ut_1, t_2) \gamma_2(du), \tag{2.5}$$

and observe that they both have effective domain  $\mathbb{D}_p := \mathbb{R} \times (-\infty, 1/p)$ . Also, let  $\Psi_p^*$  be the Legendre transform of  $\Psi_p$ :

$$\Psi_p^*(t_1, t_2) := \sup_{s_1, s_2 \in \mathbb{R}} \{t_1 s_1 + t_2 s_2 - \Psi_p(s_1, s_2)\}, \quad t_1, t_2 \in \mathbb{R}, \tag{2.6}$$

and let  $\mathbb{J}_p \subset \mathbb{R}^2$  be the effective domain of  $\Psi_p^*$ :

$$\mathbb{J}_p := \{(x_1, x_2) \in \mathbb{R}^2 : \Psi_p^*(x_1, x_2) < \infty\}. \tag{2.7}$$

Since by [16, Lemma 5.8], the function  $\Lambda_p$  defined in (2.4) is strictly convex on its effective domain, which we denote by  $\mathbb{D}_p$ ,  $\Psi_p$  is also strictly convex on  $\mathbb{D}_p$ . By [16, Lemma 5.9],  $\Psi_p$  is essentially smooth, lower-semicontinuous and hence closed. Therefore by [34, Theorem 26.5],  $\nabla \Psi_p$  is one-to-one and onto from the domain of  $\Psi_p$  to  $\mathbb{J}_p$ . Thus, for each  $(x_1, x_2) \in \mathbb{J}_p$  there exists a unique  $\lambda_x$  such that  $\lambda_x \in \mathbb{D}_p$  and  $\nabla \Psi_p(\lambda_x) = x$ . This in turn implies that  $\lambda_x$  uniquely achieves the supremum in (2.6), and hence that

$$\nabla \Psi_p(\lambda_x) = x, \tag{2.8}$$

and

$$\Psi_p^*(x) = \langle x, \lambda_x \rangle - \Psi_p(\lambda_x). \tag{2.9}$$

**Remark 2.1.** Since  $\Psi_p$  is a strictly convex infinitely differentiable function on  $\mathbb{D}_p$ , the inverse function theorem and (2.8) imply that the mapping  $\mathbb{J}_p \ni x \mapsto \lambda_x \in \mathbb{D}_p$  is also infinitely differentiable.

### 2.2 Results on projections of $\ell_p^n$ spheres

We first state quenched LDPs for the sequences  $(W^{(n,p)})_{n \in \mathbb{N}}$  from (2.1) and  $(\mathcal{W}^{(n,p)})_{n \in \mathbb{N}}$  from (2.2). It follows from [16, Theorem 2.5] that for  $\sigma$ -a.e.  $\theta$ , under  $\mathbb{P}_\theta$ , the sequence  $(\mathcal{W}^{(n,p)})_{n \in \mathbb{N}}$  satisfies an LDP with (speed  $n$  and) a quasiconvex good rate function

$$\mathbb{I}_p(t) = \inf_{\tau_1 \in \mathbb{R}, \tau_2 > 0: \tau_1 \tau_2^{-1/p} = t} \Psi_p^*(\tau_1, \tau_2), \tag{2.10}$$

where recall that a quasiconvex function is a function whose level sets are convex. Furthermore, it follows from [16, Lemmas 3.1 and 3.4] that  $(W^{(n,p)})_{n \in \mathbb{N}}$  also satisfies an LDP with the same speed and rate function. Note that the rate function  $\mathbb{I}_p$  is insensitive to the projection directions, in the sense that it is the same for  $\sigma$ -a.e.  $\theta$ .

We show in the following lemma that the infimum in (2.10) is attained uniquely at  $(t, 1)$ , yielding a simpler form for the rate function and use that to deduce it is strictly convex and has a unique minimizer. The latter is a crucial property both for obtaining sharp large deviation estimates and developing importance sampling algorithms.

**Lemma 2.2.** For  $p \in (1, \infty)$  and  $a > 0$  such that  $\Psi_p^*(a, 1) < \infty$ ,

$$\inf_{\tau_1 \in \mathbb{R}, \tau_2 > 0: \tau_1 \tau_2^{-1/p} = a} \Psi_p^*(\tau_1, \tau_2) = \Psi_p^*(a, 1) = \sup_{s_1, s_2 \in \mathbb{R}} \{as_1 + s_2 - \Psi_p(s_1, s_2)\}.$$

The proof of Lemma 2.2 is relegated to Appendix A; when combined with [16, Theorem 2.5, Lemma 3.1 and Lemma 3.4], it yields the following simpler form of the quenched LDP.

**Theorem 2.3.** Fix  $p \in (1, \infty)$ . For  $\sigma$ -a.e.  $\theta$ , under  $\mathbb{P}_\theta$ , the sequences  $(W^{(n,p)})_{n \in \mathbb{N}}$  and  $(\mathcal{W}^{(n,p)})_{n \in \mathbb{N}}$  both satisfy LDPs with the same strictly convex, symmetric, good rate function  $\mathbb{I}_p$  given by

$$\mathbb{I}_p(a) := \Psi_p^*(a, 1) = \sup_{s_1, s_2 \in \mathbb{R}} \{as_1 + s_2 - \Psi_p(s_1, s_2)\}. \tag{2.11}$$

We now introduce notation to state the sharp large deviation estimate for  $W^{(n,p)}$ . Recall the definitions of  $\Psi_p$ ,  $\Psi_p^*$ ,  $\mathbb{J}_p$  and  $\lambda_x$  from Section 2.1 and for  $x \in \mathbb{J}_p$ , define  $\mathcal{H}_x = \mathcal{H}_{p,x}$  by

$$\mathcal{H}_{p,x} := (\text{Hess } \Psi_p)(\lambda_x), \tag{2.12}$$

where we suppress the dependence on  $p$  from  $\lambda_x$  and  $\mathcal{H}_x$ . Also, fix  $a > 0$  such that  $\mathbb{I}_p(a) < \infty$ . With some abuse of notation, we write  $\lambda_a = \lambda_{a^*}$  and  $\mathcal{H}_a = \mathcal{H}_{a^*}$ , where  $a^* = (a, 1)$ . Note that then  $\lambda_a = (\lambda_{a,1}, \lambda_{a,2}) \in \mathbb{R}^2$  is the unique maximizer in (2.11), that is,

$$\Psi_p^*(a, 1) = a\lambda_{a,1} + \lambda_{a,2} - \Psi_p(\lambda_{a,1}, \lambda_{a,2}), \tag{2.13}$$

and

$$\mathcal{H}_a := (\text{Hess } \Psi_p)(\lambda_a). \tag{2.14}$$

Next, define the positive constants  $\xi_a = \xi_{p,a}$  and  $\kappa_a = \kappa_{p,a}$  via the relations

$$\xi_a^2 := \langle \mathcal{H}_a \lambda_a, \lambda_a \rangle \det \mathcal{H}_a, \tag{2.15}$$

$$\kappa_a^2 := 1 - \frac{(\lambda_{a,1}^2 + \lambda_{a,2}^2)^{3/2} p(p-1)a}{|\lambda_{a,2}^2 (\mathcal{H}_a^{-1})_{11} - 2\lambda_{a,1}\lambda_{a,2} (\mathcal{H}_a^{-1})_{12} + \lambda_{a,1}^2 (\mathcal{H}_a^{-1})_{22}| (a^2 + p^2)^{3/2}}. \tag{2.16}$$

**Remark 2.4.** Although it is not *a priori* obvious that the right-hand side of (2.16) is positive, this will become apparent from the proof of Theorem 2.5.

Finally, also define the following functions: for  $x \in \mathbb{R}$ ,

$$\begin{aligned} \ell_a(x) &:= \Lambda_p(x\lambda_{a,1}, \lambda_{a,2}), \\ \ell_{a,1}(x) &:= x\partial_1 \Lambda_p(x\lambda_{a,1}, \lambda_{a,2}), \\ \ell_{a,2}(x) &:= \partial_2 \Lambda_p(x\lambda_{a,1}, \lambda_{a,2}). \end{aligned} \tag{2.17}$$

Note that the dependence on  $p$  of these functions is again not explicitly notated.

We are now ready to state the quenched sharp large deviation estimate for scaled projections of  $\ell_p^n$  spheres. Recall for  $\theta \in \mathbb{S}$ , we denote  $\mathbb{P}_\theta$  to be the probability measure  $\mathbb{P}$  conditioned on  $\Theta = \theta$ .

**Theorem 2.5.** Fix  $p \in (1, \infty)$  and  $a > 0$  such that  $\mathbb{I}_p(a) < \infty$ . Then the following statements hold with the matrix  $\mathcal{H}_a$  as defined in (2.14) and constants  $\xi_a = \xi_{p,a}$  and  $\kappa_a = \kappa_{p,a}$  defined as in (2.15) and (2.16), respectively:

- (i) For  $n \in \mathbb{N}$ , there exist mappings  $R_a^n = R_{p,a}^n : \mathbb{S}^{n-1} \rightarrow \mathbb{R}$  and  $c_a^n = c_{p,a}^n : \mathbb{S}^{n-1} \rightarrow \mathbb{R}^2$ , defined explicitly in (5.9) and (5.8) as a centered integrated log moment generating function and its gradient, such that for  $\sigma$ -a.e.  $\theta$ ,

$$\mathbb{P}_\theta \left( W^{(n,p)} > a \right) = \frac{C_a^n(\theta^n)}{\kappa_a \xi_a \sqrt{2\pi n}} e^{-n\mathbb{I}_p(a) + \sqrt{n}R_a^n(\theta^n)} (1 + o(1)), \quad (2.18)$$

where

$$C_a^n(\theta^n) := \exp \left( \left\| \mathcal{H}_a^{-1/2} c_a^n(\theta^n) \right\|^2 \right). \quad (2.19)$$

- (ii) Moreover, there exist sequences of random variables  $(r_n = r_{p,a}^n)_{n \in \mathbb{N}}$ ,  $(s_n = s_{p,a}^n)_{n \in \mathbb{N}}$ , and  $(t_{n,i} = t_{p,a,i}^n)_{n \in \mathbb{N}}$ ,  $i = 1, 2$ , (defined on some common probability space) such that for each  $n \in \mathbb{N}$ ,

$$(R_a^n(\Theta^n), c_a^n(\Theta^n)) \stackrel{(d)}{=} \left( r_n + \frac{1}{\sqrt{n}} s_n + o\left(\frac{1}{\sqrt{n}}\right), (t_{n,1} + o(1), t_{n,2} + o(1)) \right), \quad (2.20)$$

and as  $n \rightarrow \infty$ ,

$$(r_n, s_n, t_{n,1}, t_{n,2}) \Rightarrow (\mathfrak{R}, \mathfrak{S}, \mathfrak{T}_1, \mathfrak{T}_2),$$

where

$$(\mathfrak{R}, \mathfrak{S}, \mathfrak{T}_1, \mathfrak{T}_2) :=$$

$$\left( \tilde{\mathfrak{A}} - \frac{1}{2} \mathbb{E}[\ell'_a(Z)Z] \tilde{\mathfrak{D}}, \frac{1}{8} \mathbb{E}[\ell''_a(Z)Z^2] \tilde{\mathfrak{D}}^2, \tilde{\mathfrak{E}} - \frac{1}{2} \mathbb{E}[\ell'_{a,1}(Z)Z] \tilde{\mathfrak{D}}, \tilde{\mathfrak{E}} - \frac{1}{2} \mathbb{E}[\ell'_{a,2}(Z)Z] \tilde{\mathfrak{D}} \right),$$

$Z$  is a standard Gaussian random variable, and  $(\tilde{\mathfrak{A}}, \tilde{\mathfrak{D}}, \tilde{\mathfrak{E}}, \tilde{\mathfrak{E}})$  are jointly Gaussian with mean 0 and covariance matrix  $\Sigma_a = \Sigma_{p,a}$  that takes the following explicit form:

$$\begin{pmatrix} \text{Cov}(\ell_a(Z), \ell_a(Z)) & \text{Cov}(\ell_a(Z), Z^2) & \text{Cov}(\ell_a(Z), \ell_{a,1}(Z)) & \text{Cov}(\ell_a(Z), \ell_{a,2}(Z)) \\ \text{Cov}(Z^2, \ell_a(Z)) & \text{Cov}(Z^2, Z^2) & \text{Cov}(Z^2, \ell_{a,1}(Z)) & \text{Cov}(Z^2, \ell_{a,2}(Z)) \\ \text{Cov}(\ell_{a,1}(Z), \ell_a(Z)) & \text{Cov}(\ell_{a,1}(Z), Z^2) & \text{Cov}(\ell_{a,1}(Z), \ell_{a,1}(Z)) & \text{Cov}(\ell_{a,1}(Z), \ell_{a,2}(Z)) \\ \text{Cov}(\ell_{a,2}(Z), \ell_a(Z)) & \text{Cov}(\ell_{a,2}(Z), Z^2) & \text{Cov}(\ell_{a,2}(Z), \ell_{a,1}(Z)) & \text{Cov}(\ell_{a,2}(Z), \ell_{a,2}(Z)) \end{pmatrix}. \quad (2.21)$$

An outline of the proof of Theorem 2.5 is given in Section 2.4, with full details provided in Sections 5.5 and 5.6. See also (7.12) and (7.13) for an interpretation of  $c_a^n$  and  $\mathcal{H}_a$  as the scaled mean vector and limiting covariance matrix, under a quenched tilted measure of a two-dimensional vector that arises in a convenient representation for  $W^{(n,p)}$  described in Section 2.4).

**Remark 2.6.** We will refer to the term  $C_a^n(\theta^n) e^{-n\mathbb{I}_p(a) + \sqrt{n}R_a^n(\theta^n)} / \kappa_a \xi_a \sqrt{2\pi n}$  in (2.18) as the “prefactor” since it provides a multiplicative correction to the exponentially decaying term  $e^{-n\mathbb{I}_p(a)}$ , which is identified by the LDP. In addition, it follows from (2.19)–(2.20) that (in distribution)  $R_a^n(\Theta^n)$  and  $c_a^n(\Theta^n)$  both converge to zero as  $n \rightarrow \infty$ ; see also Lemma 5.8 for more refined estimates. Further insight into the form of the prefactor can be found in Remarks 2.8 and 2.15.

As mentioned above, the most significant term in the prefactor that depends on  $\theta$  is  $e^{\sqrt{n}R_{p,a}^n(\theta^n)}$ . The following proposition describes the additional geometric information contained in this term beyond what is available in the rate function  $\mathbb{I}_p$ , which is  $\sigma$ -almost surely insensitive to the projection sequence  $\Theta$ .

**Proposition 2.7.** Fix  $p \in (1, \infty)$ ,  $a > 0$  such that  $\mathbb{I}_p(a) < \infty$  and let  $R_{p,a}^n$  be the mapping in Theorem 2.5 that is defined explicitly in (5.8). Then

1. For  $p = 2$ ,  $R_{p,a}^n(\theta^n)$  is a constant regardless of the direction  $\theta^n \in \mathbb{S}^{n-1}$ ;
2. For  $p > 2$ , the maximum of  $R_{p,a}^n(\theta^n)$  over  $\theta^n \in \mathbb{S}^{n-1}$  is attained at  $(\pm 1, \pm 1, \dots, \pm 1)/\sqrt{n}$ , while the minimum is attained at  $\pm e_j$  for  $j = 1, \dots, n$ ;
3. For  $p < 2$ , the minimum of  $R_{p,a}^n(\theta^n)$  over  $\theta^n \in \mathbb{S}^{n-1}$  is attained at  $(\pm 1, \pm 1, \dots, \pm 1)\sqrt{n}$ , while the maximum is attained at  $\pm e_j$  for  $j = 1, \dots, n$ ,

where  $(e_j)_{j=1, \dots, n}$  are defined to be the standard basis vectors in  $\mathbb{R}^n$ .

**Remark 2.8.** Proposition 2.7 in conjunction with Theorem 2.5 shows how the sharp large deviation estimates reflect the difference in the geometry of  $\ell_p^n$  spheres for  $p \in (1, 2)$  and  $p \in (2, \infty)$  with respect to the relative distribution of mass along different rays. This motivates obtaining sharp large deviation estimates for projections of more general high-dimensional objects to uncover new geometric information about these objects.

As a corollary, combining the two parts of Theorem 2.5, we obtain an alternative expression for the distribution of the conditioned tail probability:

**Corollary 2.9.** Fix  $p \in (1, \infty)$  and  $a > 0$  such that  $\mathbb{I}_p(a) < \infty$ . For  $n \in \mathbb{N}$ , recall the definitions of  $(r_n)_{n \in \mathbb{N}}$ ,  $(s_n)_{n \in \mathbb{N}}$  and  $(t_n)_{n \in \mathbb{N}}$  in Theorem 2.5 (ii), and that of  $\mathcal{H}_a$  from (2.14). Then

$$\mathbb{P}_\Theta \left( W^{(n,p)} > a \right) := \mathbb{P} \left( W^{(n,p)} > a \mid \Theta^n \right) \stackrel{(d)}{=} \frac{M_n}{\kappa_a \xi_a \sqrt{2\pi n}} e^{-n\mathbb{I}_p(a) + \sqrt{n}r_n} (1 + o(1)),$$

where

$$M_n := \exp \left( s_n + \left\| \mathcal{H}_a^{-1/2} t_n \right\|^2 \right). \tag{2.22}$$

Moreover, as  $n \rightarrow \infty$ ,

$$(M_n, r_n) \Rightarrow \left( \exp \left( \mathfrak{S} + \left\| \mathcal{H}_a^{-1/2} \mathfrak{T} \right\|^2 \right), \mathfrak{R} \right), \tag{2.23}$$

where  $(\mathfrak{R}, \mathfrak{S}, \mathfrak{T}_1, \mathfrak{T}_2)$  is as defined in Theorem 2.5(ii).

*Proof.* By (2.18), (2.19) and (2.20), the tail probability can be written as

$$\begin{aligned} \mathbb{P}_\Theta \left( W^{(n,p)} > a \right) &\stackrel{(d)}{=} \frac{e^{\left\| \mathcal{H}_a^{-1/2} t_n \right\|^2 + o(1)}}{\kappa_a \xi_a \sqrt{2\pi n}} e^{-n\mathbb{I}_p(a) + \sqrt{n}r_n + s_n + o(1)} (1 + o(1)) \\ &= \frac{M_n}{\kappa_a \xi_a \sqrt{2\pi n}} e^{-n\mathbb{I}_p(a) + \sqrt{n}r_n} (1 + o(1)), \end{aligned}$$

since  $\exp(o(1)) = 1 + o(1)$ . Also, from the relation (2.22), the mapping  $(r_n, s_n, t_{n,1}, t_{n,2}) \mapsto (M_n, r_n)$  is continuous. Therefore, we may apply the continuous mapping theorem to the last display, and invoke Theorem 2.5(ii) to obtain the joint convergence stated in (2.23).  $\square$

**Remark 2.10.** In Theorem 2.5 and Corollary 2.9 we only consider values  $p \in (1, \infty)$  because for  $p \in (0, 1)$ ,  $\ell_p^n$  balls are no longer convex and the existence of even an LDP has not been established. Moreover, as shown in [16, Theorem 2.6], when  $p = 1$  the quenched LDP of the projection exists only when the projection directions satisfy  $\lim_{n \rightarrow \infty} \sqrt{\frac{n}{\log n}} \max_{1 \leq i \leq n} \theta_i^{(n)} = c$  for some constant  $c \in (0, \infty)$ , and in that case, it is with speed  $n/\sqrt{\log n}$  and the rate function is no longer universal but depends on the limiting constant  $c$ . On the other hand, we omit the case  $p = \infty$ , or the more general case of product measures, because this is in fact simpler to analyze than the  $p \in (1, \infty)$  case; details can be found in [27, Section 4.2].

### 2.3 Results on projections of $\ell_p^n$ balls

Next, we state the corresponding sharp large deviation results for balls. For  $p \in (1, \infty)$  and  $a > 0$ , recalling that  $\lambda_{a,1}$  is the first coordinate of the maximizer  $\lambda_a$  in the expression for  $\Psi_p^*(a)$  in (2.9) and  $\mathcal{H}_a$  is as defined in (2.14), define the positive constant  $\gamma_a = \gamma_{p,a}$  via the relation

$$\gamma_a^2 := \lambda_{a,1}^2 (1 + a\lambda_{a,1})^2 (\det \mathcal{H}_a)^2 \left| -\frac{a(p-1)}{p^2} \lambda_{a,1} + \frac{2a}{p} (\mathcal{H}_a)_{12}^{-1} + (\mathcal{H}_a)_{22}^{-1} + \frac{a^2}{p^2} (\mathcal{H}_a)_{11}^{-1} \right|. \tag{2.24}$$

**Theorem 2.11.** Fix  $p \in (1, \infty)$  and  $a > 0$  such that  $\mathbb{I}_p(a) < \infty$ . Then for  $n \in \mathbb{N}$ ,

$$\mathbb{P}_\theta \left( \mathcal{W}^{(n,p)} > a \right) = \frac{C_a^n(\theta^n)}{\gamma_a \sqrt{2\pi n}} e^{-n\mathbb{I}_p(a) + \sqrt{n}R_a^n(\theta^n)} (1 + o(1)), \tag{2.25}$$

where  $\gamma_a = \gamma_{p,a}$  is the constant defined in (2.24), and  $R_a^n$  and  $C_a^n$  are the functions defined in Theorem 2.5.

**Remark 2.12.** (i) Note that the tail probability in (2.25) is a geometric quantity, equal to the volume of the  $p$ -spherical cap (at level  $a$ ) of  $\ell_p^n$  balls along the direction  $\theta^n$ .

(ii) Recall that it follows from the results of [16] (recapitulated here as Theorem 2.3) that  $\ell_p^n$  spheres and balls cannot be distinguished because the large deviation speeds and rate functions for random projections of  $\ell_p^n$  balls and spheres coincide. In contrast, we see from (2.18) and (2.25) that although the two prefactors have a similar form, their actual values differ since in general  $\gamma_a \neq \kappa_a \xi_a$ . Thus, the sharp large deviation estimates obtained here are sufficiently refined to distinguish these two objects, whereas the LDP rate function does not do so.

(iii) As in Remark 2.8, due to the appearance of  $R_a^n$  in (2.25), the sharp large deviation estimate provides more insight into the distinction between the geometry of  $\ell_p^n$  balls with  $p \in (1, 2)$  and  $\ell_p^n$  balls with  $p \in (2, \infty)$ .

Similar to Corollary 2.9, we have the following immediate corollary for balls:

**Corollary 2.13.** Fix  $p \in (1, \infty)$  and  $a > 0$  such that  $\mathbb{I}_p(a) < \infty$ . For  $n \in \mathbb{N}$ , recall the definitions of  $(M_n)_{n \in \mathbb{N}}$  and  $(r_n)_{n \in \mathbb{N}}$  in Corollary 2.9 and let  $\gamma_a$  be as in (2.24). Then for  $n \in \mathbb{N}$ ,

$$\mathbb{P}_\Theta \left( \mathcal{W}^{(n,p)} > a \right) \stackrel{(d)}{=} \frac{M_n}{\gamma_a \sqrt{2\pi n}} e^{-n\mathbb{I}_p(a) + \sqrt{n}r_n} (1 + o(1)),$$

where (2.22) and (2.23) hold.

### 2.4 Reformulation of the problem and outline of the proof

Fix  $p \in (1, \infty)$ . As mentioned in the introduction, one of the reasons the estimate (2.18) is challenging to establish is that  $W^{(n,p)}$  and  $\mathcal{W}^{(n,p)}$  are randomly weighted sums of random variables that are not independent, and furthermore, the random weights are also themselves not independent. In this section we provide a brief outline of our proof and additional insight into the form of the sharp large deviation estimates, contrasting them with existing results, and explaining the role of various constants.

The first step of the proof is to reformulate the probability of the rare event in terms of a certain multi-dimensional random vector ( $\bar{S}^{(n,p)}$  in the case of spheres and  $\bar{\mathfrak{S}}^{(n,p)}$  in the case of balls) using a well known probabilistic representation for the random vector  $X^{(n,p)}$  that we now recall. Assume without loss of generality that the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  is large enough to also support an i.i.d. sequence of generalized  $p$ -Gaussian random variables  $(Y_i^{(p)})_{i \in \mathbb{N}}$ , independent of  $\Theta$ , and define the  $n$ -dimensional random

vector  $Y^{(n,p)} := (Y_1^{(p)}, \dots, Y_n^{(p)})$ , where each  $Y_j^{(p)}$  has density  $f_p$  defined in (2.3). Then, it follows from [35, Lemma 1] (see also a statement of this property at the bottom of p. 548 in [8]) that

$$X^{(n,p)} \stackrel{(d)}{=} \frac{Y^{(n,p)}}{\|Y^{(n,p)}\|_{n,p}}, \quad n \in \mathbb{N}, \tag{2.26}$$

where recall that  $\|x\|_{n,p}$  denotes the  $p$ -norm in  $\mathbb{R}^n$ . Define the  $\mathbb{R}^2$ -valued random vector

$$\bar{S}^{(n,p)} := \frac{1}{n} \sum_{j=1}^n \left( \sqrt{n} \Theta_j^n Y_j^{(p)}, |Y_j^{(p)}|^p \right). \tag{2.27}$$

In view of (2.1) and the independence of  $X^{(n,p)}$ , (2.26), and  $\Theta$ , for  $a > 0$  and  $\theta \in \mathbb{S}$ , we may rewrite the tail probability on the left-hand side of (2.18) as

$$\begin{aligned} \mathbb{P}_\theta \left( W^{(n,p)} > a \right) &= \mathbb{P} \left( \frac{n^{1/p}}{n} \sum_{j=1}^n \frac{\sqrt{n} \theta_j^n Y_j^{(p)}}{\|Y^{(n,p)}\|_{n,p}} > a \right) \\ &= \mathbb{P} \left( \frac{1}{n} \sum_{j=1}^n \sqrt{n} \theta_j^n Y_j^{(p)} > a \left( \frac{1}{n} \sum_{j=1}^n |Y_j^{(p)}|^p \right)^{1/p} \right) \\ &= \mathbb{P}_\theta \left( \bar{S}^{(n,p)} \in \bar{D}_{p,a} \right), \end{aligned} \tag{2.28}$$

where  $\bar{D}_{p,a}$  is the two-dimensional domain defined by

$$\bar{D}_{p,a} := \left\{ (x_1, x_2) \in \mathbb{R}^2 : x_2 > 0, x_1 > ax_2^{1/p} \right\}. \tag{2.29}$$

On the other hand, again from [35, Lemma 1], we also have an equivalent representation for  $\mathcal{X}^{(n,p)}$ :

$$\mathcal{X}^{(n,p)} \stackrel{(d)}{=} \mathcal{U}^{1/n} \frac{Y^{(n,p)}}{\|Y^{(n,p)}\|_{n,p}}, \quad n \in \mathbb{N}, \tag{2.30}$$

where  $\mathcal{U}$  is a uniform random variable on  $(0, 1)$ , independent of the sequence  $(Y^{(n,p)})_{n \in \mathbb{N}}$ . Define the  $\mathbb{R}^3$ -valued random vector

$$\bar{\mathcal{S}}^{(n,p)} := \left( \frac{1}{n} \sum_{j=1}^n \sqrt{n} \Theta_j^n Y_j^{(p)}, \frac{1}{n} \sum_{j=1}^n |Y_j^{(p)}|^p, \mathcal{U}^{1/n} \right) = \left( \bar{S}^{(n,p)}, \mathcal{U}^{1/n} \right). \tag{2.31}$$

From the equivalent representation (2.30), for  $a > 0$  and  $\theta \in \mathbb{S}$ , we may rewrite the tail probability of  $\mathcal{W}^{(n,p)}$  as

$$\begin{aligned} \mathbb{P}_\theta \left( \mathcal{W}^{(n,p)} > a \right) &= \mathbb{P} \left( \frac{n^{1/p}}{n} \sum_{j=1}^n \mathcal{U}^{1/n} \frac{\sqrt{n} \theta_j^n Y_j^{(p)}}{\|Y^{(n,p)}\|_{n,p}} > a \right) \\ &= \mathbb{P} \left( \mathcal{U}^{1/n} \frac{1}{n} \sum_{j=1}^n \sqrt{n} \theta_j^n Y_j^{(p)} > a \left( \frac{1}{n} \sum_{j=1}^n |Y_j^{(p)}|^p \right)^{1/p} \right) \\ &= \mathbb{P}_\theta \left( \bar{\mathcal{S}}^{(n,p)} \in \bar{\mathcal{D}}_{p,a} \right), \end{aligned} \tag{2.32}$$

where  $\bar{\mathcal{D}}_{p,a}$  is the three-dimensional domain given by

$$\bar{\mathcal{D}}_{p,a} := \left\{ (x_1, x_2, y) \in \mathbb{R}^3 : 1 > y > 0, x_2 > 0, x_1 y > ax_2^{1/p} \right\}. \tag{2.33}$$

**Remark 2.14.** Throughout the paper, we will typically use an overline to denote quantities related to these multi-dimensional reformulations, and script fonts for quantities related to  $\ell_p^n$  balls.

While several results on sharp large deviations in multiple dimensions have been obtained (see, e.g., [2, 18] as well as [5] for a comprehensive list of references), none of these cover the cases of interest in (2.28) and (2.32). In particular, the work [2] considers empirical means of i.i.d. random vectors whereas, under  $\mathbb{P}_\theta$ ,  $\bar{S}^{(n,p)}$  is the empirical mean of non-identical random vectors. Moreover, the results of [18] also do not apply since the condition imposed in [18, Assumption (A.2)] is not satisfied here due to the additional  $\sqrt{n}$  factor in the exponent of (2.18) compared with [18, Equation (3)]. Instead, our proof proceeds by first exploiting quantitative asymptotic independence results of the weights  $(\Theta_i^n)_{i=1,\dots,n}$  obtained in Section 4, and combining them with new asymptotic estimates for certain Laplace-type integrals stated in Section 5.

**Remark 2.15.** Comparing the estimate in (2.18) with the sharp large deviation estimate for the projection of an i.i.d. sum onto the vector  $\mathcal{I}^n = (1, 1, \dots, 1)/\sqrt{n}$  given in (1.2), we see that  $\xi_a$  in (2.18) plays a role similar to  $\bar{\sigma}_a \tau_a$  in (1.2). On the other hand, the additional constant  $\kappa_a$  in (2.18) arises due to the geometry of the domain  $\bar{D}_{p,a}$  defined in (2.29) and the fact that we obtain this estimate by reformulating it in terms of a two-dimensional problem. From a technical point of view, the additional  $\theta^n$ -dependent terms  $R_a^n(\theta^n)$  and  $C_a^n(\theta^n)$  arise because we are considering (quenched) sharp large deviations of a vector  $\bar{S}^{(n,p)}$  whose independent summands are not identically distributed under  $\mathbb{P}_\theta$  on account of the different weights arising from the coordinates of  $\theta^n$ . From their exact definitions given in (5.8) and (5.9), it is easy to see that both these terms would vanish if we considered  $\theta^n \in \mathbb{S}^{n-1}$  with identical weights such as  $\theta^n = \mathcal{I}^n = (1, 1, \dots, 1)/\sqrt{n}$ .

### 3 An importance sampling algorithm

To numerically compute the tail probability  $\mathbb{P}_\theta(W^{(n,p)} > a) = \mathbb{E}_\theta[1_{\{W^{(n,p)} > a\}}]$  using standard Markov Chain Monte Carlo (MCMC), for any  $\theta^n \in \mathbb{S}^{n-1}$ , one would have to generate independent samples of  $X^{(n,p)}$  from the cone measure  $\mu_{n,p}$  defined in (1.5), and use the empirical mean as an estimate of the expectation. However, since the probability is very small, this is inefficient or computationally infeasible for even moderate values of  $n$ . In this section, we propose an alternative *importance sampling (IS) algorithm* to more efficiently compute the tail probability numerically, for a range of values of  $n$ , and compare this with the analytical estimate obtained in Theorem 2.5. For  $a > 0$ , fix  $p \in (1, \infty)$  and recall the constant  $\lambda_a$  defined in (2.13). Also, recall the definition of the density  $f_p$  in (2.3). Given  $n \in \mathbb{N}$ , let  $\tilde{Y}^{(n,p)} := (\tilde{Y}_1^{(n,p)}, \dots, \tilde{Y}_n^{(n,p)})$ , where  $\tilde{Y}_j^{(n,p)}$ ,  $j = 1, \dots, n$ , are random variables defined on  $(\Omega, \mathcal{F}, \mathbb{P})$  that are independent under  $\mathbb{P}_\theta$  for each  $\theta \in \mathbb{S}$ , and such that  $\tilde{Y}_j^{(n,p)}$  has density

$$\tilde{f}_{p,j}^n(y) := \exp(\langle \lambda_a, (\sqrt{n}\theta_j^n y, |y|^p) \rangle - \Lambda_p(\sqrt{n}\theta_j^n \lambda_{a,1}, \lambda_{a,2})) f_p(y), \quad y \in \mathbb{R}, \quad (3.1)$$

where we suppress from the notation the explicit dependence of  $\tilde{f}_{p,j}^n$  on  $\theta^n$ . Also define

$$\tilde{W}^{(n,p)} := \frac{n^{1/p}}{n^{1/2}} \sum_{j=1}^n \frac{\tilde{Y}_j^{(n,p)} \Theta_j^n}{\|\tilde{Y}^{(n,p)}\|_{n,p}}. \quad (3.2)$$

In view of (3.1) and (3.2), it then follows that

$$\mathbb{P}_\theta(W^{(n,p)} > a) = \mathbb{E}_\theta \left[ 1_{\{\tilde{W}^{(n,p)} > a\}} \prod_{j=1}^n \exp\left(-\langle \lambda_a, (\sqrt{n}\theta_j^n \tilde{Y}_j^{(n,p)}, |\tilde{Y}_j^{(n,p)}|^p) \rangle\right) \right]$$

$$+ \Lambda_p \left( \sqrt{n} \theta_i^n \lambda_{a,1}, \lambda_{a,2} \right) \Big]. \tag{3.3}$$

The IS algorithm estimates the tail probability on the left-hand side of (3.3) by first sampling a direction  $\theta^n$  according to  $\sigma_n$  and then sampling from i.i.d. copies of the vector  $\tilde{Y}^{(n,p)} := (\tilde{Y}_1^p, \dots, \tilde{Y}_n^p)$ , independently of the  $\theta^n$  sample, to approximate the expectation on the right-hand side of (3.3) by a standard Monte Carlo estimate.

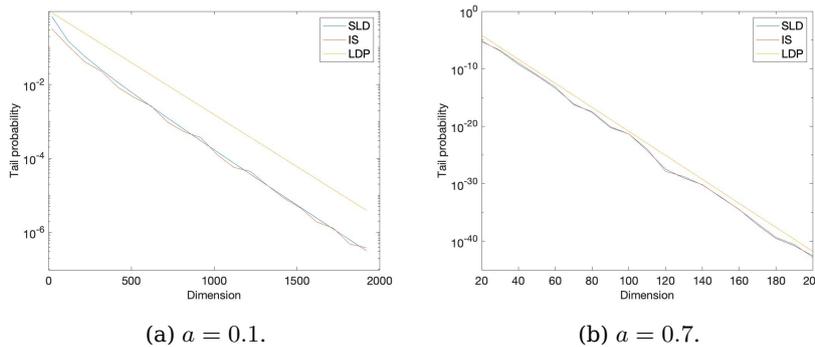


Figure 1: Log scale plot of estimates of  $P_\theta(W^{(n,3)} > a)$  vs. dimension.

The results are displayed in Figures 1–2 and Tables 1–2. In each case, the IS estimate is computed as above, the LDP estimate is  $e^{-n\mathbb{I}_p(a)}$  (i.e., with 1 as a prefactor), and the sharp large deviation (SLD) estimate is the prefactor (see Remark 2.6) times  $e^{-n\mathbb{I}_p(a)}$ . We consider  $p = 3$  with only 100 samples since we do not have closed form expressions for various functions needed in the IS simulation, thus requiring greater computational effort per sample. In Table 1 we also calculate the confidence interval of the IS estimate and tabulate the relative distance between the SLD and IS estimates, computed as  $(\text{SLD} - \text{IS}) \times 100/\text{IS}$ . First, we see from Figure 1 that the LDP estimate is not a good enough approximation, but the sharp large deviation (SLD) estimate does a much better job. For large  $a$ , namely  $a = 0.7$ , in Figure 1(B) and Table 1, we see that the SLD and IS estimates match pretty well even for small  $n$  (namely, even  $n = 20$ ). However, this is not the case for  $a$  small, namely for  $a = 0.1$ . In this case, as evident from Figure 1(A) and Table 2, the SLD estimate appears to achieve the same accuracy only for much larger  $n$ , which likely reflects the dependence of the  $o(1)$  term in (2.18) on  $a$ .

Table 1: Estimates of  $P_\theta(W^{(n,p)} > 0.7)$  for  $p = 3$ . The sample size for IS is 100.

$n$	SLD	IS	Relative distance	Confidence Interval
20	$6.8707 \times 10^{-6}$	$5.3317 \times 10^{-6}$	27.18%	$[3.4203 \times 10^{-6}, 7.2430 \times 10^{-6}]$
80	$2.5403 \times 10^{-18}$	$3.4245 \times 10^{-18}$	-25.82%	$[1.5542 \times 10^{-18}, 5.2948 \times 10^{-18}]$
140	$6.9378 \times 10^{-31}$	$6.1856 \times 10^{-31}$	12.16%	$[2.5305 \times 10^{-31}, 9.8407 \times 10^{-31}]$
200	$2.8813 \times 10^{-43}$	$1.6547 \times 10^{-43}$	74.13%	$[4.0920 \times 10^{-44}, 2.9002 \times 10^{-43}]$

Finally, we also ran simulations for different realizations  $\theta$  of the direction sequence  $\Theta$ . We see from Figure 2 that different projection direction sequences result in fluctuations around the quantity  $e^{-n\mathbb{I}_p(a)} / \kappa_a \xi_a \sqrt{2\pi n}$ , which is the basic sharp large deviation estimate obtained by ignoring the  $\theta^n$ -dependent terms in the prefactor in (2.18). As shown in Theorem 2.5(ii), these fluctuations converge in distribution to functionals of a multi-dimensional Gaussian vector with an explicit covariance matrix.

Table 2: Estimates of  $P_\theta(W^{(n,p)} > 0.1)$  for  $p = 3$ . The sample size for IS is 100.

$n$	SLD	IS	Relative distance	Confidence Interval
20	$6.9193 \times 10^{-1}$	$3.2004 \times 10^{-1}$	116.20%	$[2.5636 \times 10^{-1}, 3.8372 \times 10^{-1}]$
420	$1.1317 \times 10^{-2}$	$8.3597 \times 10^{-3}$	35.38%	$[5.6085 \times 10^{-3}, 1.1110 \times 10^{-2}]$
820	$6.0651 \times 10^{-4}$	$5.2198 \times 10^{-4}$	16.19%	$[3.2412 \times 10^{-4}, 7.1985 \times 10^{-4}]$
1220	$3.7235 \times 10^{-5}$	$4.4306 \times 10^{-5}$	15.96%	$[2.8312 \times 10^{-5}, 6.0299 \times 10^{-5}]$

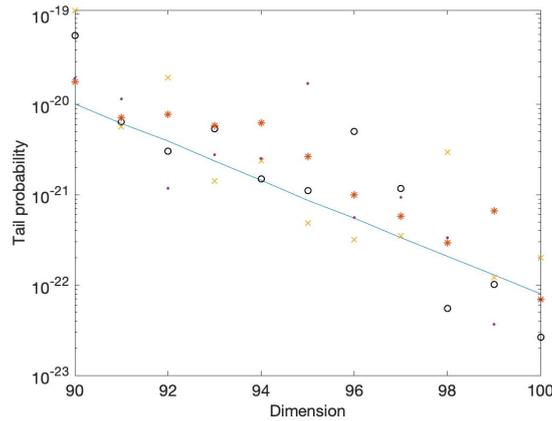


Figure 2: Log scale plot of estimates of  $P_\theta(W^{(n,p)} > 0.7)$  vs. dimension for  $p = 3$ . Solid line is  $e^{-nI_p(a)} / \kappa_a \xi_a \sqrt{2\pi n}$ . Scatter points are SLD estimates for different direction sequences.

#### 4 Asymptotic independence results for the weights

Recall that  $\mathcal{P}(\mathbb{R})$  is the set of probability measures on  $\mathbb{R}$ . For  $p \in [1, \infty)$ , denote

$$\mathcal{P}_p(\mathbb{R}) := \left\{ \nu \in \mathcal{P}(\mathbb{R}) : \int_{\mathbb{R}} |u|^p \nu(du) < \infty \right\},$$

and equip  $\mathcal{P}_p(\mathbb{R})$  with the  $p$ -Wasserstein distance defined to be

$$\mathcal{W}_p(\nu, \nu') := \inf_{\pi \in \Pi(\nu, \nu')} \int_{\mathbb{R}^2} |x - y|^p \pi(dx, dy), \quad \nu, \nu' \in \mathcal{P}_p(\mathbb{R}), \quad (4.1)$$

where  $\Pi(\nu, \nu')$  denotes the set of couplings of  $\nu$  and  $\nu'$  or equivalently, the set of probability measures on  $\mathbb{R}^2$  whose first and second marginals coincide with  $\nu$  and  $\nu'$ , respectively.

We now define a function with polynomial growth in the natural way.

**Definition 4.1.** Given  $m \in \mathbb{N}$ , we say that a function  $f : \mathbb{R} \rightarrow \mathbb{R}$  has polynomial growth of degree  $m$  if there exist  $T \in \mathbb{R}$  and  $C \in (0, \infty)$  such that

$$|f(t)| \leq C(|t|^m + 1), \quad \text{for } |t| > T.$$

We say a function  $f : \mathbb{R} \rightarrow \mathbb{R}$  has polynomial growth if it has polynomial growth of degree  $m$  for some  $m \in \mathbb{N}$ .

Next, we recall the definition of the  $p$ -Wasserstein distance on probability measures.

**Lemma 4.2** (Definition 6.8 and Theorem 6.9 of [39]). Let  $(\nu^n)_{n \in \mathbb{N}} \subset \mathcal{P}_p(\mathbb{R})$  and  $\nu \in \mathcal{P}_p(\mathbb{R})$ . Then the following two statements are equivalent:

1.  $\mathcal{W}_p(\nu^n, \nu) \rightarrow 0$ .

2. For any continuous  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  that has polynomial growth of degree  $p$

$$\int_{\mathbb{R}} \phi(x) \nu^n(dx) \rightarrow \int_{\mathbb{R}} \phi(x) \nu(dx).$$

For each  $n \in \mathbb{N}$  and  $\theta \in \mathbb{S}$ , let  $L_\theta^n$  denote the empirical measure of the coordinates of the scaled projection direction  $\sqrt{n}\theta^n$ :

$$L_\theta^n := \frac{1}{n} \sum_{i=1}^n \delta_{\sqrt{n}\theta_i^n}. \tag{4.2}$$

The following strong law of large numbers for  $(L_\theta^n)_{n \in \mathbb{N}}$  was established in [16, Lemma 5.11]. Recall that  $\gamma_2$  denotes the standard normal distribution.

**Lemma 4.3** (Lemma 5.11 of [16]). For  $p \in (1, \infty)$ , for  $\sigma$ -a.e.  $\theta \in \mathbb{S}$ ,

$$\mathcal{W}_p(L_\theta^n, \gamma_2) \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

We now establish a central limit theorem refinement of Lemma 4.3. Given an i.i.d. array  $(Z^n = (Z_j^n, j = 1 \dots, n))_{n \in \mathbb{N}}$  of standard normal random variables, for any twice continuously differentiable function  $\phi$ , define

$$\hat{s}_n(\phi) := \sum_{j=1}^n \frac{\phi''(Z_j^n)}{2} \left( \frac{\sqrt{n}Z_j^n}{\|Z^n\|_{n,2}} - Z_j^n \right)^2, \tag{4.3}$$

and set

$$\hat{r}_n(\phi) := \frac{1}{\sqrt{n}} \sum_{j=1}^n \left[ \phi(Z_j^n) - \int_{\mathbb{R}} \phi(x) \gamma_2(dx) + \phi'(Z_j^n) \left( \frac{\sqrt{n}Z_j^n}{\|Z^n\|_{n,2}} - Z_j^n \right) \right]. \tag{4.4}$$

For any probability measure  $\pi \in \mathcal{P}(\mathbb{R})$ , define  $\pi(F) := \int_{\mathbb{R}} F(x) \pi(dx)$ , for any Borel measurable function  $F : \mathbb{R} \rightarrow \mathbb{R}$ .

**Lemma 4.4.** Given a thrice continuously differentiable function  $F : \mathbb{R} \rightarrow \mathbb{R}$  and two twice continuously differentiable functions  $G_1, G_2 : \mathbb{R} \rightarrow \mathbb{R}$  such that  $F''', G_1''$  and  $G_2''$  have polynomial growth in the sense of Definition 4.1, we have the following expansion:

$$\begin{aligned} & \sqrt{n} (L_\theta^n(F) - \gamma_2(F), L_\theta^n(G_1) - \gamma_2(G_1), L_\theta^n(G_2) - \gamma_2(G_2)) \\ & \stackrel{(d)}{=} \left( \hat{r}_n(F) + \frac{1}{\sqrt{n}} \hat{s}_n(F) + o\left(\frac{1}{\sqrt{n}}\right), \hat{r}_n(G_1) + o(1), \hat{r}_n(G_2) + o(1) \right), \end{aligned}$$

where  $\hat{s}_n$  and  $\hat{r}_n$  are as defined in (4.3) and (4.4), and as  $n \rightarrow \infty$ ,

$$\begin{aligned} & (\hat{r}_n(F), \hat{s}_n(F), \hat{r}_n(G_1), \hat{r}_n(G_2)) \\ & \Rightarrow \left( \tilde{\mathfrak{A}} - \frac{1}{2} \mathbb{E}[F'(Z)Z] \tilde{\mathfrak{D}}, \frac{1}{8} \mathbb{E}[F'''(Z)Z^2] \tilde{\mathfrak{D}}^2, \tilde{\mathfrak{E}} - \frac{1}{2} \mathbb{E}[G_1'(Z)Z] \tilde{\mathfrak{D}}, \tilde{\mathfrak{E}} - \frac{1}{2} \mathbb{E}[G_2'(Z)Z] \tilde{\mathfrak{D}} \right) \end{aligned}$$

where  $(\tilde{\mathfrak{A}}, \tilde{\mathfrak{D}}, \tilde{\mathfrak{E}}, \tilde{\mathfrak{E}})$  is jointly Gaussian with mean 0 and covariance matrix

$$\begin{pmatrix} \text{Cov}(F(Z), F(Z)) & \text{Cov}(F(Z), Z^2) & \text{Cov}(F(Z), G_1(Z)) & \text{Cov}(F(Z), G_2(Z)) \\ \text{Cov}(Z^2, F(Z)) & \text{Cov}(Z^2, Z^2) & \text{Cov}(Z^2, G_1(Z)) & \text{Cov}(Z^2, G_2(Z)) \\ \text{Cov}(G_1(Z), F(Z)) & \text{Cov}(G_1(Z), Z^2) & \text{Cov}(G_1(Z), G_1(Z)) & \text{Cov}(G_1(Z), G_2(Z)) \\ \text{Cov}(G_2(Z), F(Z)) & \text{Cov}(G_2(Z), Z^2) & \text{Cov}(G_2(Z), G_1(Z)) & \text{Cov}(G_2(Z), G_2(Z)) \end{pmatrix},$$

and  $Z$  is a standard normal random variable.

This result is similar in spirit to [20, Theorem 1.1], which establishes a central limit theorem for the sequence of  $q$ -norms of  $\sqrt{n}\Theta^n$ ,  $n \in \mathbb{N}$ . Lemma 4.4 above provides fluctuation estimates for suitable joint functionals of  $\sqrt{n}\Theta^n$ , for which we first apply a Taylor expansion to the functionals. The proof of Lemma 4.4 is deferred to Appendix B.

## 5 Proof of the sharp large deviation estimate for spheres

Throughout this section, fix  $p \in (1, \infty)$  and for  $n \in \mathbb{N}$ , recall from Section 2.4 the definition of the two-dimensional random vector  $\bar{S}^n := \bar{S}^{(n,p)} = \frac{1}{n} \sum_{j=1}^n (\sqrt{n} \Theta_j^n Y_j, |Y_j|^p)$ , where  $(Y_j)_{j \in \mathbb{N}}$  is an i.i.d. sequence of random variables with common density  $f_p$  as in (2.3), and for  $\theta \in \mathbb{S}$ , let  $\bar{h}_\theta^n$  denote the (joint) density of  $\bar{S}^n$  under  $\mathbb{P}_\theta$ , where in this section we will typically suppress the dependence of  $\bar{h}_\theta^n$ ,  $\bar{S}^n$  and  $Y_j$  and other quantities on  $p$ . In view of (2.28), we then have

$$\mathbb{P}_\theta \left( W^{(n,p)} > a \right) = \mathbb{P}_\theta \left( \bar{S}^{(n,p)} \in \bar{D}_{p,a} \right) = \int_{\bar{D}_a} \bar{h}_\theta^n(x, y) dx dy, \quad (5.1)$$

where  $\bar{D}_a = \bar{D}_{p,a}$  is the domain defined in (2.29).

**Remark 5.1.** Note that  $\bar{h}_\theta^n$  depends on  $\theta$  only through  $\theta^n$ . For notational simplicity throughout we will adopt the convention that for quantities that depend on both  $n$  and  $\theta^n$ , we will use a superscript  $n$  to denote the former dependence and a subscript  $\theta$  instead of  $\theta^n$  to denote the dependence on  $\theta^n$ .

The key ingredients required to estimate the tail probability in (5.1) are an asymptotic expansion for the joint density  $\bar{h}_\theta^n$  carried out in Proposition 5.4 of Section 5.2, a multi-dimensional generalized Laplace approximation stated in Proposition 5.7 of Section 5.3, and a certain estimate that justifies the application of this Laplace approximation that is stated in Lemma 5.8 of Section 5.4. The proof of Proposition 5.4 is somewhat involved and hence deferred to Section 7. Instead, these results are first used in Sections 5.5 and 5.6 to prove Theorem 2.5. We first state a preliminary result in Section 5.1.

### 5.1 Estimates on the joint logarithmic moment generating function

We obtain an estimate on the growth of the log moment generating function  $\Lambda_p$  of  $(Y_j, |Y_j|^p)$  defined in (2.4), which will be useful in the subsequent discussion. The following expression was established in [16, Lemma 5.7]:

$$\Lambda_p(t_1, t_2) = -\frac{1}{p} \log(1 - pt_2) + \log M_{\gamma_p} \left( \frac{t_1}{(1 - pt_2)^{1/p}} \right), \quad (5.2)$$

for

$$(t_1, t_2) \in \mathbb{D}_p := \{(t_1, t_2) \in \mathbb{R}^2 : t_2 < 1/p\}, \quad (5.3)$$

where

$$M_{\gamma_p}(t) := \mathbb{E} [e^{tY_j}], \quad t \in \mathbb{R}, \quad (5.4)$$

is the moment generating function of  $Y_j$ . In order to understand the growth in  $t_1$  of the derivatives of  $\Lambda_p$ , it suffices to understand the derivatives of  $\log M_{\gamma_p}$ .

**Lemma 5.2.** For  $1 < p < \infty$ , let  $M_{\gamma_p}$  and  $\Lambda_p$  be as defined in (5.4) and (2.4), respectively. Then for every  $k \in \mathbb{N} \cup \{0\}$ , the derivative

$$t \mapsto \frac{d^k}{dt^k} \log M_{\gamma_p}(t),$$

exists and has at most polynomial growth, in the sense of Definition 4.1. Therefore, for  $j, k \in \mathbb{N} \cup \{0\}$ , and any  $t_2 < 1/p$ , the function

$$t_1 \mapsto \partial_1^j \partial_2^k \Lambda_p(t_1, t_2)$$

has at most polynomial growth.

The proof of Lemma 5.2 involves conceptually straightforward (though detailed) estimates, and is thus deferred to Appendix C.

### 5.2 An asymptotic expansion for the joint density

The main result of this section is Proposition 5.4, which provides an asymptotic expansion for the joint density  $\bar{h}_\theta^n$  of the two-dimensional random vector  $\bar{S}^n$  under  $\mathbb{P}_\theta$ . To state the result, for  $n \in \mathbb{N}$ , define

$$\bar{V}_j^n := (\sqrt{n}\Theta_j^n Y_j, |Y_j|^p), \quad j = 1, \dots, n. \tag{5.5}$$

For  $t = (t_1, t_2) \in \mathbb{C}^2$ , the Laplace transform of  $(Y_j, |Y_j|^p)$  is given by

$$\Phi_p(t_1, t_2) := \mathbb{E} \left[ e^{t_1 Y_j + t_2 |Y_j|^p} \right]. \tag{5.6}$$

The observation  $|e^{t_1 Y_j + t_2 |Y_j|^p}| = e^{\operatorname{Re}\{t_1\}Y_j + \operatorname{Re}\{t_2\}|Y_j|^p}$  shows that  $\Phi_p$  is finite precisely when  $\operatorname{Re}\{t_2\} < 1/p$ , or equivalently,  $(\operatorname{Re}\{t_1\}, \operatorname{Re}\{t_2\})$  lies in  $\mathbb{D}_p$ , the effective domain of  $\Lambda_p$  defined in (5.3). For  $t = (t_1, t_2) \in \mathbb{D}_p$  and  $\theta \in \mathbb{S}$ , also define

$$\Psi_{p,\theta}^n(t) := \frac{1}{n} \sum_{j=1}^n \log \Phi_p(\sqrt{n}\theta_j^n t_1, t_2) = \int_{\mathbb{R}} \log \Phi_p(ut_1, t_2) L_\theta^n(du), \tag{5.7}$$

where  $L_\theta^n$  is the empirical measure of the coordinates of  $\sqrt{n}\theta^n$ , as defined in (4.2).

**Remark 5.3.** Since  $\log \Phi_p = \Lambda_p$  on  $\mathbb{D}_p$ , for  $(t_1, t_2) \in \mathbb{D}_p$ ,  $\mathbb{R} \ni u \mapsto \log \Phi_p(ut_1, t_2)$  is continuous and has polynomial growth by Lemma 5.2. Hence, for every  $t = (t_1, t_2) \in \mathbb{D}_p$  and  $\sigma$ -a.e.  $\theta$ , the convergence of  $L_\theta^n$  to  $\gamma_2$  established in Lemma 4.3 shows that as  $n \rightarrow \infty$ ,

$$\Psi_{p,\theta}^n(t) \rightarrow \int_{\mathbb{R}} \log \Phi_p(ut_1, t_2) \gamma_2(du) = \Psi_p(t),$$

where the last equality holds by the definition of  $\Psi_p$  given in (2.5).

Next, recall the definition of  $\mathbb{J}_p$  from (2.7) and for  $x \in \mathbb{J}_p$ , the definition of  $\lambda_x$  from (2.7)–(2.8). Then for  $\theta \in \mathbb{S}$ , define

$$R_x^n(\theta^n) := \sqrt{n}(\Psi_{p,\theta}^n(\lambda_x) - \Psi_p(\lambda_x)), \tag{5.8}$$

and

$$c_x^n(\theta^n) := \sqrt{n} \nabla (\Psi_{p,\theta}^n(\lambda_x) - \Psi_p(\lambda_x)), \quad \mathcal{H}_x^n(\theta^n) := \operatorname{Hess} \Psi_{p,\theta}^n(\lambda_x), \tag{5.9}$$

where we drop the explicit dependence on  $p$  from  $c_x^n$ ,  $\mathcal{H}_x^n$  and  $R_x^n$ , and note that the right-hand sides above depend on  $\theta$  only through  $\theta^n$  (see Remark 5.1).

For  $a > 0$ , with the same abuse of notation used for  $\mathcal{H}_a$  in Section 2.2, we let  $c_a^n$  and  $R_a^n$  denote the functions  $c_{a^*}^n$  and  $R_{a^*}^n$ , respectively, where  $a^* = (a, 1)$ . We show in Section 7.2 that  $c_x^n(\theta^n)$  and  $\mathcal{H}_x^n(\theta^n)$  are the mean vector and covariance matrix, respectively, of  $\frac{1}{\sqrt{n}} \sum_{j=1}^n (\bar{V}_j^n - x)$ , with  $\bar{V}_j^n$  as in (5.5), under a certain quenched tilted measure; see (7.12) and (7.13).

**Proposition 5.4.** Fix  $p \in (1, \infty)$ ,  $n \in \mathbb{N}$ , and recall the definitions of  $\Psi_p$ ,  $\Psi_p^*$ ,  $\mathbb{J}_p$  and  $\Psi_{p,\theta}^n$  given in (2.5), (2.6), (2.7) and (5.7), respectively, and for  $x \in \mathbb{J}_p$ , recall the definitions of  $\mathcal{H}_x$ ,  $c_x^n(\cdot)$  and  $R_x^n(\cdot)$  from (2.12), (5.9) and (5.8), respectively. Then for  $\sigma$ -a.e.  $\theta$ ,

$$\bar{h}_\theta^n(x) = \frac{n}{2\pi} \bar{g}_\theta^n(x) e^{-n\Psi_p^*(x)} (1 + o(1)), \tag{5.10}$$

and the expansion in (5.10) is uniform on any compact subset of  $\mathbb{J}_p$ , where  $\bar{g}_\theta^n$  is the infinitely differentiable function defined by

$$\bar{g}_\theta^n(x) := (\det \mathcal{H}_x)^{-1/2} e^{\sqrt{n}R_x^n(\theta^n)} e^{-\|\mathcal{H}_x^{-1/2} c_x^n(\theta^n)\|^2}. \tag{5.11}$$

Section 7 is devoted to establishing Proposition 5.4, with the final proof given in Section 7.4.

### 5.3 A multi-dimensional generalized Laplace approximation

The formula (5.1) and the expression for  $\bar{h}_\theta^n$  in (5.10)–(5.11) show that the tail probability can be expressed as a Laplace-type integral over the domain  $\bar{D}_a$  defined in (2.29). However, to estimate this integral, we cannot directly apply conventional Laplace approximations such as those in [6, Chapter 8] or [42, Chapter V] due to the additional dependence of  $n$  in  $\bar{g}_\theta^n$ . Instead, in Propositions 5.6 and 5.7, we first establish a generalization of multi-dimensional Laplace approximations that can accommodate such  $n$ -dependent terms, which may be of independent interest.

**Definition 5.5.** Given  $m, d \in \mathbb{N}$ ,  $\alpha \in (0, 1)$  and a bounded domain  $D \subset \mathbb{R}^{m+d}$ , we say that the sequence  $h^n : \mathbb{R}^{m+d} \rightarrow \mathbb{R}$ ,  $n \in \mathbb{N}$ , admits a  $(f, x^*, \alpha, g^n)$ -representation if for each  $n \in \mathbb{N}$ ,

$$h^n(x) = g^n(x)e^{-nf(x)}, \quad x \in \mathbb{R}^{m+d},$$

where

1.  $f$  is a nonnegative function that is twice continuously differentiable in  $D$  and achieves its minimum on  $\text{cl}(D)$ , the closure of  $D$ , at a unique point  $x^*$ ,
2. there exists  $C \in (0, \infty)$  such that for each  $n \in \mathbb{N}$  sufficiently large,  $g^n(x) = \exp(r^n(x))$  is continuously differentiable with

$$|r^n(x)| \leq Cn^\alpha \|x\|_2 \quad \text{for all } x \text{ in a neighborhood of } x^*.$$

We start by establishing a Laplace asymptotics result, which extends the one-dimensional result in [31, Chapter 9.2].

**Proposition 5.6.** Given  $m, d \in \mathbb{N}$ , and a bounded domain  $D \subset \mathbb{R}^m \times \mathbb{R}_+^d$  containing the origin. Suppose the sequence  $h^n : \mathbb{R}^{m+d} \rightarrow \mathbb{R}$ ,  $n \in \mathbb{N}$ , admits a  $(f, x^*, \alpha, g^n)$ -representation on  $D$  with  $x^* = (0, 0, \dots, 0)$ . Then we have the following asymptotic expansion:

$$\int_D h^n(x) dx = \frac{(2\pi)^{\frac{m}{2}}}{n^{d+\frac{m}{2}}} \frac{g^n(x^*)}{\prod_{i=1}^d \partial_{m+i} f(x^*) \sqrt{\prod_{j=1}^m |\partial_{j,j}^2 f(x^*)|}} e^{-nf(x^*)} (1 + o(1)). \quad (5.12)$$

The proof is deferred to Appendix E. We now obtain an alternative representation for this integral. To state the result we need to introduce the definition of Weingarten maps. Let  $\mathcal{D}$  be a hypersurface in  $\mathbb{R}^d$ . Denote the tangent space at a point  $x \in \mathcal{D}$  to be  $T_x(\mathcal{D})$  and the normal vector field at  $x$  to be  $N_x$ . Then the Weingarten map at  $x$  is defined to be the linear map  $L_x : T_x(\mathcal{D}) \rightarrow T_x(\mathcal{D})$  where  $L_x(v) := \partial_v N_x$  and  $\partial_v$  is the directional derivative in the direction of  $v$ . Also, for a map  $L$ , let  $L^{-1}$  denote its inverse and recall that  $\det(A)$  denotes the determinant of a matrix  $A$ . (See also [2, Section 4] for more information on Weingarten maps).w

**Proposition 5.7.** For  $m, n \in \mathbb{N}$ , let  $\mathcal{D} \subset \mathbb{R}^{m+d}$  be a bounded domain whose boundary is a differentiable  $(d - 1)$ -dimensional hypersurface. Let  $h^n : \mathbb{R}^{m+d} \rightarrow \mathbb{R}$ ,  $n \in \mathbb{N}$ , be a sequence of functions that admits a  $(f, x^*, \alpha, g^n)$ -representation on  $\mathcal{D}$  in the sense of Definition 5.5. Then

$$\mathcal{I}^n := \int_{\mathcal{D}} h^n(x) dx = \frac{(2\pi)^{(d-1)/2} \det(L_1^{-1}(L_1 - L_2))^{-1/2}}{n^{(d+1)/2} \langle (\text{Hess } f(x^*))^{-1} \nabla f(x^*), \nabla f(x^*) \rangle^{1/2}} g^n(x^*) e^{-nf(x^*)} (1 + o(1)),$$

where for  $i = 1, 2$ ,  $L_i$  is the Weingarten map at  $x^* \in \partial\mathcal{D}$  of the surface  $\mathcal{C}_i$ , given by

$$\mathcal{C}_1 := \{y : f(y) = f(x^*)\} \quad \text{and} \quad \mathcal{C}_2 := \partial\mathcal{D}.$$

*Proof.* The proof will make use of arguments from [6] as well as a result from [2]. Since  $\partial\mathcal{D}$  is a differentiable  $(d - 1)$ -dimensional hypersurface, there exists a one-to-one continuously differentiable transformation  $\Gamma : \mathcal{N} \rightarrow \tilde{\mathcal{N}} \subset \mathbb{R} \times \mathbb{R}_+^{d-1}$  such that  $F$  maps  $x^*$  to the origin. Setting  $J_\Gamma(x) \in \mathbb{R}^d \times \mathbb{R}^d$  to be the Jacobian matrix of  $\Gamma$  at  $x$ , we can write

$$\mathcal{I}^n := \int_{\mathcal{D}} g^n(x) e^{-nf(x)} dx = \int_{\Gamma(\mathcal{D})} |\det J_\Gamma(x)| g^n(\Gamma^{-1}(x)) e^{-nf(\Gamma^{-1}(x))} dx.$$

By the assumption in Definition 5.5 there exist  $\alpha \in (0, 1)$  and  $C \in (0, \infty)$  such that  $g^n(x) = \exp(r^n(x))$  with  $|r^n(x)| \leq Cn^\alpha \|x\|_2$  on a neighborhood of  $\mathcal{D}$ . By the differentiability of  $\Gamma$ , we have  $|r^n(\Gamma^{-1}(x))| \leq Cn^\alpha \|x\|_2$ . Hence, Proposition 5.6 with  $m, d, g^n, f$  and  $\mathcal{D}$  therein replaced with  $d - 1, 1, |\det J_\Gamma(x)| g^n(\Gamma^{-1}(x)), f \circ \Gamma^{-1}$  and  $\Gamma(\mathcal{D})$ , respectively, implies there exists a constant  $C' = C'(\Gamma, \mathcal{D}, f) \in (0, \infty)$  that does not depend on  $g^n$  such that

$$\mathcal{I}^n = \frac{(2\pi)^{(d-1)/2} C'}{n^{(d+1)/2}} g^n(x^*) e^{-nf(x^*)} (1 + o(1)). \tag{5.13}$$

In order to deduce the constant  $C'$ , we note that the same formula also holds when  $g^n \equiv 1$  and hence it follows that

$$\tilde{\mathcal{I}}^n := \int_{\mathcal{D}} e^{-nf(x)} dx = \frac{(2\pi)^{(d-1)/2} C'}{n^{(d+1)/2}} e^{-nf(x^*)} (1 + o(1)). \tag{5.14}$$

Also note that  $\tilde{\mathcal{I}}^n$  coincides with the integral in [6, Equation (8.3.63)] when  $\lambda, n, \phi$  and  $g_0$  therein are replaced with  $n, d, -f$  and 1 here. By the stated properties of  $\mathcal{D}$ , there exists a local chart of a coordinate system  $\mathcal{G} : \mathcal{N}_x \rightarrow \mathcal{U}$  of  $\mathcal{N}_x \subset \partial\mathcal{D}$  around  $x^*$ , for some subset  $\mathcal{U} \subset \mathbb{R}^{d-1}$ . Let  $\mathcal{J}_*$  be the Jacobian matrix of the transformation  $\mathcal{G}$  at  $x^*$ , and let  $\mathcal{J}_*^T$  denote its transpose. Then, under the stated conditions on  $f$  and  $\mathcal{D}$ , the formula [6, Equation (8.3.63)] yields the following estimate:

$$\tilde{\mathcal{I}}^n = \frac{(2\pi)^{(d-1)/2} |\det(\mathcal{J}_*^T \mathcal{J}_*)|^{1/2}}{n^{(d+1)/2} |\det \text{Hess}(f \circ \mathcal{G}(x^*))|^{1/2} |\nabla f(x^*)|} e^{-nf(x^*)} (1 + o(1)). \tag{5.15}$$

Next, to further simplify the expression in the last display, by [2, Equations (4.5) and (4.6)] it follows, after identifying  $DG(0), I$  and  $A$  therein with  $\mathcal{J}_*, f$  and  $\text{Hess}(f \circ \mathcal{G}(x^*))$ , respectively, that

$$\frac{|\det(\mathcal{J}_*^T \mathcal{J}_*)|^{1/2}}{|\det \text{Hess}(f \circ \mathcal{G}(x^*))|^{1/2} |\nabla f(x^*)|} = \frac{\det(L_1^{-1}(L_1 - L_2))^{-1/2}}{\langle (\text{Hess } f(x^*))^{-1} \nabla f(x^*), \nabla f(x^*) \rangle^{1/2}}, \tag{5.16}$$

with  $L_1, L_2$  as in the proposition. (Note that there is an erroneous additional factor of  $\sqrt{2\pi n}$  in the denominator of the expression in [2, Equation (4.6)], which we have corrected). Comparing (5.15), (5.16) and (5.14), we see that

$$C' = \frac{\det(L_1^{-1}(L_1 - L_2))^{-1/2}}{\langle (\text{Hess } f(x^*))^{-1} \nabla f(x^*), \nabla f(x^*) \rangle^{1/2}}. \tag{5.17}$$

The proposition then follows on substituting the above expression for  $C'$  into (5.13).  $\square$

#### 5.4 Continuity estimates for terms in the prefactor

In order to apply Proposition 5.7 to the expression for  $\bar{h}_\theta^n$  given in (5.10)–(5.11), we need to verify that  $\bar{h}_\theta^n$  satisfies Definition 5.5. The following lemma will be useful in verifying property (2) of Definition 5.5.

**Lemma 5.8.** Fix  $p \in (1, \infty)$ . For every  $y \in \mathbb{J}_p$ ,  $\alpha \in (1/2, 1)$  and  $\varepsilon > 0$  with a fixed  $B_\varepsilon(y) \subset \mathbb{J}_p$ , there exists  $\Xi = \Xi(y, \alpha, \varepsilon) \in (0, \infty)$  such that for  $\sigma$ -a.e.  $\theta$ , there exists  $N = N(y, \alpha, \varepsilon, \theta) \in \mathbb{N}$  such that for  $x \in B_\varepsilon(y)$  and  $n \geq N$ ,

$$|\sqrt{n}R_x^n(\theta^n) - \sqrt{n}R_y^n(\theta^n)| \leq \Xi n^\alpha \|x - y\|_2, \tag{5.18}$$

and

$$\left| \left\| \mathcal{H}_x^{-1/2} c_x^n(\theta^n) \right\|_2^2 - \left\| \mathcal{H}_y^{-1/2} c_y^n(\theta^n) \right\|_2^2 \right| \leq \Xi n^{2\alpha-1} \|x - y\|_2. \tag{5.19}$$

Before presenting the proof of the lemma, we provide an alternative formulation of (5.18) and a related result. Fix  $\alpha \in (1/2, 1)$ ,  $p \in (1, \infty)$ ,  $y \in \mathbb{J}_p$  and  $\varepsilon > 0$  such that  $B_\varepsilon(y) \subset \mathbb{J}_p$ . From (2.4), (5.6) and the fact that  $Y_j$  is a  $p$ -Gaussian random variable, it follows that the equality  $\Lambda_p = \log \Phi_p$  holds on the domain  $\mathbb{D}_p$  of  $\Phi_p$  defined in (5.3). When combined with (5.8), (5.7) and (2.5), this shows that

$$\sqrt{n}R_x^n(\theta^n) = \sum_{j=1}^n (\log \Phi_p(\sqrt{n}\theta_j^n \lambda_{x,1}, \lambda_{x,2}) - \mathbb{E}[\log \Phi_p(Z\lambda_{x,1}, \lambda_{x,2})]),$$

where  $Z \sim \gamma_2$  is a standard normal random variable. Since  $x \mapsto \lambda_x$  is infinitely differentiable by Remark 2.1, there exists  $C' \in (0, \infty)$  such that  $\|\lambda_x - \lambda_y\| \leq C' \|x - y\|$  for  $x \in B_\varepsilon(y)$ . Therefore, to show (5.18), it suffices to show that given any fixed  $s = (s_1, s_2) \in \mathbb{D}_p$ , for every  $\varepsilon' > 0$  such that  $B_{\varepsilon'}(s) \subset \mathbb{D}_p$ , there exist  $C = C(s, \alpha, \varepsilon') \in (0, \infty)$  and a random integer  $N = N(s, \alpha, \varepsilon')$  such that  $\mathbb{P}$ -almost surely,

$$\left| \sum_{j=1}^n (\mathcal{K}_s(\sqrt{n}\Theta_j^n t_1, t_2) - \mathbb{E}[\mathcal{K}_s(Zt_1, t_2)]) \right| \leq C n^\alpha \|t\|, \quad \text{for } \|t\| < \varepsilon' \text{ and } n \geq N, \tag{5.20}$$

where

$$\mathcal{K}_s(t_1, t_2) := \log \Phi_p((s_1 + t_1), (s_2 + t_2)) - \log \Phi_p(s_1, s_2), \tag{5.21}$$

for

$$(t_1, t_2) \in \mathbb{D}_{p,s} := \{(u_1, u_2) \in \mathbb{R}^2 : u_2 < 1/p - s_2\}. \tag{5.22}$$

Let  $Z, (Z_j)_{j \in \mathbb{N}}$  be independent standard Gaussian random variables on the probability space  $(\Omega', \mathcal{F}', \mathbb{P}')$ . Then letting  $Z^{(n)} := (Z_1, \dots, Z_n)$  we have (e.g. see Section 2.4 or [35, Lemma 1]),

$$(\Theta_1^n, \dots, \Theta_n^n) \stackrel{(d)}{=} \frac{(Z_1, \dots, Z_n)}{\|Z^{(n)}\|}. \tag{5.23}$$

The following result on Gaussian vectors will be used to prove Lemma 5.8.

**Lemma 5.9.** With the notation above, fix  $\alpha \in (1/2, 1)$ , let  $\mathbb{D} = \mathbb{R} \times (-\infty, T)$  and let  $\mathcal{K} : \mathbb{D} \rightarrow \mathbb{R}$  be a twice continuously differentiable function. Suppose for  $t_2 \in (-\infty, T)$ , the mappings  $t_1 \mapsto \partial_1 \mathcal{K}(t_1, t_2)$ ,  $t_1 \mapsto \partial_{12} \mathcal{K}(t_1, t_2)$  and  $t_1 \mapsto \partial_{11} \mathcal{K}(t_1, t_2)$  have polynomial growth in the sense of Definition 4.1. Then for every  $\varepsilon > 0$  such that  $B_\varepsilon(0) \subset \mathbb{D}$ , there exist  $C = C(s, \alpha, \varepsilon) \in (0, \infty)$  and a random integer  $N = N(s, \alpha, \varepsilon)$  on  $(\Omega', \mathcal{F}', \mathbb{P}')$  such that  $\mathbb{P}'$ -almost surely, for

$$\left| \sum_{j=1}^n \left( \mathcal{K} \left( \frac{\sqrt{n}Z_j}{\|Z^{(n)}\|} t_1, t_2 \right) - \mathbb{E}[\mathcal{K}(Zt_1, t_2)] \right) \right| \leq C n^\alpha \|t\|, \quad \text{for } \|t\| < \varepsilon \text{ and } n \geq N. \tag{5.24}$$

Deferring the proof of Lemma 5.9 to Appendix F, we now use it to prove Lemma 5.8.

**Remark 5.10.** From (5.23), we will use in our proof the following equivalence that a statement about  $Z^{(n)}/\|Z^{(n)}\|$  holds  $\mathbb{P}'$ -almost surely if and only if the same statement with  $Z^{(n)}/\|Z^{(n)}\|$  replaced by  $\Theta^n$  holds  $\mathbb{P}$ -almost surely.

*Proof of Lemma 5.8.* We begin with the proof of (5.18). Since  $\Phi_p$  is finite near the origin,  $\Phi_p$  is infinitely differentiable on its domain  $\mathbb{D}_p$  and hence, for any  $s \in \mathbb{D}_p$ , the functional  $\mathcal{K}_s$  from (5.21) is twice continuously differentiable on its domain  $\mathbb{D}_{p,s}$  defined in (5.22). Since  $\log \Phi_p = \Lambda_p$  on  $\mathbb{D}_p$ , the expression in (5.21) and Lemma 5.2 imply that the mappings  $t_1 \mapsto \partial_1 \mathcal{K}_s(t_1, t_2)$ ,  $t_1 \mapsto \partial_{12} \mathcal{K}_s(t_1, t_2)$  and  $t_1 \mapsto \partial_{11} \mathcal{K}_s(t_1, t_2)$  have polynomial growth for  $t_2 < 1/p - s_2$ . Therefore, Lemma 5.9 implies that for each  $s \in \mathbb{D}_p$  and  $\alpha \in (1/2, 1)$ , for  $\varepsilon > 0$  with  $B_\varepsilon(0) \subset \mathbb{D}$ , there exist  $C = C(s, \alpha, \varepsilon) \in (0, \infty)$  and a random integer  $N = N(s, \alpha, \varepsilon)$  such that (5.24) holds  $\mathbb{P}'$ -almost surely. Due to the relation  $\sigma_n = \mathbb{P} \circ (\Theta^{(n)})^{-1}$ , Remark 5.10 implies that (5.20) holds and thus, that (5.18) also holds.

We now turn to the proof of (5.19). Fix  $s \in \mathbb{D}_p$ . For  $i = 1, 2$ , and  $(t_1, t_2) \in \mathbb{D}_{p,s}$ , define

$$\tilde{\mathcal{K}}_i(t_1, t_2) := \partial_i \log \Phi_p(t_1, t_2), \tag{5.25}$$

and

$$\mathcal{K}_{s,i}(t_1, t_2) := \tilde{\mathcal{K}}_i((s_1 + t_1), (s_2 + t_2)) - \tilde{\mathcal{K}}_i(s_1, s_2). \tag{5.26}$$

Note that for  $i = 1, 2$ , by the smoothness of  $\Phi_p$ ,  $\mathcal{K}_{s,i}$  is twice continuously differentiable on its domain  $\mathbb{D}_{p,s}$  and also the mappings  $t_1 \mapsto \partial_1 \mathcal{K}_{s,i}(t_1, t_2)$ ,  $t_1 \mapsto \partial_{12} \mathcal{K}_{s,i}(t_1, t_2)$  and  $t_1 \mapsto \partial_{11} \mathcal{K}_{s,i}(t_1, t_2)$  have polynomial growth for  $t_2 \in \mathbb{D}_{p,s}$  by Lemma 5.2 and the fact that  $\log \Phi_p = \Lambda_p$  on  $\mathbb{D}_p$ . Thus, for  $i = 1, 2$ ,  $\mathcal{K}_{s,i}$  satisfies the assumption in Lemma 5.9. Hence, (5.9), (5.7), (5.23), Lemma 5.9 and the equivalence between statements about  $Z^{(n)}/\|Z^{(n)}\|$  and  $\Theta^n$ , imply that for every  $y \in \mathbb{J}_p$ ,  $\alpha \in (1/2, 1)$  and  $\varepsilon > 0$  with  $B_\varepsilon(y) \subset \mathbb{J}_p$ , there exists  $\Xi = \Xi(y, \varepsilon, \alpha) \in (0, \infty)$ , and for  $\sigma$ -a.e.  $\theta$ , there exists  $N = N(y, \alpha, \varepsilon, \theta) \in \mathbb{N}$  such that for  $x \in B_\varepsilon(y)$  and  $n \geq N$ ,

$$|\sqrt{n}c_{x,i}^n(\theta^n) - \sqrt{n}c_{y,i}^n(\theta^n)| \leq \Xi n^\alpha \|x - y\|_2, \quad i = 1, 2, \tag{5.27}$$

where  $c_{y,i}^n(\theta^n)$  denotes the  $i$ -th coordinate of  $c_x^n(\theta^n)$ . By the smoothness of  $x \mapsto \lambda_x$  in Remark 2.1, there exists  $C' \in (0, \infty)$  such that  $\|\lambda_x\| \leq C'$  for  $x \in B_{\varepsilon'}(y)$ . Hence, (5.9) and Lemma 5.9, with  $\mathcal{K}$  replaced by  $\tilde{\mathcal{K}}_i$ , imply that for every  $y \in \mathbb{J}_p$  and  $\alpha \in (1/2, 1)$ , there exists  $\Xi' = \Xi'(y, \varepsilon, \alpha) \in (0, \infty)$ , and for  $\sigma$ -a.e.  $\theta$ , there exists  $N' = N'(y, \alpha, \varepsilon, \theta) \in \mathbb{N}$  such that for  $x \in B_\varepsilon(y)$  and  $n \geq N'$ ,

$$|\sqrt{n}c_{x,i}^n(\theta^n)| \leq \Xi' n^\alpha \|\lambda_x\|_2 \leq C' \Xi' n^\alpha, \quad i = 1, 2. \tag{5.28}$$

Recall the definition of  $\mathcal{H}_x$  in (2.12). The following relation

$$\left\| \mathcal{H}_x^{-1/2} c_x^n(\theta^n) \right\|_2^2 = \sum_{i=1}^2 \left( \sum_{j=1}^2 (\mathcal{H}_x^{-1/2})_{ij} c_{x,j}^n(\theta^n) \right)^2$$

implies that the left-hand side of (5.19) can be written as

$$\left| \left\| \mathcal{H}_x^{-1/2} c_x^n(\theta^n) \right\|_2^2 - \left\| \mathcal{H}_y^{-1/2} c_y^n(\theta^n) \right\|_2^2 \right| \leq \sum_{i=1}^2 A_i^n \tilde{A}_i^n, \tag{5.29}$$

where for  $i = 1, 2$ ,

$$A_i^n := \left| \sum_{j=1}^2 ((\mathcal{H}_x^{-1/2})_{ij} c_{x,j}^n(\theta^n) - (\mathcal{H}_y^{-1/2})_{ij} c_{y,j}^n(\theta^n)) \right|,$$

$$\tilde{A}_i^n := \left| \sum_{j=1}^2 ((\mathcal{H}_x^{-1/2})_{ij} c_{x,j}^n(\theta^n) + (\mathcal{H}_y^{-1/2})_{ij} c_{y,j}^n(\theta^n)) \right|.$$

By the smoothness of  $\Psi_p$  and  $\lambda_x$  in Remark 2.1,  $x \mapsto \mathcal{H}_x^{-1/2}$  is also infinitely differentiable and there exists  $C''' \in (0, \infty)$  such that  $|(\mathcal{H}_x^{-1/2})_{ij} - (\mathcal{H}_y^{-1/2})_{ij}| \leq C''' \|x - y\|$  and  $|(\mathcal{H}_x^{-1/2})_{ij}| \leq C'''$  for  $x \in B_\varepsilon(y)$  and  $i, j = 1, 2$ . Hence, by (5.27) and (5.28), for  $\sigma$ -a.e.  $\theta$ ,  $x \in B_\varepsilon(y)$  and  $n \geq \max\{N, N'\}$ ,

$$\begin{aligned} A_i^n &\leq \sum_{j=1}^2 \left| ((\mathcal{H}_x^{-1/2})_{ij} c_{x,j}^n(\theta^n) - (\mathcal{H}_y^{-1/2})_{ij} c_{y,j}^n(\theta^n)) \right| \\ &\leq \left| (\mathcal{H}_x^{-1/2})_{ij} - (\mathcal{H}_y^{-1/2})_{ij} \right| \sum_{j=1}^2 |c_{x,j}^n(\theta^n)| + \left| (\mathcal{H}_x^{-1/2})_{ij} \right| \sum_{j=1}^2 |c_{x,j}^n(\theta^n) - c_{y,j}^n(\theta^n)| \\ &\leq 2(C''' C' \Xi' + C'' \Xi) n^{\alpha-1/2} \|x - y\|_2. \end{aligned} \tag{5.30}$$

Next, by (5.28), for  $\sigma$ -a.e.  $\theta$ ,  $x \in B_\varepsilon(y)$  and  $n \geq N'$ ,

$$\tilde{A}_i^n \leq 2C'' C' \Xi' n^{\alpha-1/2}. \tag{5.31}$$

Therefore (5.19) follows from (5.29)–(5.31).  $\square$

### 5.5 Proof of Theorem 2.5(i)

We are now ready to prove the main estimate (2.18). Fix  $p \in (1, \infty)$  and  $a > 0$  such that  $\mathbb{I}_p(a) < \infty$  and recall the definition of the domain  $\bar{D}_a = \bar{D}_{p,a}$  given in (2.29). Since  $\mathbb{I}_p(a)$  is convex and symmetric,  $\mathbb{I}_p(a)$  is increasing for  $a \in \mathbb{R}_+$ . Thus, (2.10) and Lemma 2.2 imply that

$$\inf_{x \in \bar{D}_a} \Psi_p^*(x) = \inf_{t > a} \mathbb{I}_p(t) = \mathbb{I}_p(a) = \inf_{\tau_1 \in \mathbb{R}, \tau_2 > 0: \tau_1 \tau_2^{-1/p} = a} \Psi_p^*(\tau_1, \tau_2) = \Psi_p^*(a, 1).$$

Hence, the infimum of  $\Psi_p^*$  over the closure  $\text{cl}(\bar{D}_a)$  of  $\bar{D}_a$  is attained at  $a^* := (a, 1)$ . Moreover due to (2.11), the assumption  $\mathbb{I}_p(a) < \infty$  implies  $\Psi_p^*(a, 1) < \infty$ , and hence,  $a^* = (a, 1) \in \mathbb{J}_p$ , defined in (2.7). Further, by (2.29),  $a^*$  is a point on the smooth part of the the boundary  $\partial \bar{D}_a$  of  $\bar{D}_a$ . Let  $U \subset \mathbb{R}_+^2 := \{(x, y) \in \mathbb{R}^2 : x > 0, y > 0\}$  be an open neighborhood of  $a^*$  to be chosen below and note that the boundary of  $U \cap \bar{D}_a$  is also smooth at  $a^*$ . Then, for  $\theta \in \mathbb{S}$ , we can split the probability of interest from (5.1) into two parts:

$$\mathbb{P}_\theta(\bar{S}^n \in \bar{D}_a) = \mathbb{P}_\theta(\bar{S}^n \in \bar{D}_a \cap U) + \mathbb{P}_\theta(\bar{S}^n \in \bar{D}_a \cap U^c). \tag{5.32}$$

The proof will proceed in two steps. In the key first step, we will estimate the first term on the right-hand side of (5.32) by integrating the estimate of the density  $\bar{h}_\theta^n$  of  $\bar{S}^n$  obtained in Proposition 5.4 over the domain  $\bar{D}_a \cap U$ , and then analyze the asymptotics of the resulting Laplace type integral, as  $n \rightarrow \infty$  using Proposition 5.4, Proposition 5.7 and Lemma 5.8. The second step will involve using the LDP for  $(\bar{S}^n)_{n \in \mathbb{N}}$  to show that the second term on the right-hand side of (5.32) is negligible.

**Step 1.** Using the expressions for  $\bar{h}_\theta^n$  from (5.10) and the fact that the domain  $\bar{D}_a \cap U \subset \mathbb{R}^2$  is bounded, we have for  $\sigma$ -a.e.  $\theta$ ,

$$\mathbb{P}_\theta(\bar{S}^n \in \bar{D}_a \cap U) = \int_{\bar{D}_a \cap U} \bar{h}_\theta^n(x) dx = \frac{n}{2\pi} \mathcal{I}_\theta^n(1 + o(1)), \tag{5.33}$$

where

$$\mathcal{I}_\theta^n := \int_{\bar{D}_a \cap U} \bar{g}_\theta^n(x) e^{-n\Psi_p^*(x)} dx, \tag{5.34}$$

where  $\bar{g}_\theta^n$  is as defined in (5.11).

To apply Proposition 5.7, we first prove the following:

**Lemma 5.11.** *For  $\alpha \in (1/2, 1)$ , the function  $\bar{g}_\theta^n(x) e^{-n\Psi_p^*(x)}$  admits a  $(\Psi_p^*, a^*, \alpha, \bar{g}_\theta^n)$ -representation on the bounded region  $\bar{D}_a \cap U$ .*

*Proof.* To verify property (1) of Definition 5.5, first note that  $\Psi_p^*$  is nonnegative since it is a rate function by Theorem 2.3. Next note that by (2.5), (5.2) and Lemma 5.2,  $\Psi_p$  is twice (in fact infinitely) differentiable on  $\mathbb{D}_p = \mathbb{R} \times \{t_2 : t_2 < 1/p\}$ . Hence, by the duality of the Legendre transform [43, Section III.D], it follows that  $\Psi_p^*$  is twice differentiable in  $\bar{D}_a$  and achieves its minimum uniquely at  $a^* = (a, 1) \in \partial(\bar{D}_a \cap U)$ . Thus, property (1) of Definition 5.5 holds.

We next turn to the verification of property (2) of Definition 5.5. From (5.11), it follows that  $\bar{g}_\theta^n(x) = \exp(r_\theta^n(x))$ , where

$$r_\theta^n(x) := \log \det \mathcal{H}_x^{-1/2} + \sqrt{n} R_x^n(\theta^n) + \left\| \mathcal{H}_x^{-1/2} c_x^n(\theta^n) \right\|^2, \quad x \in \bar{D}_a \cap U.$$

Lemma 5.8 and the smoothness of  $x \mapsto \log \mathcal{H}_x^{-1/2}$ , which follows from Remark 2.1 and (2.12), imply that for a sufficiently small neighborhood of  $x^*$ , for any  $\alpha \in (1/2, 1)$ , there exist  $C \in (1, \infty)$  and a finite random variable  $N$  such that for  $\sigma$ -a.e.  $\theta$ ,  $r_\theta^n$  satisfies property (2) of Definition 5.5 and the claim follows.  $\square$

Given the claim, Proposition 5.7 applied with  $d = 2$ ,  $\mathcal{D} = \bar{D}_a \cap U$  and  $h^n(x) = \bar{g}_\theta^n(x) e^{-n\Psi_p^*(x)}$  shows that for  $\sigma$ -a.e.  $\theta$ ,

$$\mathcal{I}_\theta^n = \frac{(2\pi)^{1/2}}{n^{3/2}} \frac{(L_{a,1}^{-1}(L_{a,1} - L_{a,2}))^{-1/2}}{\langle (\text{Hess } \Psi_p^*(a^*))^{-1} \nabla \Psi_p^*(a^*), \nabla \Psi_p^*(a^*) \rangle^{1/2}} \bar{g}_\theta^n(a^*) e^{-n\Psi_p^*(a^*)} (1 + o(1)), \tag{5.35}$$

where  $L_{a,1}$  and  $L_{a,2}$  are the Weingarten maps of the curves  $\mathcal{C}_1 := \{x \in \mathbb{R}^2 : \Psi_p^*(x) = \Psi_p^*(a, 1)\}$  and  $\mathcal{C}_2 := \{x \in \mathbb{R}^2 : x_1 = ax_2^{1/p}\}$ , evaluated at  $a^* = (a, 1)$ . To further simplify (5.35), first note that by the duality of the Legendre transform, and the definition of  $\lambda_{a,j}$  in (2.13), we have

$$\partial_j \Psi_p^*(a^*) = \lambda_{a,j}, \quad \text{for } j = 1, 2, \tag{5.36}$$

and

$$(\text{Hess } \Psi_p^*(a^*))^{-1} = \text{Hess } \Psi_p(\lambda_a) = \mathcal{H}_a. \tag{5.37}$$

Hence,

$$\mathcal{I}_\theta^n = \frac{(2\pi)^{1/2}}{n^{3/2}} \frac{(L_{a,1}^{-1}(L_{a,1} - L_{a,2}))^{-1/2}}{\langle \mathcal{H}_a \lambda_a, \lambda_a \rangle^{1/2}} \bar{g}_\theta^n(a^*) e^{-n\Psi_p^*(a^*)} (1 + o(1)). \tag{5.38}$$

Next, observe that [2, Example 4.3] shows that in  $\mathbb{R}^2$ , the Weingarten map is reduced to multiplication by the inverse of the radius of the osculating circle, which is equal to the absolute value of the curvature. Recall that for a curve in  $\mathbb{R}^2$  defined by the equation  $T(x, y) = 0$  for a sufficiently smooth map  $T : \mathbb{R}^2 \rightarrow \mathbb{R}$ , the curvature at a point  $x^*$  on the curve is given by the formula

$$\frac{T_y^2 T_{xx} - 2T_x T_y T_{xy} + T_x^2 T_{yy}}{(T_x^2 + T_y^2)^{3/2}}(x^*).$$

Thus, to calculate the curvature of the curve  $\mathcal{C}_1$  at  $a^*$ , use the above formula with  $T(x, y) = \Psi_p^*(x, y) - \Psi_p^*(a, 1)$  and  $x^* = a^*$ , and substitute the relations  $\partial_j \Psi_p^*(a^*) = \lambda_{a,j}$ ,  $j = 1, 2$ , and the definition of  $\mathcal{H}_a$  mentioned above to conclude that

$$L_{a,1} = \frac{|\lambda_{a,2}^2(\mathcal{H}_a^{-1})_{11} - 2\lambda_{a,1}\lambda_{a,2}(\mathcal{H}_a^{-1})_{12} + \lambda_{a,1}^2(\mathcal{H}_a^{-1})_{22}|}{(\lambda_{a,1}^2 + \lambda_{a,2}^2)^{3/2}}. \tag{5.39}$$

On the other hand, the curvature of the graph of a function  $y = \tilde{T}(x)$  at the point  $(x, \tilde{T}(x))$  for sufficiently smooth  $\tilde{T} : \mathbb{R} \rightarrow \mathbb{R}$  is given by  $|\tilde{T}''(x)| / (1 + (\tilde{T}'(x))^2)^{3/2}$ . Recalling the definition of  $\bar{D}_a$  from (2.29), we can apply this with  $\tilde{T}(x) = (x/a)^p$  to compute the curvature of  $\mathcal{C}_2 = \partial \bar{D}_a$  at  $a^*$  as:

$$L_{a,2} = \frac{p(p-1)a}{(a^2 + p^2)^{3/2}}. \tag{5.40}$$

Substituting these calculations back into the expressions (5.33) and (5.38), and recalling the definitions of  $\bar{g}_\theta^n$  from (5.11),  $C_a^n(\theta^n)$  from (2.19) and  $\xi_a$  and  $\kappa_a$  from (2.15) and (2.16), we conclude that for  $\sigma$ -a.e.  $\theta$ ,

$$\mathbb{P}_\theta(\bar{S}^n \in \bar{D}_a \cap U) = \frac{C_a^n(\theta^n)}{\kappa_a \xi_a \sqrt{2\pi n}} e^{-n\mathbb{I}_p(a) + \sqrt{n}R_a^n(\theta^n)}(1 + o(1)). \tag{5.41}$$

**Step 2.** We now turn to the second term in (5.32). Note that by the continuity of  $\Psi_p^*$ , there exists  $\eta > 0$  such that

$$\inf_{y \in \bar{D}_a \cap U^c} \Psi_p^*(y) > \Psi_p^*(a^*) + \eta.$$

By the refinement in Lemma 2.2 of the (quenched) large deviation principle for  $\bar{S}^n$  established in [16, Proposition 5.3],  $\Psi_p^*$  achieves its unique minimum in  $\bar{D}_a$  at  $a^* = (a, 1)$ . Thus, for  $\sigma$ -a.e.  $\theta$ ,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_\theta(\bar{S}^n \in \bar{D}_a \cap U^c) \leq -\Psi_p^*(a^*) - \eta, \tag{5.42}$$

which shows that the term in (5.42) is negligible with respect to (5.41).

When combined, (2.19), (5.8), (5.11), (5.32), (5.41) and (5.42) together yield (2.18). This completes the proof of Theorem (i).

### 5.6 Proof of Theorem 2.5(ii)

We start by obtaining expansions for  $R_a^n(\Theta^n)$  and  $c_a^n(\Theta^n)$ . First, note that the functions  $\ell_a$ ,  $\ell_{a,1}$  and  $\ell_{a,2}$  defined in (2.17) and their derivatives up to second order (for  $\ell_{a,1}$  and  $\ell_{a,2}$ ) and third order (for  $\ell_a$ ) are continuous and have at most polynomial growth by Lemma 5.2. Therefore, setting

$$r_n := \hat{r}_n(\ell_a), \quad s_n := \hat{s}_n(\ell_a), \quad t_{n,1} := \hat{r}_n(\ell_{a,1}), \quad t_{n,2} := \hat{r}_n(\ell_{a,2}),$$

where  $\hat{r}_n$  and  $\hat{s}_n$  are defined in (4.4) and (4.3), respectively, we can apply (5.9), (5.8) and Lemma 4.4 to obtain

$$\begin{aligned} R_a^n(\Theta^n) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\ell_a(\sqrt{n}\Theta_i^n) - \mathbb{E}[\ell_a(Z)]) \\ &\stackrel{(d)}{=} r_n + \frac{1}{\sqrt{n}} s_n + o\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

$$c_a^n(\Theta^n) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} \ell_{a,1}(\sqrt{n}\Theta_i^n) - \mathbb{E}[\ell_{a,1}(Z)] \\ \ell_{a,2}(\sqrt{n}\Theta_i^n) - \mathbb{E}[\ell_{a,2}(Z)] \end{pmatrix} \\ \stackrel{(d)}{=} \begin{pmatrix} t_{n,1} \\ t_{n,2} \end{pmatrix} + o(1).$$

Moreover, Lemma 4.4 also shows that we have the convergence

$$(r_n, s_n, t_{n,1}, t_{n,2}) \Rightarrow \left( \tilde{\mathfrak{A}} - \frac{1}{2} \mathbb{E}[\ell'_a(Z)Z] \tilde{\mathfrak{D}}, \frac{1}{8} \mathbb{E}[\ell''_a(Z)Z^2] \tilde{\mathfrak{D}}^2, \tilde{\mathfrak{E}} - \frac{1}{2} \mathbb{E}[\ell'_{a,1}(Z)Z] \tilde{\mathfrak{D}}, \tilde{\mathfrak{F}} - \frac{1}{2} \mathbb{E}[\ell'_{a,2}(Z)Z] \tilde{\mathfrak{D}} \right),$$

where  $(\tilde{\mathfrak{A}}, \tilde{\mathfrak{D}}, \tilde{\mathfrak{E}}, \tilde{\mathfrak{F}})$  is jointly Gaussian with mean 0 and covariance matrix (2.21).

## 6 Proof of the sharp large deviation estimate for balls

### 6.1 Preliminary notation

Fix  $p \in (1, \infty)$  and  $a > 0$  such that  $\mathbb{I}_p(a) < \infty$ . The definitions in Section 2.4, specifically (2.32), yield the following expression for the tail probability of projections of  $\ell_p^n$  balls:

$$\mathbb{P}_\theta \left( W^{(n,p)} > a \right) = \int_{\bar{\mathcal{D}}_a} \bar{h}_\theta^n(x_1, x_2, y) dx_1 dx_2 dy, \quad (6.1)$$

where for  $\theta \in \mathbb{S}$ ,  $\bar{h}_\theta^n(x_1, x_2, y)$  is the density under  $\mathbb{P}_\theta$  of the random vector  $\bar{S}^{(n,p)} = (\bar{S}^{(n,p)}, U^{1/n})$  defined in (2.31), and  $\bar{\mathcal{D}}_a := \bar{\mathcal{D}}_{p,a} \subset \mathbb{R}^3$  is the domain defined in (2.33). By the independence of  $\mathcal{U}$  and  $Y^{(n,p)}$ , for  $x \in \mathbb{R}^2$  and  $y \in (0, 1]$ ,  $\bar{h}_\theta^n(x_1, x_2, y)$  is the product of  $\bar{h}_\theta^n(x_1, x_2)$ , the density of  $\bar{S}^{(n,p)}$  under  $\mathbb{P}_\theta$  evaluated at  $(x_1, x_2)$ , and the density of  $U^{1/n}$  at  $y$ , which is equal to  $\frac{n}{y} e^{n \log y}$ . Hence, by Proposition 5.4, we have the following uniform estimate for  $\bar{h}_\theta^n$ : for  $\sigma$  a.e.  $\theta$ ,

$$\bar{h}_\theta^n(x_1, x_2, y) = \frac{n^2}{2\pi} \mathfrak{g}_\theta^n(x_1, x_2, y) e^{-nF(x_1, x_2, y)} (1 + o(1)), \quad (x_1, x_2) \in \mathbb{R}^2, \quad y \in (0, 1], \quad (6.2)$$

where

$$\mathfrak{g}_\theta^n(x_1, x_2, y) := \frac{1}{y} \bar{g}_\theta^n(x_1, x_2), \quad (6.3)$$

with  $\bar{g}_\theta^n$  defined in (5.11), and

$$F(x, y) := \Psi_p^*(x) - \log y, \quad x = (x_1, x_2) \in \mathbb{R}^2, y \in (0, 1]. \quad (6.4)$$

Thus, as in Section 5.5, the integral (6.1) of interest is once again a Laplace-type integral, and so one expects the significant contribution to come from the value of the integrand in a neighborhood of the point where the minimum of  $F$  over  $\bar{\mathcal{D}}_a$  is achieved. Now, for any  $x \in \mathbb{R}^2$ , the minimum of  $F(x, y)$  over  $y \in (0, 1]$  is clearly attained when  $y = 1$ , and by Lemma 2.2 the minimum of  $F(x, 1)$  over the region  $\{x \in \mathbb{R}^2 : x_2 > 0, x_1 x_2^{1/p} = a\}$  is attained at  $x = (a, 1)$ . Together with the strict convexity of the function  $\Psi_p^*(a, 1)$  established in Theorem 2.3 and the fact that its minimum is attained at 0, this shows that for  $a > 0$ , the minimizing point is given by

$$\arg \min_{(x_1, x_2, y) \in \bar{\mathcal{D}}_a} F(x_1, x_2, y) = \arg \min_{(x_1, x_2, y): 0 \leq y \leq 1, x_2 \geq 0, x_1 \geq ax_2^{1/p}} F(x_1, x_2, y) = (a, 1, 1). \quad (6.5)$$

However, in this case, the boundary of the domain  $\bar{\mathcal{D}}_a$  is not smooth at the minimizing point  $(a, 1, 1)$ , and so instead of Proposition 5.7, we apply Proposition 5.6 to prove Theorem 2.11.

**6.2 Proof of the sharp quenched estimate for  $\ell_p^n$ -balls**

We now prove Theorem 2.11.

*Proof of Theorem 2.11.* Fix  $p \in (1, \infty)$  and  $a > 0$  such that  $\mathbb{I}_p(a) < \infty$ . For  $\theta \in \mathbb{S}$ , recall that the density of  $\bar{S}^n$  can be expressed as in (6.1) and (6.2), and recall the assertion in (6.5) that the minimum of the function  $F$  in (6.4) on  $\bar{\mathcal{D}}_a$  is attained at  $(a, 1, 1)$ . Thus, for any open neighborhood  $\mathcal{U}$  of  $(a, 1, 1)$  whose closure does not intersect the plane  $y = 0$ , we split the probability into two parts. Fix  $\theta \in \mathbb{S}$ . Then

$$\mathbb{P}_\theta(\bar{S}^n \in \bar{\mathcal{D}}_a) = \mathbb{P}_\theta(\bar{S}^n \in \bar{\mathcal{D}}_a \cap \mathcal{U}) + \mathbb{P}_\theta(\bar{S}^n \in \bar{\mathcal{D}}_a \cap \mathcal{U}^c). \tag{6.6}$$

For the first term in (6.6), we have the following estimate from (6.1) and (6.2):

$$\mathbb{P}_\theta(\bar{S}^n \in \bar{\mathcal{D}}_a \cap \mathcal{U}) = \frac{n^2}{2\pi} \int_{\bar{\mathcal{D}}_a \cap \mathcal{U}} \mathfrak{g}_\theta^n(x_1, x_2, y) e^{-nF(x_1, x_2, y)} dx_1 dx_2 dy, \tag{6.7}$$

where  $\mathfrak{g}_\theta^n$  and  $F$  are given in (6.3) and (6.4).

The bulk of the proof is devoted to the asymptotics of the Laplace type integral in (6.7). In order to apply Proposition 5.6, we first perform a change of variables to transform the domain of integration. Let  $\mathfrak{T} : \bar{\mathcal{D}}_a \rightarrow \mathbb{R}^3$  be the mapping that takes  $(x_1, x_2, y)$  to  $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$  such that

$$\mathcal{X} = x_1 y - a x_2^{1/p}, \quad \mathcal{Y} = 1 - y, \quad \mathcal{Z} = x_2 - 1. \tag{6.8}$$

Note that the transformation  $\mathfrak{T}$  is invertible in a neighborhood of  $(a, 1, 1)$ , the Jacobian of this transformation at  $(a, 1, 1)$  is 1, the image of  $\bar{\mathcal{D}}_a$  under this transformation is

$$\tilde{\mathcal{D}}_a := \{(\mathcal{X}, \mathcal{Y}, \mathcal{Z}) \in \mathbb{R}^3 : 0 < \mathcal{Y} < 1, \mathcal{Z} > -1, \mathcal{X} > 0\},$$

and  $\mathfrak{T}$  maps the minimizer  $(a, 1, 1)$  of  $F$  to  $(0, 0, 0)$ . Hence, under the transformation  $\mathfrak{T}$ , setting  $\tilde{\mathcal{U}} := \mathfrak{T}(\mathcal{U})$ , we rewrite (6.7) as

$$\mathbb{P}_\theta(\bar{S}^n \in \bar{\mathcal{D}}_a \cap \mathcal{U}) = \frac{n^2}{2\pi} \int_{\tilde{\mathcal{D}}_a \cap \tilde{\mathcal{U}}} \mathfrak{g}_\theta^n \circ \mathfrak{T}^{-1}(\mathcal{X}, \mathcal{Y}, \mathcal{Z}) e^{-nF \circ \mathfrak{T}^{-1}(\mathcal{X}, \mathcal{Y}, \mathcal{Z})} d\mathcal{X} d\mathcal{Y} d\mathcal{Z}. \tag{6.9}$$

Let  $v_{ijk} := \partial_1^i \partial_2^j \partial_3^k F(a, 1, 1)$ . Then, from (6.8), we have

$$\begin{aligned} \frac{\partial F \circ \mathfrak{T}^{-1}}{\partial \mathcal{X}}(0, 0, 0) &= v_{100} \frac{\partial x_1}{\partial \mathcal{X}} + v_{010} \frac{\partial x_2}{\partial \mathcal{X}} + v_{001} \frac{\partial y}{\partial \mathcal{X}} \Big|_{(0,0,0)} \\ &= v_{100} \frac{1}{1 - \mathcal{Y}} \Big|_{(0,0,0)} \\ &= v_{100}; \\ \frac{\partial F \circ \mathfrak{T}^{-1}}{\partial \mathcal{Y}}(0, 0, 0) &= v_{100} \frac{\mathcal{X} + a(1 + \mathcal{Z})^{1/p}}{(1 - \mathcal{Y})^2} - v_{001} \Big|_{(0,0,0)} \\ &= av_{100} - v_{001}; \\ \frac{\partial F \circ \mathfrak{T}^{-1}}{\partial \mathcal{Z}}(0, 0, 0) &= v_{100} \frac{a(1 + \mathcal{Z})^{(1-p)/p}}{1 - \mathcal{Y}} + v_{010} \Big|_{(0,0,0)} \\ &= 0; \\ \frac{\partial^2 F \circ \mathfrak{T}^{-1}}{\partial \mathcal{Z}^2}(0, 0, 0) &= \left( v_{200} \frac{a(1 + \mathcal{Z})^{(1-p)/p}}{1 - \mathcal{Y}} + v_{110} \right) \frac{a(1 + \mathcal{Z})^{(1-p)/p}}{1 - \mathcal{Y}} \\ &\quad + v_{100} \left( -\frac{a(p-1)}{p^2} \right) \frac{(1 + \mathcal{Z})^{(1-2p)/p}}{1 - \mathcal{Y}} \end{aligned}$$

$$\begin{aligned}
 & + v_{110} \frac{a(1+\mathcal{Z})^{(1-p)/p}}{p(1-\mathcal{Y})} + v_{020} \Big|_{(0,0,0)} \\
 & = \frac{a^2}{p^2} v_{200} + \frac{2a}{p} v_{110} - \frac{a(p-1)}{p^2} v_{100} + v_{020}.
 \end{aligned}$$

Combining (6.4) with the duality relations (5.36)–(5.37), imply the following identities:

$$\begin{aligned}
 v_{100} & = \partial_{x_1} \Psi_p^*(a^*) = \lambda_{a,1}, \\
 v_{001} & = -1, \\
 v_{110} & = \partial_{x_1, x_2}^2 \Psi_p^*(a^*) = (\mathcal{H}_a)_{12}^{-1}, \\
 v_{020} & = \partial_{x_2, x_2}^2 \Psi_p^*(a^*) = (\mathcal{H}_a)_{22}^{-1}, \\
 v_{200} & = \partial_{x_1, x_1}^2 \Psi_p^*(a^*) = (\mathcal{H}_a)_{11}^{-1}.
 \end{aligned}$$

To apply Proposition 5.6, we first prove the following:

**Claim.** For  $\alpha \in (1/2, 1)$ , the integrand  $\mathfrak{g}_\theta^n \circ \mathfrak{T}^{-1}(\mathcal{X}, \mathcal{Y}, \mathcal{Z}) e^{-nF \circ \mathfrak{T}^{-1}(\mathcal{X}, \mathcal{Y}, \mathcal{Z})}$  in (6.9) admits a  $(F \circ \mathfrak{T}^{-1}, (0, 0, 0), \alpha, \mathfrak{g}_\theta^n \circ \mathfrak{T}^{-1})$ -representation on the bounded region  $\tilde{\mathcal{D}}_a \cap \tilde{\mathcal{U}}$ .

*Proof of Claim.* Let  $U$  be a neighborhood of  $a^* = (a, 1)$ . Then it follows from Lemma 5.11 that the function  $x \mapsto \bar{g}_\theta^n(x) e^{-n\Psi_p^*(x)}$  admits a  $(\Psi_p^*, (a, 1), \alpha, \bar{g}_\theta^n)$ -representation on  $\bar{\mathcal{D}}_a \cap U$ , and thus properties (1) and (2) of Definition 5.5 hold with  $f = \Psi_p^*$ ,  $g^n = \bar{g}_\theta^n$  and  $x^* = (a, 1)$ . It is easy to see that this implies that the corresponding properties also hold for the representation stated in the claim. Indeed, since  $F(x, y) = \Psi_p^*(x) - \ln y$  by (6.4),  $\bar{\mathcal{D}}_a = \{(x, y) : x \in \bar{\mathcal{D}}_a, 1 \geq y \geq 0\}$  by (2.29) and (2.33), and  $\tilde{\mathcal{U}}$  does not intersect the hyperplane  $\{y = 0\}$ , this implies that  $F$  is twice continuously differentiable on  $\bar{\mathcal{D}}_a \cap \mathcal{U}$  with unique minimizer on the closure of  $\bar{\mathcal{D}}_a \cap \mathcal{U}$  achieved at  $(a, 1, 1)$ . On the other hand, Lemma 5.11 and property (2) of Definition 5.5 imply that  $\bar{g}_\theta^n$  is continuously differentiable and there exists  $C < \infty$  such that  $\ln \bar{g}_\theta^n(x) \leq Cn^\alpha \|x\|_2$  for all  $x$  in a neighborhood  $U'$  of  $(a, 1)$ . By (6.3), it follows that on any neighborhood  $\mathcal{U}'$  of  $(a, 1, 1)$  of the form  $\{(x, y) : x \in U', 1 > y > \varepsilon\}$  for some  $\varepsilon > 0$ ,  $\mathfrak{g}_\theta^n$  is twice continuously differentiable and satisfies  $\ln \mathfrak{g}_\theta^n(x_1, x_2, y) \leq \frac{C}{\varepsilon} n^\alpha \|(x_1, x_2, y)\|_2$ . Combining these two properties with the fact that  $\mathfrak{T}$  is infinitely differentiable, maps  $(a, 1, 1)$  to  $(0, 0, 0)$  and  $\bar{\mathcal{D}}_a \cap \mathcal{U}$  to  $\tilde{\mathcal{D}}_a \cap \tilde{\mathcal{U}}$ , and has an infinitely differentiable inverse on  $\tilde{\mathcal{D}}_a \cap \tilde{\mathcal{U}}$ , it follows that properties (1) and (2) of Definition 5.5 hold with  $f = F \circ \mathfrak{T}^{-1}$ ,  $D = \tilde{\mathcal{D}}_a \cap \tilde{\mathcal{U}}$ ,  $g^n = \mathfrak{g}_\theta^n \circ \mathfrak{T}^{-1}$  and  $x^* = (0, 0, 0)$ . This proves the claim.  $\square$

Note that  $\tilde{\mathcal{D}}_a \subset \mathbb{R}_+ \times \mathbb{R}^2$ . Thus, we apply Proposition 5.6 with  $m = 1$  and  $d = 2$  to the transformed integral (6.9) with  $g^n$ ,  $f$ ,  $x^*$  and  $D$  therein replaced with  $\mathfrak{g}_\theta^n \circ \mathfrak{T}^{-1}$ ,  $F \circ \mathfrak{T}^{-1}$ ,  $(0, 0, 0)$  and  $\tilde{\mathcal{D}}_a$ , to obtain

$$\begin{aligned}
 & \mathbb{P}_\theta(\bar{\mathcal{S}}^n \in \tilde{\mathcal{D}}_a \cap U) \\
 & = \frac{n^2}{2\pi} \times \frac{\sqrt{2\pi}}{n^{5/2}} \frac{\mathfrak{g}_\theta^n \circ \mathfrak{T}^{-1}(0, 0, 0) e^{-nF \circ \mathfrak{T}^{-1}(0, 0, 0)}}{\partial_{\mathcal{X}} F \circ \mathfrak{T}^{-1}(0, 0, 0) \partial_{\mathcal{Y}} F \circ \mathfrak{T}^{-1}(0, 0, 0) \sqrt{|\partial_{\mathcal{Z}}^2 F \circ \mathfrak{T}^{-1}(0, 0, 0)|}} (1 + o(1)).
 \end{aligned}$$

The expression on the right-hand side can be simplified further using first the relations  $\mathfrak{T}^{-1}(0, 0, 0) = (a, 1, 1)$ ,  $F(a, 1, 1) = \Psi_p^*(a^*)$ ,  $\mathfrak{g}_\theta^n(a, 1, 1) = \bar{g}_\theta^n(a^*)$ , which follow from (6.3) and (6.4), together with the expressions for the partial derivatives of  $F$  calculated above, to obtain

$$\mathbb{P}_\theta(\bar{\mathcal{S}}^n \in \tilde{\mathcal{D}}_a \cap U) = \frac{1}{\sqrt{2\pi n \tilde{\gamma}_a}} \bar{g}_\theta^n(a^*) e^{-n\Psi_p^*(a^*)} (1 + o(1)),$$

where

$$\tilde{\gamma}_a := \lambda_{a,1} (a\lambda_{a,1} + 1) \sqrt{\left| -\frac{a(p-1)}{p^2} \lambda_{a,1} + \frac{2a}{p} (\mathcal{H}_a)_{12}^{-1} + (\mathcal{H}_a)_{22}^{-1} + \frac{a^2}{p^2} (\mathcal{H}_a)_{11}^{-1} \right|}.$$

Substituting for  $\bar{g}_\theta^n$  and  $\Psi_p^*$  using the relations (5.11), (2.19) and (2.11), we then obtain

$$\mathbb{P}_\theta(\bar{S}^n \in \bar{D}_a \cap U) = \frac{C_a^n(\theta^n)}{\gamma_a \sqrt{2\pi n}} e^{-n\mathbb{I}_p(a) + \sqrt{n}R_a^n(\theta^n)}(1 + o(1)), \tag{6.10}$$

where  $\gamma_a = (\det \mathcal{H}_a)\tilde{\gamma}_a$ , which coincides with the definition given in (2.24).

For the second term in (6.6), as in the proof of  $\ell_p^n$  spheres in (5.42), one can invoke the quenched large deviation principle for  $\bar{S}^n$  established in [16, Proposition 5.3] along with the fact that the rate function has a unique minimum, as proved in Lemma 2.2 to show that it is negligible with respect to (6.10). When combined with (2.32), (6.6) and (6.10), this yields (2.25).  $\square$

## 7 The joint density estimate

This section is devoted to the proof of the density estimate stated in Proposition 5.4. As usual, throughout fix  $p \in (1, \infty)$ . In Section 7.1 an identity for the joint density is established in terms of an integral. This integral is then shown in Section 7.2 to admit an alternative representation as an expectation with respect to a tilted measure. The latter representation is used in Section 7.3 to obtain certain asymptotic estimates. These results are finally combined in Section 7.4 to prove Proposition 5.4.

### 7.1 An integral representation for the joint density

**Lemma 7.1** (Representation for the density of  $\bar{S}^n$  under  $\mathbb{P}_\theta$ ). *Fix  $n \in \mathbb{N}$  and  $\theta \in \mathcal{S}$ , and recall the definitions of  $\Psi_p, \mathbb{J}_p, \lambda_x, \Phi_p$  and  $\Psi_{p,\theta}^n$  in (2.5), (2.7), (2.9), (5.6) and (5.7), respectively, and recall that  $\bar{h}_\theta^n$  is the density, under  $\mathbb{P}_\theta$ , of  $\bar{S}^n$  defined in (2.27). Then for all sufficiently large  $n$ , and  $x \in \mathbb{J}_p$ , the following identity holds,*

$$\bar{h}_\theta^n(x) = \left(\frac{n}{2\pi}\right)^2 e^{-n\Psi_p^*(x)} e^{n(\Psi_{p,\theta}^n(\lambda_x) - \Psi_p(\lambda_x))} \mathcal{I}_\theta^n(x), \tag{7.1}$$

where

$$\mathcal{I}_\theta^n(x) := \int_{\mathbb{R}^2} e^{-i\langle t, nx \rangle} \prod_{j=1}^n \frac{\Phi_p(\sqrt{n}\theta_j^n(\lambda_{x,1} + it_1), \lambda_{x,2} + it_2)}{\Phi_p(\sqrt{n}\theta_j^n \lambda_{x,1}, \lambda_{x,2})} dt. \tag{7.2}$$

Moreover, there exists  $s > 1$  such that  $(t_1, t_2) \mapsto (\prod_{j=1}^n \Phi_p(\sqrt{n}\theta_j^n(\lambda_{x,1} + it_1), \lambda_{x,2} + it_2))^{s/n}$  lies in  $\mathbb{L}_1(\mathbb{R}^2)$  for all sufficiently large  $n$ .

*Proof.* Let  $\mathbb{D}_p$  be as in (5.3), fix  $x \in \mathbb{J}_p$  and omit the subscript  $x$  from  $\lambda_x \in \mathbb{D}_p \subset \mathbb{R}^2$  and the superscript  $p$  from many quantities for notational simplicity. Recall the definition of  $\bar{V}_j^n$  in (5.5) and for  $\theta \in \mathcal{S}$ , let  $\bar{l}_\theta^n$  be the density of the sum  $\sum_{j=1}^n \bar{V}_j^n$  under  $\mathbb{P}_\theta$ . The moment generating function of this sum is given by

$$\begin{aligned} \int_{\mathbb{R}^2} e^{\langle \lambda, y \rangle} \bar{l}_\theta^n(y) dy &= \mathbb{E}_\theta \left[ e^{\langle \lambda, \sum_{j=1}^n \bar{V}_j^n \rangle} \right] \\ &= \prod_{j=1}^n \mathbb{E}_\theta \left[ e^{\lambda_1 \sqrt{n}\theta_j^n Y_j + \lambda_2 |Y_j|^p} \right] < \infty, \end{aligned}$$

where  $Y_1, \dots, Y_n$  are i.i.d. with density  $f_p$  defined in (2.3) and the finiteness follows because  $\lambda \in \mathbb{D}_p$  and thus  $\lambda_2 < 1/p$ . Then the Fourier transform of the integrable function  $y \mapsto e^{\langle \lambda, y \rangle} \bar{l}_\theta^n(y)$  is given as follows<sup>1</sup>: for  $t \in \mathbb{R}^2$ ,

$$\int_{\mathbb{R}^2} e^{\langle \lambda + it, y \rangle} \bar{l}_\theta^n(y) dy = \mathbb{E}_\theta \left[ e^{\langle \lambda + it, \sum_{j=1}^n \bar{V}_j^n \rangle} \right]$$

<sup>1</sup>Note that we use the convention for characteristic functions and thus put  $i$  in place of  $-2\pi i$  in the Fourier transform.

$$\begin{aligned}
 &= \prod_{j=1}^n \mathbb{E}_\theta \left[ e^{\langle \lambda + it, \bar{V}_j^n \rangle} \right] \\
 &= \prod_{j=1}^n \Phi_p(\sqrt{n}\theta_j^n(\lambda_1 + it_1), \lambda_2 + it_2). \tag{7.3}
 \end{aligned}$$

We now make the following claim:

**Claim.** There exists  $s > 1$  such that for any  $\lambda \in \mathbb{D}_p$ ,  $t \in \mathbb{R}^2$  and  $j, k \in \{1, \dots, n\}$ ,  $j \neq k$ , we have

$$K_\theta^{n,j,k}(\lambda, t) := \int_{\mathbb{R}^2} \left| \Phi_p(\sqrt{n}\theta_k^n(\lambda_1 + it_1), \lambda_2 + it_2) \Phi_p(\sqrt{n}\theta_j^n(\lambda_1 + it_1), \lambda_2 + it_2) \right|^s dt < \infty. \tag{7.4}$$

We defer the proof of the claim, first showing how the lemma follows from the claim. Let  $s > 1$  be as in the claim. Since the moment generating function  $\Phi_p$  is bounded, the claim holds for any  $s' > s$ . Now, pick any integer  $n > 2s$ . Then Hölder’s inequality and the claim imply that the right-hand side of (7.3) lies in  $\mathbb{L}_1(\mathbb{R}^2)$ . Hence, the second assertion of the lemma holds for any such  $n$ . We may then apply the inverse Fourier transform formula to conclude that, for all sufficiently large  $n$ ,

$$\bar{t}_\theta^n(x) = \left( \frac{1}{2\pi} \right)^2 \int_{\mathbb{R}^2} e^{-\langle \lambda + it, x \rangle} \prod_{j=1}^n \Phi_p(\sqrt{n}\theta_j^n(\lambda_1 + it_1), \lambda_2 + it_2) dt. \tag{7.5}$$

Next, recall that for any  $x \in \mathbb{J}_p$ ,  $\lambda = \lambda_x$  is chosen so that (2.9) is satisfied. Also, by (2.27) and (5.5), we have

$$\bar{S}^n = \frac{1}{n} \sum_{j=1}^n \bar{V}_j^n.$$

Hence, using (7.5), (2.9) and (5.7), we see that the density  $\bar{h}_\theta^n$  of  $\bar{S}^n$  under  $\mathbb{P}_\theta$  is given by

$$\begin{aligned}
 \bar{h}_\theta^n(x) &= n^2 \bar{t}_\theta^n(nx) \\
 &= \left( \frac{n}{2\pi} \right)^2 \int_{\mathbb{R}^2} e^{-\langle \lambda + it, nx \rangle} \prod_{j=1}^n \Phi_p(\sqrt{n}\theta_j^n(\lambda_1 + it_1), \lambda_2 + it_2) dt \\
 &= \left( \frac{n}{2\pi} \right)^2 e^{-n\Psi_p^*(x)} \int_{\mathbb{R}^2} e^{n(\Psi_p^*(x) - \langle \lambda, x \rangle)} e^{-i\langle t, nx \rangle} \prod_{j=1}^n \Phi_p(\sqrt{n}\theta_j^n(\lambda_1 + it_1), \lambda_2 + it_2) dt \\
 &= \left( \frac{n}{2\pi} \right)^2 e^{-n\Psi_p^*(x)} \int_{\mathbb{R}^2} e^{-n\Psi_p(\lambda)} e^{-i\langle t, nx \rangle} \prod_{j=1}^n \Phi_p(\sqrt{n}\theta_j^n(\lambda_1 + it_1), \lambda_2 + it_2) dt \\
 &= \left( \frac{n}{2\pi} \right)^2 e^{-n\Psi_p^*(x)} e^{n(\Psi_{p,\theta}^n(\lambda) - \Psi_p(\lambda))} \int_{\mathbb{R}^2} e^{-i\langle t, nx \rangle} \prod_{j=1}^n \frac{\Phi_p(\sqrt{n}\theta_j^n(\lambda_1 + it_1), \lambda_2 + it_2)}{\Phi_p(\sqrt{n}\theta_j^n(\lambda_1), \lambda_2)} dt,
 \end{aligned}$$

for  $x \in \mathbb{J}_p$ . Since the right-hand side above coincides with the expression for  $\bar{h}_\theta^n$  given in (7.1) and (7.2), this proves the first part of the lemma given the claim.

To complete the proof of the lemma, it only remains to prove the claim.

*Proof of the Claim.* Fix  $n \in \mathbb{N}$ ,  $j, k \in \{1, \dots, n\}$ ,  $j \neq k$ , and set  $\theta_1 := \theta_j^n$  and  $\theta_2 := \theta_k^n$ . Let  $\bar{v} := \bar{v}_\theta^{n,j,k}$  denote the density of  $\bar{V}_j^n + \bar{V}_k^n$  under  $\mathbb{P}_\theta$ . We assert that to prove the claim it suffices to show that the function  $\mathbb{R}^2 \ni z \mapsto e^{\langle \lambda, z \rangle} \bar{v}(z)$  lies in  $\mathbb{L}_{1+r}(\mathbb{R}^2)$  for some  $r \in (0, \infty)$ . Indeed, then by the Hausdorff-Young inequality [14, Theorem 8.21], the Fourier transform of  $z \mapsto e^{\langle \lambda, z \rangle} \bar{v}(z)$  lies in  $\mathbb{L}_s$ , where  $s$  is the “conjugate exponent” of  $1 + r$ . By (5.5) and (5.6), this is equivalent to saying that (7.4) holds with  $s = 1 + 1/r > 0$ .

To this end, we start by obtaining a convenient expression for  $\bar{v}$ . Note from (5.5) that  $\bar{V}_j^n + \bar{V}_k^n = T(Y_j, Y_k)$ , where  $(Y_j)_{j \in \mathbb{N}}$  are i.i.d. with common density  $f_p$  and  $T := T^{n,j,k} : \mathbb{R}^2 \rightarrow \mathbb{R} \times \mathbb{R}_+$  is the differentiable transformation given by,

$$T(y_1, y_2) = (\sqrt{n}(\theta_1 y_1 + \theta_2 y_2), |y_1|^p + |y_2|^p), \quad (y_1, y_2) \in \mathbb{R}^2.$$

Given  $(z_1, z_2) \in \mathbb{R} \times \mathbb{R}_+$ , we solve for  $(z_1, z_2) = T(y_1, y_2)$ . For  $z_2 > 0$ , consider the curves  $\{y \in \mathbb{R}^2 : z_1 = \sqrt{n}(\theta_1 y_1 + \theta_2 y_2)\}$  and  $\{y \in \mathbb{R}^2 : z_2 = |y_1|^p + |y_2|^p\}$ , which describe a line and an  $\ell_p^2$  sphere, respectively. Using Lagrange multipliers, it is straightforward to deduce that, the equation  $(z_1, z_2) = T(y_1, y_2)$ , which describes the intersection of these two curves, then has two solutions when  $|z_1| < z_2^{1/p} \sqrt{n}(|\theta_1|^{p/(p-1)} + |\theta_2|^{p/(p-1)})^{(p-1)/p} =: M(z_2)$ , one solution when  $|z_1| = M(z_2)$  and no  $y \in \mathbb{R}^2$  such that  $T(y) = (z_1, z_2)$  and when  $|z_1| > M(z_2)$ .

For  $|z_1| < M(z_2)$ , we define  $y^+$  and  $y^-$  to be the two solutions to  $T(y) = z$ . Thus,  $T$  is locally invertible on its range and hence, by the change of variables formula and the differentiability of  $T$ , we may write the density  $\bar{v}$  as

$$\begin{aligned} \bar{v}(z_1, z_2) &= \left( f_p(y_1^+) f_p(y_2^+) \left| \frac{\partial(y_1^+, y_2^+)}{\partial(z_1, z_2)} \right| + f_p(y_1^-) f_p(y_2^-) \left| \frac{\partial(y_1^-, y_2^-)}{\partial(z_1, z_2)} \right| \right) \mathbf{1}_{\{z_2 > 0, |z_1| < M(z_2)\}} \\ &= \left( \left| \frac{\partial(y_1^+, y_2^+)}{\partial(z_1, z_2)} \right| + \left| \frac{\partial(y_1^-, y_2^-)}{\partial(z_1, z_2)} \right| \right) e^{-z_2/p} \mathbf{1}_{\{z_2 > 0, |z_1| < M(z_2)\}}. \end{aligned}$$

Here,  $|\partial(y_1, y_2)/\partial(z_1, z_2)|$  is the Jacobian of the transformation  $T$  at  $(y_1, y_2)$ , which is given by the explicit formula

$$\mathcal{J}_T(y) := \left| \frac{\partial(y_1, y_2)}{\partial(z_1, z_2)} \right| = \frac{1}{\sqrt{np}} \left| \frac{1}{\theta_2 \operatorname{sgn}(y_1) |y_1|^{p-1} - \theta_1 \operatorname{sgn}(y_2) |y_2|^{p-1}} \right|, \quad (7.6)$$

where  $\operatorname{sgn}(\cdot)$  denotes the sign function.

For  $r > 0$ , the above discussion shows that

$$\begin{aligned} &\int_{\mathbb{R}^2} |e^{\lambda_1 z_1 + \lambda_2 z_2} \bar{v}(z_1, z_2)|^{1+r} dz_1 dz_2 \\ &= \int_{\mathbb{R}^2} |e^{\lambda_1 z_1 + \lambda_2 z_2 - z_2/p}|^{1+r} \left( \left| \frac{\partial(y_1^+, y_2^+)}{\partial(z_1, z_2)} \right| + \left| \frac{\partial(y_1^-, y_2^-)}{\partial(z_1, z_2)} \right| \right)^{1+r} \mathbf{1}_{\{z_2 > 0, |z_1| < M(z_2)\}} dz_1 dz_2 \\ &\leq 2^r \int_{\mathbb{R}^2} |e^{\lambda_1 z_1 + \lambda_2 z_2 - z_2/p}|^{1+r} \left( \left| \frac{\partial(y_1^+, y_2^+)}{\partial(z_1, z_2)} \right|^{1+r} + \left| \frac{\partial(y_1^-, y_2^-)}{\partial(z_1, z_2)} \right|^{1+r} \right) \mathbf{1}_{\{z_2 > 0, |z_1| < M(z_2)\}} dz_1 dz_2 \\ &= 2^r \int_{\mathbb{R}^2} |e^{\lambda_1 \sqrt{n}(\theta_1 y_1 + \theta_2 y_2) + (\lambda_2 - \frac{1}{p})(|y_1|^p + |y_2|^p)}|^{1+r} \left| \frac{\partial(y_1, y_2)}{\partial(z_1, z_2)} \right|^r dy_1 dy_2, \end{aligned}$$

where the inequality follows from  $(a + b)^{1+r} \leq 2^r (a^{1+r} + b^{1+r})$  for  $a, b \in \mathbb{R}_+$ , and the last equality uses the definition of  $T$ . Next, let  $\mathcal{N} \subset \mathbb{R}^2$  be a neighborhood of the origin. Then

$$\begin{aligned} &\int_{\mathbb{R}^2} |e^{\lambda_1 z_1 + \lambda_2 z_2} \bar{v}(z_1, z_2)|^{1+r} dz_1 dz_2 \\ &= 2^r \int_{\mathbb{R}^2 \cap \mathcal{N}} |e^{\lambda_1 \sqrt{n}(\theta_1 y_1 + \theta_2 y_2) + (\lambda_2 - \frac{1}{p})(|y_1|^p + |y_2|^p)}|^{1+r} \left| \frac{\partial(y_1, y_2)}{\partial(z_1, z_2)} \right|^r dy_1 dy_2 \\ &\quad + 2^r \int_{\mathbb{R}^2 \cap \mathcal{N}^c} |e^{\lambda_1 \sqrt{n}(\theta_1 y_1 + \theta_2 y_2) + (\lambda_2 - \frac{1}{p})(|y_1|^p + |y_2|^p)}|^{1+r} \left| \frac{\partial(y_1, y_2)}{\partial(z_1, z_2)} \right|^r dy_1 dy_2. \end{aligned}$$

Since  $p \in (1, \infty)$  and  $x \in \mathbb{J}_p$  implies  $p\lambda_2 = p\lambda_{x,2} < 1$ , it follows that  $e^{\lambda_1 \sqrt{n}(\theta_1 y_1 + \theta_2 y_2) + (\lambda_2 - \frac{1}{p})(|y_1|^p + |y_2|^p)}$  lies in  $\mathbb{L}_r(\mathbb{R}^2)$  for any  $r > 0$ . Moreover, since  $p > 1$ ,

by (7.6), there exists  $r_1 > 0$  small enough such that the Jacobian  $\mathcal{J}_T$  lies in  $\mathbb{L}_{r_1}(\mathcal{N})$ . On the other hand, there exists  $0 < r_2 < \infty$  large enough such that the Jacobian  $\mathcal{J}_T$  lies in  $\mathbb{L}_{r_2}(\mathcal{N}^c)$ . Thus, by Hölder's inequality, there exists  $r > 0$  such that the last display is finite. This completes the proof of the claim, and therefore of the lemma.  $\square$

$\square$

### 7.2 Representation of the integrand in terms of a tilted measure

We next obtain a representation for the integrand of the integral  $\mathcal{I}_\theta^n$  in (7.2) using a change of measure. Once again, from Section 2.4, recall the i.i.d. sequence of random variables  $(Y_j)_{j \in \mathbb{N}}$  defined on  $(\Omega, \mathcal{F}, \mathbb{P})$  that have density  $f_p$  and are independent of  $\Theta = (\Theta^n)_{n \in \mathbb{N}}$ . Fix  $a > 0$  such that  $\mathbb{L}_p(a) < \infty$ , recall the definition of  $\lambda = \lambda_a$  from (2.9). Fix  $n \in \mathbb{N}$ , and consider a “tilted” measure  $\tilde{\mathbb{P}}^n = \tilde{\mathbb{P}}^{n,a}$  on  $(\Omega, \mathcal{F})$  such that the (marginal) distribution of  $\Theta^n$  remains unchanged but conditioned on  $\Theta = \theta \in \mathbb{S}$ ,  $\{Y_j^n, j = 1, \dots, n\}$  are still independent, but not identically distributed, with  $Y_j^n$  having density  $\tilde{f}_j^n = \tilde{f}_{\theta,j}^{n,a}$  given by

$$\tilde{f}_j^n(y) := \exp(\langle \lambda_a, (\sqrt{n}\theta_j^n y, |y|^p) \rangle - \Lambda_p(\sqrt{n}\theta_j^n \lambda_1, \lambda_2)) f_p(y), \quad y \in \mathbb{R}, \quad (7.7)$$

with  $\Lambda_p$  as defined in (2.4) and as before we omit the explicit dependence and other quantities of  $\tilde{f}_j^n$  on  $p$  and  $a$ . For  $\theta \in \mathbb{S}$ , denote by  $\tilde{\mathbb{P}}_\theta^n$  and  $\tilde{\mathbb{E}}_\theta^n$  the probability and the expectation taken with respect to  $\tilde{\mathbb{P}}^n$ , conditioned on  $\theta$ , and likewise, let  $\widetilde{\text{Var}}_\theta^n(\cdot)$  and  $\widetilde{\text{Cov}}_\theta^n(\cdot, \cdot)$  denote the conditional variance and conditional covariance, respectively, under  $\tilde{\mathbb{P}}_\theta^n$ .

Recall from (2.4) and (5.6) that  $\Lambda_p(t) = \log \Phi_p(t)$  for  $t \in \mathbb{R}^2$ . Then, by (5.5), (5.6) and (7.7), it follows that for  $j = 1, \dots, n$  and  $\beta = (\beta_1, \beta_2) \in \mathbb{R}^2$ ,

$$\tilde{\mathbb{E}}_\theta^n \left[ e^{\langle \beta, \bar{V}_j^n \rangle} \right] = \frac{\Phi_p(\sqrt{n}\theta_j^n (\beta_1 + \lambda_1), \beta_2 + \lambda_2)}{\Phi_p(\sqrt{n}\theta_j^n \lambda_1, \lambda_2)}, \quad (7.8)$$

and hence,

$$\tilde{\mathbb{E}}_\theta^n [\bar{V}_j^n] = \nabla_\beta \tilde{\mathbb{E}}_\theta^n \left[ e^{\langle \beta, \bar{V}_j^n \rangle} \right] \Big|_{\beta=(0,0)} = \nabla \log \Phi_p(\sqrt{n}\theta_j^n \lambda_1, \lambda_2). \quad (7.9)$$

Denoting  $\bar{V}_j^n = (\bar{V}_{j,1}^n, \bar{V}_{j,2}^n)$ , by (7.8), we also have for  $k, l = 1, 2$ ,

$$\begin{aligned} \widetilde{\text{Cov}}_\theta^n (\bar{V}_{j,k}^n, \bar{V}_{j,l}^n) &= \tilde{\mathbb{E}}_\theta^n [\bar{V}_{j,k}^n \bar{V}_{j,l}^n] - \tilde{\mathbb{E}}_\theta^n [\bar{V}_{j,k}^n] \tilde{\mathbb{E}}_\theta^n [\bar{V}_{j,l}^n] \\ &= \partial_{\beta_k, \beta_l}^2 \tilde{\mathbb{E}}_\theta^n \left[ e^{\langle \beta, \bar{V}_j^n \rangle} \right] \Big|_{\beta=(0,0)} - \tilde{\mathbb{E}}_\theta^n [\bar{V}_{j,k}^n] \tilde{\mathbb{E}}_\theta^n [\bar{V}_{j,l}^n] \\ &= \partial_{k,l}^2 \log \Phi_p(\sqrt{n}\theta_j^n \lambda_1, \lambda_2). \end{aligned} \quad (7.10)$$

For  $x \in \mathbb{J}_p$ , define  $\widehat{V}_x^n$  to be

$$\widehat{V}_x^n := \frac{1}{\sqrt{n}} \sum_{j=1}^n (\bar{V}_j^n - x). \quad (7.11)$$

**Lemma 7.2.** For  $x \in \mathbb{J}_p$  and  $\theta \in \mathbb{S}$ , recall the definitions of  $\bar{V}_j^n$ ,  $\Phi_p$ ,  $c_x^n$ ,  $\mathcal{H}_x^n$  and  $\widehat{V}_x^n$  given in (5.5), (5.6), (5.9) and (7.11). Then

$$c_x^n(\theta^n) = \tilde{\mathbb{E}}_\theta^n [\widehat{V}_x^n], \quad (7.12)$$

$$\langle \mathcal{H}_x^n(\theta^n) t, t \rangle = \widetilde{\text{Var}}_\theta^n \left( \langle t, \widehat{V}_x^n \rangle \right), \quad \text{for all } t \in \mathbb{R}^2. \quad (7.13)$$

Moreover, for  $t = (t_1, t_2) \in \mathbb{R}^2$ ,

$$\hat{\mu}_{x,\theta}^n(t) := \tilde{\mathbb{E}}_\theta^n \left[ e^{i\langle t, \sqrt{n}\widehat{V}_x^n \rangle} \right] = e^{-i\langle t, nx \rangle} \prod_{j=1}^n \frac{\Phi_p(\sqrt{n}\theta_j^n(\lambda_{x,1} + it_1), \lambda_{x,2} + it_2)}{\Phi_p(\sqrt{n}\theta_j^n \lambda_{x,1}, \lambda_{x,2})}. \quad (7.14)$$

Furthermore, for  $\sigma$ -a.e.  $\theta$ , as  $n \rightarrow \infty$ ,  $\mathcal{H}_x^n(\theta^n)$  converges to the quantity  $\mathcal{H}_x$  defined in (2.12).

*Proof.* We fix  $\theta \in \mathbb{S}$  and  $x$  in the domain  $\mathbb{J}_p$  of  $\Psi_p^*$  defined in (2.7) and omit the subscript  $x$  from  $\lambda_x$  for notational simplicity. By (7.9), (7.11), the definition of  $\Psi_{p,\theta}^n$  in (5.7) and (2.8), we have,

$$\begin{aligned} \tilde{\mathbb{E}}_\theta^n \left[ \widehat{V}_x^n \right] &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \left( -x + \tilde{\mathbb{E}}_\theta^n[\bar{V}_j^n] \right) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \left( -x + \nabla \log \left( \Phi_p(\sqrt{n}\theta_j^n \lambda_1, \lambda_2) \right) \right) \\ &= \frac{1}{\sqrt{n}} \left( -nx + n \nabla \Psi_{p,\theta}^n(\lambda) \right) \\ &= \sqrt{n} \nabla \left( \Psi_{p,\theta}^n(\lambda) - \Psi_p(\lambda) \right). \end{aligned}$$

When combined with (5.9), this proves (7.12). Similarly, by the independence of  $\bar{V}_j^n, j = 1, \dots, n$ , under  $\tilde{\mathbb{P}}_\theta^n$ , (7.10), the definition of  $\Psi_{p,\theta}^n$  in (5.7) and the definition of  $\mathcal{H}_x^n$  in (5.9), it follows that

$$\widehat{\text{Var}}_\theta^n \left( \langle t, \widehat{V}_x^n \rangle \right) = \frac{1}{n} \sum_{j=1}^n \widehat{\text{Var}}_\theta^n \left( \langle t, \bar{V}_j^n \rangle \right) = \langle \mathcal{H}_x^n(\theta^n) t, t \rangle,$$

which proves (7.13). Also, by the definitions of  $\hat{\mu}_{x,\theta}^n$  and  $\widehat{V}_x^n$  in (7.14) and (7.11), respectively, the independence of  $\bar{V}_j^n, j = 1, \dots, n$ , under  $\tilde{\mathbb{P}}_\theta^n$  and the relation (7.8), it follows that for  $t \in \mathbb{R}^2$ ,

$$\hat{\mu}_{x,\theta}^n(t) = e^{-i\langle t, nx \rangle} \prod_{j=1}^n \tilde{\mathbb{E}}_\theta^n \left[ e^{i\langle t, \bar{V}_j^n \rangle} \right] = e^{-i\langle t, nx \rangle} \prod_{j=1}^n \frac{\Phi_p(\sqrt{n}\theta_j^n(\lambda_1 + it_1), \lambda_2 + it_2)}{\Phi_p(\sqrt{n}\theta_j^n \lambda_1, \lambda_2)},$$

which proves (7.14).

It only remains to establish the convergence stated in the last assertion of the lemma. By (5.9) and (5.7), it follows that for each  $i, j = 1, 2$ , there exists  $\alpha, \beta \in \mathbb{N}$  such that the entry  $(\mathcal{H}_x^n(\theta^n))_{ij}$  can be written as

$$\begin{aligned} (\mathcal{H}_x^n(\theta^n))_{ij} &= \frac{1}{n} \sum_{j=1}^n (\sqrt{n}\theta_j^n)^\alpha \partial_1^\alpha \partial_2^\beta \log \Phi_p(\sqrt{n}\theta_j^n \lambda_1, \lambda_2) \\ &= \int_{\mathbb{R}} u^\alpha \partial_1^\alpha \partial_2^\beta \log \Phi_p(u\lambda_1, \lambda_2) L_\theta^n(du). \end{aligned}$$

Since, the moment generating function  $\Phi_p$  is infinitely differentiable, the mapping  $u \mapsto \phi(u) := u^\alpha \partial_1^\alpha \partial_2^\beta \log \Phi_p(u\lambda_1, \lambda_2)$  is continuous. Moreover,  $\phi$  has polynomial growth by Lemma 5.2. Since Lemma 4.3 implies that  $\mathcal{W}_p(L_\theta^n, \gamma_2) \rightarrow 0$  as  $n \rightarrow \infty$ , it follows that

$$(\mathcal{H}_x^n(\theta^n))_{ij} \rightarrow \int_{\mathbb{R}} u^\alpha \partial_1^\alpha \partial_2^\beta \log \Phi_p(u\lambda_1, \lambda_2) \gamma_2(du) = (\mathcal{H}_x)_{ij},$$

where from Lemma 4.2(2) that, as  $n$  tends to infinity, the last equality follows by the definition of  $\mathcal{H}_x$  in (2.12).  $\square$

**7.3 Estimates of the integrand**

**Lemma 7.3.** Fix  $x \in \mathbb{J}_p$ . Recall the definitions of  $\widehat{V}_x^n$  and  $(\bar{V}_j^n)_{j=1,\dots,n}$  given in (7.11) and (5.5), respectively. There exist constants  $\tilde{C} < \infty$  such that for all  $n \in \mathbb{N}$  and for  $\sigma$ -a.e.  $\theta$ ,

$$\frac{1}{n} \sum_{j=1}^n \tilde{\mathbb{E}}_\theta^n \left[ \left\| \bar{V}_j^n - \tilde{\mathbb{E}}_\theta^n[\bar{V}_j^n] \right\|^3 \right] < \tilde{C}, \quad \frac{1}{n} \sum_{j=1}^n \tilde{\mathbb{E}}_\theta^n \left[ \left\| \bar{V}_j^n - \tilde{\mathbb{E}}_\theta^n[\bar{V}_j^n] \right\|^4 \right] < \tilde{C}, \tag{7.15}$$

and for all  $n \in \mathbb{N}$ ,

$$\tilde{\mathbb{E}}_\theta^n \left[ \left\| \widehat{V}_x^n - \tilde{\mathbb{E}}_\theta^n[\widehat{V}_x^n] \right\|^3 \right] < \tilde{C}. \tag{7.16}$$

*Proof.* Due to the following standard inequalities,  $(a^2 + b^2)^{3/2} \leq C'(|a|^3 + |b|^3)$  and  $\frac{1}{n} \sum_{j=1}^n |a_j|^3 \leq (\frac{1}{n} \sum_{j=1}^n |a_j|^4)^{3/4}$ , to show (7.15) it suffices to show the boundedness of

$$\frac{1}{n} \sum_{j=1}^n \tilde{\mathbb{E}}_\theta^n \left[ \left( \bar{V}_{j,1}^n - \tilde{\mathbb{E}}_\theta^n[\bar{V}_{j,1}^n] \right)^4 \right] \quad \text{and} \quad \frac{1}{n} \sum_{j=1}^n \tilde{\mathbb{E}}_\theta^n \left[ \left( \bar{V}_{j,2}^n - \tilde{\mathbb{E}}_\theta^n[\bar{V}_{j,2}^n] \right)^4 \right].$$

We show boundedness of just the first term; boundedness of the second can be shown analogously. Using following relation between cumulants and central moments, by simple calculation we have

$$\begin{aligned} & \frac{1}{n} \sum_{j=1}^n \tilde{\mathbb{E}}_\theta^n \left[ \left( \bar{V}_{j,1}^n - \tilde{\mathbb{E}}_\theta^n[\bar{V}_{j,1}^n] \right)^4 \right] \\ &= \frac{3}{n} \sum_{j=1}^n \tilde{\mathbb{E}}_\theta^n \left[ \left( \bar{V}_{j,1}^n - \tilde{\mathbb{E}}_\theta^n[\bar{V}_{j,1}^n] \right)^2 \right] + \int_{\mathbb{R}} \partial_1^4(\log \Phi_p(u\lambda_{x,1}, \lambda_{x,2})) L_\theta^n(du) \\ &= 3\widetilde{\text{Var}}_\theta^n(\widehat{V}_{x,1}^n) + \int_{\mathbb{R}} \partial_1^4(\log \Phi_p(u\lambda_{x,1}, \lambda_{x,2})) L_\theta^n(du). \end{aligned} \tag{7.17}$$

Now, by (7.13),  $\widetilde{\text{Var}}_\theta^n(\widehat{V}_{x,1}^n) = (\mathcal{H}_x^n(\theta^n))_{11}$  and so by the last assertion of Lemma 7.2, for  $\sigma$ -a.e.  $\theta$ , as  $n \rightarrow \infty$ ,  $\widetilde{\text{Var}}_\theta^n(\widehat{V}_{x,1}^n)$  converges to  $(\mathcal{H}_x)_{11}$ . Also, since the function  $\mathbb{R} \ni \partial_1^4(\log \Phi_p(u\lambda_{x,1}, \lambda_{x,2}))$  is continuous and has polynomial growth (the latter by Lemma 5.2), Lemma 4.3 and Lemma 4.2(2) together show that for  $\sigma$ -a.e.  $\theta$ , the second term on the right-hand-side of (7.17) also has a finite limit as  $n \rightarrow \infty$ . Therefore, for  $\sigma$ -a.e.  $\theta$ , the sum of the two terms is uniformly bounded.

Next, we deal with the second inequality. By (7.11), we have

$$\widehat{V}_x^n - \tilde{\mathbb{E}}_\theta^n[\widehat{V}_x^n] = \frac{1}{\sqrt{n}} \sum_{j=1}^n \left( \bar{V}_j^n - \tilde{\mathbb{E}}_\theta^n[\bar{V}_j^n] \right).$$

By Jensen’s inequality, we further obtain

$$\begin{aligned} & \tilde{\mathbb{E}}_\theta^n \left[ \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n \left( \bar{V}_j^n - \tilde{\mathbb{E}}_\theta^n[\bar{V}_j^n] \right) \right\|^3 \right] \\ & \leq \left( \tilde{\mathbb{E}}_\theta^n \left[ \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n \left( \bar{V}_j^n - \tilde{\mathbb{E}}_\theta^n[\bar{V}_j^n] \right) \right\|^4 \right] \right)^{3/4} \end{aligned}$$

$$\leq \left( \frac{2}{n^2} \tilde{\mathbb{E}}_\theta^n \left[ \left( \sum_{j=1}^n (\bar{V}_{j,1}^n - \tilde{\mathbb{E}}_\theta^n[\bar{V}_{j,1}^n]) \right)^4 \right] + \frac{2}{n^2} \tilde{\mathbb{E}}_\theta^n \left[ \left( \sum_{j=1}^n (\bar{V}_{j,2}^n - \tilde{\mathbb{E}}_\theta^n[\bar{V}_{j,2}^n]) \right)^4 \right] \right)^{3/4}.$$

Now, to show the boundedness of the last display, it suffices to show the boundedness of

$$\frac{1}{n^2} \tilde{\mathbb{E}}_\theta^n \left[ \left( \sum_{j=1}^n (\bar{V}_{j,m}^n - \tilde{\mathbb{E}}_\theta^n[\bar{V}_{j,m}^n]) \right)^4 \right] \quad \text{for } m = 1, 2.$$

We show the boundedness of the first term above, and the second follows similarly. For  $m \in \{1, 2\}$ , by the independence of  $(\bar{V}_{j,1}^n)_{j=1, \dots, n}$ , we have

$$\begin{aligned} & \frac{1}{n^2} \tilde{\mathbb{E}}_\theta^n \left[ \left( \sum_{j=1}^n (\bar{V}_{j,m}^n - \tilde{\mathbb{E}}_\theta^n[\bar{V}_{j,m}^n]) \right)^4 \right] \\ &= \frac{1}{n^2} \sum_{j=1}^n \tilde{\mathbb{E}}_\theta^n \left[ (\bar{V}_{j,m}^n - \tilde{\mathbb{E}}_\theta^n[\bar{V}_{j,m}^n])^4 \right] \\ & \quad + \frac{6}{n^2} \sum_{1 \leq i < j \leq n} \tilde{\mathbb{E}}_\theta^n \left[ (\bar{V}_{i,1}^n - \tilde{\mathbb{E}}_\theta^n[\bar{V}_{i,1}^n])^2 \right] \tilde{\mathbb{E}}_\theta^n \left[ (\bar{V}_{j,m}^n - \tilde{\mathbb{E}}_\theta^n[\bar{V}_{j,m}^n])^2 \right] \\ & \leq \frac{1}{n^2} \sum_{j=1}^n \tilde{\mathbb{E}}_\theta^n \left[ (\bar{V}_{j,m}^n - \tilde{\mathbb{E}}_\theta^n[\bar{V}_{j,m}^n])^4 \right] + 6 \left( \frac{2}{n} \sum_{j=1}^n \tilde{\mathbb{E}}_\theta^n \left[ (\bar{V}_{j,m}^n - \tilde{\mathbb{E}}_\theta^n[\bar{V}_{j,m}^n])^4 \right] \right)^2 \end{aligned}$$

which is bounded above by (7.15). This proves (7.16).  $\square$

**Lemma 7.4.** Fix  $x \in \mathbb{J}_p$  and recall the definitions of  $\mathcal{H}_x$ ,  $\Phi_p$ ,  $c_x^n$ ,  $\mathcal{H}_x^n$ ,  $\hat{V}_x^n$  and  $\hat{\mu}_{x,\theta}^n$  given in (2.12), (5.6), (5.9) (7.11) and (7.14), respectively. Then for  $\sigma$ -a.e.  $\theta$  and every neighborhood  $U \subset \mathbb{R}^2$  of the origin, there exist a neighborhood  $\tilde{U}$  of  $x$  and a constant  $C \in (0, 1)$  such that for all sufficiently large  $n$ ,

$$\sup_{t \in U^c} |\hat{\mu}_{y,\theta}^n(t)|^{1/n} < C, \quad y \in \tilde{U}. \tag{7.18}$$

Furthermore, for  $\sigma$ -a.e.  $\theta$ , there exist a neighborhood  $U \subset \mathbb{R}^2$  of the origin and a neighborhood  $\tilde{U}$  of  $x$  such that for all sufficiently large  $n$ ,

$$\left| \hat{\mu}_{y,\theta}^n \left( \frac{t}{\sqrt{n}} \right) e^{-itc_y^n(\theta^n)} \right| \leq \exp \left( -\frac{1}{2} \langle (\mathcal{H}_y - \varepsilon I)t, t \rangle \right), \quad y \in \tilde{U}, \quad t \in U. \tag{7.19}$$

*Proof.* We omit the subscript  $x$  of  $\lambda_x$  for notational simplicity. Now, for  $\theta \in \mathbb{S}$ , and  $t \in \mathbb{R}^2$ , the relation (7.8) yields the inequality

$$\left| \frac{\Phi_p(\sqrt{n}\theta_j^n(\lambda_1 + it_1), \lambda_2 + it_2)}{\Phi_p(\sqrt{n}\theta_j^n\lambda_1, \lambda_2)} \right| = \left| \tilde{\mathbb{E}}_\theta^n \left[ e^{i\langle t, \bar{V}_j^n \rangle} \right] \right| \leq \tilde{\mathbb{E}}_\theta^n \left[ \left| e^{i\langle t, \bar{V}_j^n \rangle} \right| \right] \leq 1. \tag{7.20}$$

Noting from (5.6) that  $\Phi_p(t)$  is the Fourier transform of the joint density of  $(Y_1, |Y_1|^p)$ , evaluated at  $+it$ , we can apply the Riemann-Lebesgue lemma [14, Theorem 8.22] to obtain

$$\|(\sqrt{n}\theta_j^n(i\lambda_1 - t_1), i\lambda_2 - t_2)\| \rightarrow \infty \quad \Rightarrow \quad |\Phi_p(\sqrt{n}\theta_j^n(\lambda_1 + it_1), \lambda_2 + it_2)| \rightarrow 0.$$

Now for  $\theta_j^n \neq 0$ ,  $\|t\| \rightarrow \infty$  implies  $\|(\sqrt{n}\theta_j^n(i\lambda_1 - t_1), i\lambda_2 - t_2)\| \rightarrow \infty$ . Thus, under the assumption that  $\theta_j^n \neq 0$ , we see that

$$\lim_{\|t\| \rightarrow \infty} \left| \frac{\Phi_p(\sqrt{n}\theta_j^n(\lambda_1 + it_1), \lambda_2 + it_2)}{\Phi_p(\sqrt{n}\theta_j^n \lambda_1, \lambda_2)} \right| = 0.$$

Since  $\Phi_p$  is a moment generating function which converges to 0 at infinity,  $\Phi_p$  is strictly smaller than 1 other than at the origin. For any neighborhood of the origin  $U \subset \mathbb{R}^2$  and any  $0 < K < \infty$ , there exists  $0 < r < 1$  such that for all  $t \in U^c$ , if  $K^{-1} \leq |\sqrt{n}\theta_j^n| \leq K$  and  $\theta_j^n \neq 0$ , then

$$\left| \frac{\Phi_p(\sqrt{n}\theta_j^n(\lambda_1 + it_1), \lambda_2 + it_2)}{\Phi_p(\sqrt{n}\theta_j^n \lambda_1, \lambda_2)} \right| < r.$$

This implies

$$\left| \frac{\Phi_p(\sqrt{n}\theta_j^n(\lambda_1 + it_1), \lambda_2 + it_2)}{\Phi_p(\sqrt{n}\theta_j^n \lambda_1, \lambda_2)} \right| < r^{1_{\{K^{-1} \leq |\sqrt{n}\theta_j^n| \leq K, \theta_j^n \neq 0\}}}.$$

Combining this with (7.14) yields the inequality

$$\sup_{t \in U^c} |\hat{\mu}_{x,\theta}^n(t)|^{1/n} \leq r^{\frac{1}{n} \sum_{j=1}^n 1_{\{K^{-1} \leq |\sqrt{n}\theta_j^n| \leq K, \theta_j^n \neq 0\}}}.$$

Since  $\frac{1}{n} \sum_{j=1}^n 1_{\{K^{-1} \leq |\sqrt{n}\theta_j^n| \leq K\}} = L_\theta^n([K^{-1}, K] \setminus \{0\})$  whose limit, as  $n \rightarrow \infty$ , is dominated by  $c_K := \gamma_2([K^{-1}, K]) > 0$  due to Lemma 4.3, we have for  $\sigma$ -a.e.  $\theta$ ,

$$\limsup_{n \rightarrow \infty} \sup_{t \in U^c} |\hat{\mu}_{x,\theta}^n(t)|^{1/n} \leq r^{c_K} < 1.$$

Thus, for  $\sigma$ -a.e.  $\theta$ , we have a uniform bound  $0 < C < 1$  such that for all sufficiently large  $n$ ,

$$\sup_{t \in U^c} |\hat{\mu}_{x,\theta}^n(t)|^{1/n} < C. \tag{7.21}$$

Since  $\Phi_p$  is uniformly continuous in  $\lambda_x$  by definition and  $\lambda_x$  is a infinitely differentiable function of  $x$  by the inverse function theorem applied to (2.9), we may choose a neighborhood  $\tilde{U}$  of  $x$  such that for  $y \in \tilde{U}$ ,

$$\sup_{t \in U^c} |\hat{\mu}_{y,\theta}^n(t)|^{1/n} < C,$$

i.e., for  $\sigma$ -a.e.  $\theta$  and all sufficiently large  $n$  (possibly depending on  $\theta$ ), (7.18) holds.

Next, note that by (7.14) and (7.12), for  $t \in \mathbb{R}^2$ ,

$$\hat{\mu}_{x,\theta}^n \left( \frac{t}{\sqrt{n}} \right) e^{-i\langle t, c_x^n(\theta^n) \rangle} = \tilde{\mathbb{E}}_\theta^n \left[ e^{i\langle t, \hat{V}_x^n - \tilde{\mathbb{E}}_\theta^n[\hat{V}_x^n] \rangle} \right].$$

Thus, for  $\theta \in \mathcal{S}$ , by (7.13) and [12, Lemma 3.3.7], we have the following expansion:

$$\begin{aligned} \left| \hat{\mu}_{x,\theta}^n \left( \frac{t}{\sqrt{n}} \right) e^{-i\langle t, c_x^n(\theta^n) \rangle} - 1 + \frac{1}{2} \langle \mathcal{H}_x^n(\theta^n) t, t \rangle \right| &\leq \tilde{\mathbb{E}}_\theta^n \left[ \left| \langle t, \hat{V}_x^n - \tilde{\mathbb{E}}_\theta^n[\hat{V}_x^n] \rangle \right|^3 \right] \\ &\leq \|t\|^3 \tilde{\mathbb{E}}_\theta^n \left[ \left\| \hat{V}_x^n - \tilde{\mathbb{E}}_\theta^n[\hat{V}_x^n] \right\|^3 \right]. \end{aligned}$$

For  $\varepsilon > 0$ , by (7.16) of Lemma 7.3, we may choose a neighborhood  $U \subset \mathbb{R}^2$  of the origin with small enough radius so that the right-hand-side of the last display is bounded by  $\varepsilon \|t\|^2$  for  $t \in U$ . On the other hand, by the convergence of  $\mathcal{H}_x^n(\theta^n)$  to  $\mathcal{H}_x$  established in

Lemma 7.2, for  $\sigma$ -a.e.  $\theta$ , there exists  $\varepsilon > 0$  such that  $\mathcal{H}_x^n(\theta^n) - \varepsilon I$  is positive definite for all sufficiently large  $n$  (possibly depending on  $\theta$ ) and for  $t \in U$ ,

$$\left| \hat{\mu}_{x,\theta}^n \left( \frac{t}{\sqrt{n}} \right) e^{-i\langle t, c_x^n(\theta^n) \rangle} \right| \leq 1 - \frac{1}{2} \langle (\mathcal{H}_x^n(\theta^n) - \varepsilon I)t, t \rangle \leq \exp \left( -\frac{1}{2} \langle (\mathcal{H}_x^n(\theta^n) - \varepsilon I)t, t \rangle \right).$$

Note that the right-hand side of the last display converges to the integrable function  $\exp(-\frac{1}{2} \langle (\mathcal{H}_x - \varepsilon I)t, t \rangle)$  as  $n$  tends to infinity. Similar to the proof of (7.18), the uniformity of the bound in (7.19) follows from the definition in (5.9), (5.7) and the aforementioned uniform continuity of  $\Phi_p$  in  $x$ .  $\square$

#### 7.4 Proof of the joint density estimate

We now combine the lemmas established in Sections 7.1–7.3 to prove the estimate for the density  $\bar{h}_\theta^n$  of  $\bar{S}^n$  obtained in Proposition 5.4.

*Proof of Proposition 5.4.* Fix  $n \in \mathbb{N}$ . Combining Lemma 7.1, (5.8) and (7.14) of Lemma 7.2, we see that for  $x \in \mathbb{J}_p$  and  $\sigma$ -a.e.  $\theta$ ,

$$\bar{h}_\theta^n(x) = \frac{n}{2\pi} e^{-n\Psi_p^*(x)} e^{\sqrt{n}R_x^n(\theta^n)} \frac{n}{2\pi} \int_{\mathbb{R}^2} \hat{\mu}_{x,\theta}^n(t) dt. \tag{7.22}$$

When compared with (5.10) and (5.11), to prove the proposition, it suffices to show that

$$\frac{n}{2\pi} \int_{\mathbb{R}^2} \hat{\mu}_{x,\theta}^n(t) dt = \det \mathcal{H}_x^{-1/2} \exp \left( \left\| \mathcal{H}_x^{-1/2} c_x^n(\theta^n) \right\|^2 \right) (1 + o(1)),$$

with the approximation uniformly for  $x$  in any compact set of  $\mathbb{J}_p$ .

Let  $U \subset \mathbb{R}^2$  be a neighborhood of the origin. We split the integral in the last display into two parts

$$\int_{\mathbb{R}^2} \hat{\mu}_{x,\theta}^n(t) dt = \int_U \hat{\mu}_{x,\theta}^n(t) dt + \int_{U^c} \hat{\mu}_{x,\theta}^n(t) dt. \tag{7.23}$$

Now, by the estimate (7.18) in Lemma 7.4, we have for  $C \in (0, 1)$  and  $s > 1$ ,

$$\left| \int_{U^c} \hat{\mu}_{x,\theta}^n(t) dt \right| \leq \int_{U^c} |\hat{\mu}_{x,\theta}^n(t)| dt \leq C^{n-s} \int_{U^c} |\hat{\mu}_{x,\theta}^n(t)|^{s/n} dt. \tag{7.24}$$

From the definition of  $\hat{\mu}_{x,\theta}^n$  in (7.14) and Lemma 7.1, we see that  $|\hat{\mu}_{x,\theta}^n(t)|^{s/n}$  is integrable. Hence, the right hand side of (7.24) tends to zero exponentially fast as  $n$  tends to infinity. Moreover, the convergence is uniform in a neighborhood of  $x$  by (7.18) from Lemma 7.4.

Recall the definition of  $\hat{\mu}_{x,\theta}^n$  in (7.14). By (7.11) and (7.12), the characteristic function of  $\hat{V}_x^n$  is given by  $\hat{\mu}_{x,\theta}^n \left( \frac{t}{\sqrt{n}} \right) e^{-itc_x^n(\theta^n)}$ . Since the sequence  $(\hat{V}_x^n)_{n \in \mathbb{N}}$  satisfies the Lyapunov-type condition stated in (7.15) of Lemma 7.3, the central limit theorem implies that it converges weakly to a centered Gaussian distribution with covariance matrix  $\mathcal{H}_x$ . Thus, the corresponding characteristic functions satisfy

$$\hat{\mu}_{x,\theta}^n \left( \frac{t}{\sqrt{n}} \right) e^{-itc_x^n(\theta^n)} \rightarrow \exp \left( -\frac{1}{2} \langle \mathcal{H}_x t, t \rangle \right). \tag{7.25}$$

Now, by (7.19) of Lemma 7.4 and (7.25), we may apply the dominated convergence theorem, and use (7.25) to obtain for  $\sigma$  a.e.  $\theta$ ,

$$\begin{aligned} \int_U \hat{\mu}_{x,\theta}^n(t) dt &= \frac{1}{n} \int_{\sqrt{n}U} \hat{\mu}_{x,\theta}^n \left( \frac{t}{\sqrt{n}} \right) dt \\ &= \frac{1}{n} \int_{\sqrt{n}U} \exp \left( itc_x^n(\theta^n) - \frac{1}{2} \langle \mathcal{H}_x t, t \rangle \right) dt \end{aligned}$$

## Sharp large deviations for random projections of $\ell_p^n$ balls

$$\begin{aligned} & + \frac{1}{n} \int_{\sqrt{n}U} e^{itc_x^n(\theta^n)} \left( \hat{\mu}_{x,\theta}^n \left( \frac{t}{\sqrt{n}} \right) e^{-itc_x^n(\theta^n)} - \exp \left( -\frac{1}{2} \langle \mathcal{H}_x(\theta^n)t, t \rangle \right) \right) dt \\ & = \frac{1}{n} \int_{\mathbb{R}^2} \exp \left( itc_x^n(\theta^n) - \frac{1}{2} \langle \mathcal{H}_x t, t \rangle \right) dt (1 + o(1)), \end{aligned}$$

with  $\mathcal{H}_x$  as in (2.12). Using standard properties of Gaussian integrals, this implies that

$$\int_U \hat{\mu}_{x,\theta}^n(t) dt = \frac{2\pi}{n} \det \mathcal{H}_x^{-1/2} \exp \left( \left\| \mathcal{H}_x^{-1/2} c_x^n(\theta^n) \right\|^2 \right) (1 + o(1)), \quad (7.26)$$

Combining (5.11), (7.22), (7.23), (7.26) and the estimate of the integral over  $U^c$  in (7.24), we conclude that the asymptotic expansion for the density  $\bar{h}_\theta^n(x)$  given in (5.10) holds uniformly for  $x$  in any compact subset of  $\mathbb{J}_p$ .

Finally, by the definition of  $\lambda_x$  in (2.9) and the inverse function theorem, the mapping  $x \mapsto \lambda_x$  is infinitely differentiable. Therefore, combining (5.11), (2.12), (5.9) and (5.8), we conclude  $\bar{g}_\theta^n$  is infinitely differentiable.  $\square$

### A Infimum of the rate function

In this section, we analyze the infimum of the rate function.

*Proof of Lemma 2.2.* Recall from (2.6) and (2.10), that we have the following expression for the rate function: for  $t \in \mathbb{R}$ ,

$$\begin{aligned} \mathbb{I}_p(t) & = \inf_{\tau_1 \in \mathbb{R}, \tau_2 > 0: \tau_1 \tau_2^{-1/p} = t} \Psi_p^*(\tau_1, \tau_2) \\ & = \inf_{\tau_1 \in \mathbb{R}, \tilde{\tau}_2 > 0: \tau_1 \tilde{\tau}_2^{-1} = t} \Psi_p^*(\tau_1, \tilde{\tau}_2^p) \\ & = \inf_{\tilde{\tau}_2 > 0} \Psi_p^*(\tilde{\tau}_2 t, \tilde{\tau}_2^p), \end{aligned} \quad (A.1)$$

where  $\Psi_p^*(\tilde{\tau}_2 t, \tilde{\tau}_2^p) = \sup_{s_1, s_2 \in \mathbb{R}} \{s_1 \tilde{\tau}_2 t + s_2 \tilde{\tau}_2^p - \Psi_p(s_1, s_2)\}$ .

By Lemmas 5.8 and 5.9 of [16],  $\Psi_p$  is essentially smooth, convex and lower semi-continuous; see Definition 2.3.5 of [10] for the definition of essential smoothness. Thus, by convexity, for  $t, \tau \in \mathbb{R}$ , when  $\Psi_p^*(\tau t, \tau^p) < \infty$ , there exist  $s_i = s_i(\tau t, \tau^p)$ ,  $i = 1, 2$ , that attain the supremum in the definition of  $\Psi_p^*(\tau t, \tau^p)$ , i.e.,

$$\Psi_p^*(\tau t, \tau^p) = s_1 \tau t + s_2 \tau^p - \Psi_p(s_1, s_2), \quad (A.2)$$

where, by (2.5),  $\Psi_p(s_1, s_2) = \int \Lambda_p(us_1, s_2) \gamma_2(du)$ , with  $\gamma_2$  being the standard Gaussian measure and  $\Lambda_p$  defined as in (2.4). Note that  $s_1, s_2$  satisfy the following first order conditions:

$$\tau t = \partial_1 \Psi_p(s_1, s_2) \quad \text{and} \quad \tau^p = \partial_2 \Psi_p(s_1, s_2),$$

where  $\partial_i$  represents the partial derivative with respect to  $s_i$ , for  $i = 1, 2$ . From [16, Lemma 5.9], we can exchange the order of differentiation and integration to obtain

$$\begin{aligned} \partial_1 \Psi_p(s_1, s_2) & = \int_{\mathbb{R}} u \partial_1 \Lambda_p(us_1, s_2) \gamma_2(du), \\ \partial_2 \Psi_p(s_1, s_2) & = \int_{\mathbb{R}} \partial_2 \Lambda_p(us_1, s_2) \gamma_2(du). \end{aligned} \quad (A.3)$$

To calculate these integrals, we first recall the expression for  $\Lambda_p$  established in [16, Lemma 5.7],

$$\Lambda_p(s_1, s_2) = -\frac{1}{p} \log(1 - ps_2) + \log M_{\gamma_p} \left( \frac{s_1}{(1 - ps_2)^{1/p}} \right), \quad (A.4)$$

where  $M_{\gamma_p}$  denotes the moment generating function of the measure  $\gamma_p$  with density defined in (2.3). Differentiation yields

$$\begin{aligned} \partial_1 \Lambda_p(us_1, s_2) &= \frac{M'_{\gamma_p} \left( \frac{us_1}{(1-ps_2)^{1/p}} \right)}{M_{\gamma_p} \left( \frac{us_1}{(1-ps_2)^{1/p}} \right)} \frac{1}{(1-ps_2)^{1/p}}, \\ \partial_2 \Lambda_p(us_1, s_2) &= \frac{1}{1-ps_2} + \frac{M'_{\gamma_p} \left( \frac{us_1}{(1-ps_2)^{1/p}} \right)}{M_{\gamma_p} \left( \frac{us_1}{(1-ps_2)^{1/p}} \right)} \frac{us_1}{(1-ps_2)^{(p+1)/p}}. \end{aligned} \tag{A.5}$$

Combining all the above relations, we obtain

$$\tau t = \int_{\mathbb{R}} \frac{M'_{\gamma_p} \left( \frac{us_1}{(1-ps_2)^{1/p}} \right)}{M_{\gamma_p} \left( \frac{us_1}{(1-ps_2)^{1/p}} \right)} \frac{u}{(1-ps_2)^{1/p}} \gamma_2(du), \tag{A.6}$$

$$\begin{aligned} \tau^p &= \int_{\mathbb{R}} \left( \frac{1}{1-ps_2} + \frac{M'_{\gamma_p} \left( \frac{us_1}{(1-ps_2)^{1/p}} \right)}{M_{\gamma_p} \left( \frac{us_1}{(1-ps_2)^{1/p}} \right)} \frac{us_1}{(1-ps_2)^{(p+1)/p}} \right) \gamma_2(du) \\ &= \frac{1}{1-ps_2} + \frac{\tau t s_1}{1-ps_2}, \end{aligned} \tag{A.7}$$

and note that (A.7) implies

$$\tau^p ps_2 + \tau t s_1 = \tau^p - 1. \tag{A.8}$$

Now, in view of (A.1), to compute  $\mathbb{I}_p(t)$  we have to first take the derivative of  $\Psi_p^*(\tau t, \tau^p)$  with respect to  $\tau$  and set it to 0. Note that in the following,  $s_1, s_2$  are functions of  $\tau$  and  $t$  satisfying (A.6) and (A.7). Using (2.5) and (A.1), we first rewrite  $\Psi_p(s_1, s_2)$  as

$$\begin{aligned} \Psi_p(s_1, s_2) &= \int_{\mathbb{R}} \Lambda_p(us_1, s_2) \gamma_2(du) \\ &= -\frac{1}{p} \log(1-ps_2) + \int_{\mathbb{R}} \log M_{\gamma_p} \left( \frac{us_1}{(1-ps_2)^{1/p}} \right) \gamma_2(du). \end{aligned}$$

From equations (A.2)-(A.8), we obtain

$$\begin{aligned} \frac{d}{d\tau} \Psi_p^*(\tau t, \tau^p) &= \frac{d}{d\tau} (s_1 \tau t + s_2 \tau^p - \Psi_p(s_1, s_2)) \\ &= \frac{\partial s_1}{\partial \tau} \tau t + s_1 t + \frac{\partial s_2}{\partial \tau} \tau^p + ps_2 \tau^{p-1} - \frac{\partial s_2}{\partial \tau} \frac{1}{1-ps_2} \\ &\quad - \int_{\mathbb{R}} \frac{M'_{\gamma_p} \left( \frac{us_1}{(1-ps_2)^{1/p}} \right)}{M_{\gamma_p} \left( \frac{us_1}{(1-ps_2)^{1/p}} \right)} \left[ \frac{\partial s_1}{\partial \tau} \frac{u}{(1-ps_2)^{1/p}} + \frac{\partial s_2}{\partial \tau} \frac{us_1}{(1-ps_2)^{1/p+1}} \right] \gamma_2(du) \\ &= \frac{\partial s_1}{\partial \tau} \tau t + s_1 t + \frac{\partial s_2}{\partial \tau} \tau^p + ps_2 \tau^{p-1} - \frac{\partial s_2}{\partial \tau} \frac{1}{1-ps_2} - \tau t \frac{\partial s_1}{\partial \tau} - \frac{s_1 \tau t}{1-ps_2} \frac{\partial s_2}{\partial \tau} \\ &= s_1 t + \frac{\partial s_2}{\partial \tau} \frac{\tau^p (1-ps_2) - s_1 \tau t - 1}{1-ps_2} + ps_2 \tau^{p-1} \\ &= s_1 t + ps_2 \tau^{p-1} \\ &= \tau^{p-1} - \frac{1}{\tau}. \end{aligned}$$

Setting the derivative computed above to 0, we conclude that the minimum over  $\tau > 0$  in (A.1) is attained at  $\tau = 1$ . Substituting this back into the definition of  $\mathbb{I}_p$ , we conclude that  $\mathbb{I}_p(t) = \Psi_p^*(t, 1)$  which, along with (2.6), proves Lemma 2.2.  $\square$

## B Proof of the Central Limit Theorem for the empirical measure

*Proof of Lemma 4.4.* Let  $(Z_j^n, j = 1, \dots, n)_{n \in \mathbb{N}}$  be independent standard Gaussian random variables. Then note that (e.g. see Section 2.4 or [35, Lemma 1])

$$\Theta_j^n \stackrel{(d)}{=} \frac{Z_j^n}{\|Z^n\|}, \tag{B.1}$$

where we use  $\|Z^n\| = \|Z^n\|_{n,2}$  to denote the Euclidean norm of the vector  $Z^n := (Z_1^n, \dots, Z_n^n)$ .

Since  $F$  is a thrice continuously differentiable function, we may apply Taylor's theorem, for  $x \in \mathbb{R}$  and  $h > 0$  to obtain

$$F(x+h) = F(x) + F'(x)h + \frac{F''(x)}{2}h^2 + \frac{F'''(\tilde{x})}{6}h^3,$$

for some  $\tilde{x} \in (x, x+h)$ . With the expansion above, we obtain

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{j=1}^n \left[ F\left(\sqrt{n} \frac{Z_j^n}{\|Z^n\|}\right) - \mathbb{E}[F(Z)] \right] \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \left[ F(Z_j^n) - \mathbb{E}[F(Z)] + F'(Z_j^n) \left(\frac{\sqrt{n}Z_j^n}{\|Z^n\|} - Z_j^n\right) + \frac{F''(Z_j^n)}{2} \left(\frac{\sqrt{n}Z_j^n}{\|Z^n\|} - Z_j^n\right)^2 \right. \\ & \quad \left. + \frac{F'''(\tilde{Z}_j^n)}{6} \left(\frac{\sqrt{n}Z_j^n}{\|Z^n\|} - Z_j^n\right)^3 \right], \\ &= \hat{r}_n(F) + \frac{1}{\sqrt{n}} \hat{s}_n(F) + \frac{1}{\sqrt{n}} \sum_{j=1}^n \frac{F'''(\tilde{Z}_j^n)}{6} \left(\frac{\sqrt{n}Z_j^n}{\|Z^n\|} - Z_j^n\right)^3 \end{aligned} \tag{B.2}$$

where  $\hat{r}_n(\cdot)$  and  $\hat{s}_n(\cdot)$  are defined in (4.4) and (4.3), respectively, and  $\tilde{Z}_j^n \in \mathbb{R}$  lies between  $Z_j^n$  and  $\sqrt{n}Z_j^n/\|Z^n\|$ .

In the following, the notation  $o(1)$  means having order  $o(1)$  in probability  $\mathbb{P}$ . We first show that the last term in (B.2) is of order  $o(1/n)$  in probability. By assumption,  $|F'''|$  has polynomial growth, so there exist  $q > 0$  and  $C < \infty$  such that

$$|F'''(t)| < C(1 + |t|^q), \quad \forall t \in \mathbb{R}.$$

Therefore, for each  $n \in \mathbb{N}$ ,

$$\sum_{j=1}^n \frac{|F'''(\tilde{Z}_j^n)|}{6} \left(\frac{\sqrt{n}Z_j^n}{\|Z^n\|} - Z_j^n\right)^3 \leq \frac{C}{6} \sum_{j=1}^n \left(1 + |\tilde{Z}_j^n|^q\right) \left|\frac{\sqrt{n}Z_j^n}{\|Z^n\|} - Z_j^n\right|^3.$$

Since  $\tilde{Z}_j^n$  lies between  $Z_j^n$  and  $\sqrt{n}Z_j^n/\|Z^n\|$ , and  $\sqrt{n}/\|Z^n\|$  converges to 1 almost surely. For each  $0 < \bar{C} < \infty$ , there exists  $N = N(\bar{C})$  such that a.s. for all  $n > N$ ,

$$|\tilde{Z}_j^n| < |Z_j^n| (1 + \bar{C}).$$

Combining the last two inequalities above, we obtain for some constant  $C' < \infty$ , and all  $n > N$ ,

$$\sum_{j=1}^n \frac{|F'''(\tilde{Z}_j^n)|}{6} \left(\frac{\sqrt{n}Z_j^n}{\|Z^n\|} - Z_j^n\right)^3 \leq C' \sum_{j=1}^n \left(1 + |Z_j^n|^q\right) \left|\frac{\sqrt{n}Z_j^n}{\|Z^n\|} - Z_j^n\right|^3$$

$$= C' \frac{\left| \|Z^n\| - \sqrt{n} \right|^3}{\sqrt{n}} \frac{n^{3/2}}{\|Z^n\|^3} \left[ \frac{1}{n} \sum_{j=1}^n |Z_j^n|^3 (1 + |Z_j^n|^q) \right].$$

From the Gaussian concentration inequality (see [38, Theorem 3.1.1]), there exists a universal constant  $c$  such that for  $\delta > 0$ ,

$$\mathbb{P} \left( \left| \|Z^n\| - \sqrt{n} \right| > \delta \right) \leq 2e^{-c\delta^2},$$

Given  $\epsilon > 0$ , we have

$$\begin{aligned} \mathbb{P} \left( \frac{1}{\sqrt{n}} \left| \|Z^n\| - \sqrt{n} \right|^3 > \epsilon \right) &= \mathbb{P} \left( \left| \|Z^n\| - \sqrt{n} \right| > n^{1/6} \epsilon^{1/3} \right) \\ &\leq 2e^{-c\epsilon^{2/3} n^{1/3}} \\ &\rightarrow 0, \quad \text{as } n \rightarrow \infty. \end{aligned} \tag{B.3}$$

On the other hand, since  $(Z_j^n)_{j=1, \dots, n}$  are independent, by the strong law of large numbers for triangular arrays, as  $n$  tends to infinity, almost surely

$$\frac{1}{n} \sum_{j=1}^n |Z_j^n|^3 (1 + |Z_j^n|^q) \rightarrow \mathbb{E} \left[ |Z|^3 (1 + |Z|^q) \right]. \tag{B.4}$$

Similarly, the strong law of large numbers also ensures that as  $n$  tends to infinity,

$$\frac{\|Z^n\|}{\sqrt{n}} \rightarrow 1, \quad \text{a.s.} \tag{B.5}$$

Together, (B.3), (B.4) and (B.5) show that

$$\sum_{j=1}^n \frac{|F'''(\tilde{Z}_j^n)|}{6} \left( \frac{\sqrt{n} Z_j^n}{\|Z^n\|} - Z_j^n \right)^3 = o(1).$$

We may then rewrite (B.2) as follows:

$$\frac{1}{\sqrt{n}} \left( \sum_{j=1}^n F \left( \sqrt{n} \frac{Z_j^n}{\|Z^n\|} \right) - \mathbb{E}[F(Z)] \right) = \hat{r}_n(F) + \frac{1}{\sqrt{n}} \hat{s}_n(F) + o \left( \frac{1}{\sqrt{n}} \right). \tag{B.6}$$

Due to the assumption that  $F'''$ ,  $G_1''$  and  $G_2''$  all have polynomial growth, the variances of  $F(Z)$ ,  $F'(Z)Z$ ,  $F''(Z)Z^2$ ,  $G_1(Z)$ ,  $G_1'(Z)Z$ ,  $G_2(Z)$  and  $G_2'(Z)Z$  are all finite. Define sequences  $(\mathfrak{A}_n)$ ,  $(\mathfrak{B}_n)$ ,  $(\mathfrak{C}_n)$ ,  $(\mathfrak{D}_n)$ ,  $(\mathfrak{E}_n)$ ,  $(\mathfrak{F}_n)$ ,  $(\mathfrak{G}_n)$  and  $(\mathfrak{H}_n)$  as follows:

$$\begin{aligned} \mathfrak{A}_n &:= \frac{1}{\sqrt{n}} \sum_{j=1}^n (F(Z_j^n) - \mathbb{E}[F(Z)]), & \mathfrak{B}_n &:= \frac{1}{\sqrt{n}} \sum_{j=1}^n (F'(Z_j^n)Z_j^n - \mathbb{E}[F'(Z)Z]), \\ \mathfrak{C}_n &:= \frac{1}{\sqrt{n}} \sum_{j=1}^n (F''(Z_j^n)(Z_j^n)^2 - \mathbb{E}[F''(Z)Z^2]), & \mathfrak{D}_n &:= \frac{1}{\sqrt{n}} \sum_{j=1}^n (|Z_j^n|^2 - 1), \\ \mathfrak{E}_n &:= \frac{1}{\sqrt{n}} \sum_{j=1}^n (G_1(Z_j^n) - \mathbb{E}[G_1(Z)]), & \mathfrak{F}_n &:= \frac{1}{\sqrt{n}} \sum_{j=1}^n (G_1'(Z_j^n)Z_j^n - \mathbb{E}[G_1'(Z)Z]), \\ \mathfrak{G}_n &:= \frac{1}{\sqrt{n}} \sum_{j=1}^n (G_2(Z_j^n) - \mathbb{E}[G_2(Z)]), & \mathfrak{H}_n &:= \frac{1}{\sqrt{n}} \sum_{j=1}^n (G_2'(Z_j^n)Z_j^n - \mathbb{E}[G_2'(Z)Z]). \end{aligned}$$

By the multivariate central limit theorem,  $(\mathfrak{A}_n, \mathfrak{B}_n, \mathfrak{C}_n, \mathfrak{D}_n, \mathfrak{E}_n, \mathfrak{F}_n, \mathfrak{G}_n, \mathfrak{H}_n)$  converges in distribution to a jointly Gaussian random vector  $M := (\mathfrak{A}, \mathfrak{B}, \mathfrak{C}, \mathfrak{D}, \mathfrak{E}, \mathfrak{F}, \mathfrak{G}, \mathfrak{H})$  in  $\mathbb{R}^8$  with mean 0 and covariance matrix

$$\left(\tilde{\Sigma}\right)_{ij} := \text{Cov}(M_i, M_j), \quad \text{for } i, j = 1, \dots, 6, \tag{B.7}$$

where

$$\begin{aligned} &(M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_8) \\ &:= (F(Z), F'(Z)Z, F''(Z)Z^2, Z^2, G_1(Z), G'_1(Z)Z, G_2(Z), G'_2(Z)Z). \end{aligned}$$

By the Skorokhod representation theorem, we can find  $(\tilde{\mathfrak{A}}_n, \tilde{\mathfrak{B}}_n, \tilde{\mathfrak{C}}_n, \tilde{\mathfrak{D}}_n, \tilde{\mathfrak{E}}_n, \tilde{\mathfrak{F}}_n, \tilde{\mathfrak{G}}_n, \tilde{\mathfrak{H}}_n)$  and  $\tilde{M} := (\tilde{\mathfrak{A}}, \tilde{\mathfrak{B}}, \tilde{\mathfrak{C}}, \tilde{\mathfrak{D}}, \tilde{\mathfrak{E}}, \tilde{\mathfrak{F}}, \tilde{\mathfrak{G}}, \tilde{\mathfrak{H}})$  all defined on some common probability space, such that

$$(\mathfrak{A}_n, \mathfrak{B}_n, \mathfrak{C}_n, \mathfrak{D}_n, \mathfrak{E}_n, \mathfrak{F}_n, \mathfrak{G}_n, \mathfrak{H}_n, M) \stackrel{(d)}{=} (\tilde{\mathfrak{A}}_n, \tilde{\mathfrak{B}}_n, \tilde{\mathfrak{C}}_n, \tilde{\mathfrak{D}}_n, \tilde{\mathfrak{E}}_n, \tilde{\mathfrak{F}}_n, \tilde{\mathfrak{G}}_n, \tilde{\mathfrak{H}}_n, \tilde{M}),$$

and

$$(\tilde{\mathfrak{A}}_n, \tilde{\mathfrak{B}}_n, \tilde{\mathfrak{C}}_n, \tilde{\mathfrak{D}}_n, \tilde{\mathfrak{E}}_n, \tilde{\mathfrak{F}}_n, \tilde{\mathfrak{G}}_n, \tilde{\mathfrak{H}}_n) \rightarrow \tilde{M} \text{ a.s.} \tag{B.8}$$

Now, we substitute  $(\tilde{\mathfrak{A}}_n, \tilde{\mathfrak{B}}_n, \tilde{\mathfrak{C}}_n, \tilde{\mathfrak{D}}_n, \tilde{\mathfrak{E}}_n, \tilde{\mathfrak{F}}_n, \tilde{\mathfrak{G}}_n, \tilde{\mathfrak{H}}_n)$  into (B.6), and we first take care of  $r_n$

$$\begin{aligned} \hat{r}_n(F) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \left[ F(Z_j^n) - \mathbb{E}[F(Z)] + F'(Z_j^n) \left( \frac{\sqrt{n}Z_j^n}{\|Z^n\|} - Z_j^n \right) \right] \\ &\stackrel{(d)}{=} \frac{1}{\sqrt{n}} \left( \sqrt{n}\tilde{\mathfrak{A}}_n + (\sqrt{n}\tilde{\mathfrak{B}}_n + n\mathbb{E}[F'(Z)Z]) \frac{\sqrt{n} - (\sqrt{n}\tilde{\mathfrak{D}}_n + n)^{1/2}}{(\sqrt{n}\tilde{\mathfrak{D}}_n + n)^{1/2}} \right) \\ &= \tilde{\mathfrak{A}}_n + \sqrt{n} \left( \mathbb{E}[F'(Z)Z] + \frac{\tilde{\mathfrak{B}}_n}{\sqrt{n}} \right) \left( \frac{1 - (1 + \tilde{\mathfrak{D}}_n/\sqrt{n})^{1/2}}{(1 + \tilde{\mathfrak{D}}_n/\sqrt{n})^{1/2}} \right) \\ &= \tilde{\mathfrak{A}}_n + \sqrt{n}H_1 \left( \frac{\tilde{\mathfrak{B}}_n}{\sqrt{n}}, \frac{\tilde{\mathfrak{D}}_n}{\sqrt{n}} \right), \end{aligned}$$

where  $H_1 : \mathbb{R}^2 \rightarrow \mathbb{R}$  is the mapping

$$H_1(x, y) := (\mathbb{E}[F'(Z)Z] + x) \frac{1 - (1 + y)^{1/2}}{(1 + y)^{1/2}}.$$

Since  $\tilde{\mathfrak{B}}_n/\sqrt{n}$  and  $\tilde{\mathfrak{D}}_n/\sqrt{n}$  converge to 0 almost surely by (B.8), we consider the Taylor expansion of  $H_1$  at  $(0, 0)$ :

$$\begin{aligned} H_1(x, y) &= \frac{1 - (1 + y)^{1/2}}{(1 + y)^{1/2}} \Big|_{(x,y)=(0,0)} x \\ &\quad + (\mathbb{E}[F'(Z)Z] + x) \frac{-1}{2(1 + y)^{3/2}} \Big|_{(x,y)=(0,0)} y \\ &\quad + O(x^2 + y^2) \\ &= -\frac{y}{2}\mathbb{E}[F'(Z)Z] + O(x^2 + y^2). \end{aligned}$$

Combining the last three displays, we obtain

$$\hat{r}_n(F) \stackrel{(d)}{=} \tilde{\mathfrak{A}}_n + \sqrt{n} \left( -\frac{\tilde{\mathfrak{D}}_n}{2\sqrt{n}}\mathbb{E}[F'(Z)Z] + O\left(\frac{\tilde{\mathfrak{B}}_n^2}{n} + \frac{\tilde{\mathfrak{D}}_n^2}{n}\right) \right)$$

$$= \tilde{\mathfrak{A}}_n - \frac{1}{2} \mathbb{E}[F'(Z)Z] \tilde{\mathfrak{D}}_n + \mathbb{E}[F'(Z)Z] O\left(\frac{\tilde{\mathfrak{B}}_n^2}{\sqrt{n}} + \frac{\tilde{\mathfrak{D}}_n^2}{\sqrt{n}}\right).$$

By the a.s. convergence,  $(\tilde{\mathfrak{A}}_n, \tilde{\mathfrak{D}}_n) \rightarrow (\tilde{\mathfrak{A}}, \tilde{\mathfrak{D}})$ , we see that as  $n$  tends to infinity,

$$\frac{\tilde{\mathfrak{B}}_n^2}{\sqrt{n}} + \frac{\tilde{\mathfrak{D}}_n^2}{\sqrt{n}} \rightarrow 0, \quad \text{a.s.}$$

Applying Slutsky's lemma and the almost sure convergence above, we obtain

$$\hat{r}_n(F) \Rightarrow \tilde{\mathfrak{A}} - \frac{1}{2} \mathbb{E}[F'(Z)Z] \tilde{\mathfrak{D}}, \tag{B.9}$$

as  $n \rightarrow \infty$ .

Similarly, for  $s_n$  we have

$$\begin{aligned} \hat{s}_n(F) &= \frac{1}{2} \sum_{j=1}^n F''(Z_j^n) (Z_j^n)^2 \left(\frac{\sqrt{n}}{\|Z^n\|} - 1\right)^2 \\ &= \frac{1}{2} n \left(\mathbb{E}[F''(Z)Z^2] + \frac{\mathfrak{C}_n}{\sqrt{n}}\right) \left(\frac{1}{(1 + \mathfrak{D}_n/\sqrt{n})^{1/2}} - 1\right)^2 \\ &\stackrel{(d)}{=} \frac{1}{2} n H_2 \left(\frac{\tilde{\mathfrak{C}}_n}{\sqrt{n}}, \frac{\tilde{\mathfrak{D}}_n}{\sqrt{n}}\right), \end{aligned}$$

where  $H_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$  is the mapping

$$H_2(x, y) := (\mathbb{E}[F''(Z)Z^2] + x) \left(\frac{1}{(1 + y)^{1/2}} - 1\right)^2, \quad (x, y) \in \mathbb{R}^2.$$

Note that  $\tilde{\mathfrak{C}}_n/\sqrt{n}$  and  $\tilde{\mathfrak{D}}_n/\sqrt{n}$  converge to 0 almost surely by (B.8). We now apply the Taylor expansion to  $H_2$  at  $(0, 0)$  and obtain

$$H_2(x, y) = \frac{1}{4} \mathbb{E}[F''(Z)Z^2] y^2 + O(x^3 + y^3).$$

With the above expansion for  $H_2$ , we write

$$\begin{aligned} \hat{s}_n(F) &\stackrel{(d)}{=} \frac{1}{8} \mathbb{E}[F''(Z)Z^2] \tilde{\mathfrak{D}}_n^2 + O\left(\frac{\tilde{\mathfrak{C}}_n^3}{\sqrt{n}} + \frac{\tilde{\mathfrak{D}}_n^3}{\sqrt{n}}\right) \\ &\Rightarrow \frac{1}{8} \mathbb{E}[F''(Z)Z^2] \tilde{\mathfrak{D}}^2, \end{aligned} \tag{B.10}$$

as  $n$  tends to infinity, which holds since  $\tilde{\mathfrak{D}}_n \rightarrow \tilde{\mathfrak{D}}$  almost surely. This completes the analysis of the expansion for  $F$ . Fix  $i = 1, 2$ , we next consider the expansion for  $G_i$ . Following the same method, we can write

$$\begin{aligned} &\sqrt{n} \left[ \frac{1}{n} \sum_{j=1}^n G_i \left( \sqrt{n} \frac{Z_j^n}{\|Z^n\|} \right) - \mathbb{E}[G_i(Z)] \right] \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \left[ G_i(Z_j^n) - \mathbb{E}[G_i(Z)] + G_i'(Z_j^n) \left( \frac{\sqrt{n} Z_j^n}{\|Z^n\|} - Z_j^n \right) + \frac{1}{2} G_i''(\tilde{Z}_i^n) \left( \frac{\sqrt{n} Z_j^n}{\|Z^n\|} - Z_j^n \right)^2 \right]. \end{aligned}$$

Again by assumption,  $G_i''$  has polynomial growth, and thus the last term is of order  $o(1)$ . Hence, we may rewrite the terms above as follows:

$$\sqrt{n} \left[ \frac{1}{n} \sum_{j=1}^n G_i \left( \sqrt{n} \frac{Z_j^n}{\|Z^n\|} \right) - \mathbb{E}[G_i(Z)] \right]$$

$$\begin{aligned}
 &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \left[ G_i(Z_j^n) - \mathbb{E}[G_i(Z)] + G_i'(Z_j^n) \left( \frac{\sqrt{n}Z_j^n}{\|Z^n\|} - Z_j^n \right) \right] + o(1) \\
 &= \hat{r}_n(G_i) + o(1).
 \end{aligned} \tag{B.11}$$

Thus, the expansion in Lemma 4.4 follows from (B.1), (B.9), (B.10) and (B.11). The second assertion of the lemma is a consequence of (B.9), (B.10), the analog of (B.9) with  $F$  replaced with  $G_i$  and the joint convergence of  $(\tilde{\mathfrak{A}}_n, \tilde{\mathfrak{D}}_n, \tilde{\mathfrak{E}}_n) \Rightarrow (\tilde{\mathfrak{A}}, \tilde{\mathfrak{D}}, \tilde{\mathfrak{E}})$  and  $(\tilde{\mathfrak{A}}_n, \tilde{\mathfrak{D}}_n, \tilde{\mathfrak{E}}_n) \Rightarrow (\tilde{\mathfrak{A}}, \tilde{\mathfrak{D}}, \tilde{\mathfrak{E}})$ .  $\square$

### C Proof of Lemma 5.2

*Proof.* For  $p = 2$ ,  $\gamma_2$  is the standard Gaussian,  $\log M_{\gamma_2}(t) = t^2/2$  and so the lemma follows.

Next, we consider the case  $p > 2$ . Let  $Y$  be a generalized  $p$ -Gaussian random variable with density as in (2.3). The moments of  $Y$  are given in [30] by

$$\mathbb{E}[Y^m] = \begin{cases} 0, & m \text{ odd,} \\ \frac{p^{m/p} \Gamma(\frac{m+1}{p})}{\Gamma(\frac{1}{p})}, & m \text{ even.} \end{cases} \tag{C.1}$$

Note that  $d \log M_{\gamma_p}(t)/dt = \mathbb{E}[Y e^{tY}]/\mathbb{E}[e^{tY}]$ , and for each  $k > 1$ ,  $d^k \log M_{\gamma_p}(t)/dt^k$  is a linear combination of products of functions the form

$$t \mapsto \frac{\mathbb{E}[Y^n e^{tY}]}{\mathbb{E}[e^{tY}]}, \quad \text{for } n = 1, \dots, k.$$

Therefore, we only need to show that these functions have at most polynomial growth. The case when  $k = 0$  then follows since the derivative of  $\log M_{\gamma_p}$  has polynomial growth, thus,  $\log M_{\gamma_p}$  also has polynomial growth.

We first consider the case when  $n$  is odd and the case when  $n$  is even can be deduced analogously. Note that for  $t \in \mathbb{R}$  and  $n$  odd,

$$\begin{aligned}
 \frac{\mathbb{E}[Y^n e^{tY}]}{\mathbb{E}[e^{tY}]} &= \frac{\sum_{m=0}^{\infty} t^{2m+1} \frac{(p^{1/p})^{2m+1+n} \Gamma(\frac{2m+2+n}{p})}{\Gamma(2m+1+n)}}{\sum_{m=0}^{\infty} t^{2m} \frac{(p^{1/p})^{2m} \Gamma(\frac{2m+1}{p})}{\Gamma(2m+1)}} \\
 &\leq \frac{\sum_{m=0}^{\infty} t^{2m+1} \frac{(p^{1/p})^{2m+1+n} \Gamma(\frac{2m+2+n}{p})}{\Gamma(2m+1+n)}}{\sum_{m=n'}^{\infty} t^{2m} \frac{(p^{1/p})^{2m} \Gamma(\frac{2m+1}{p})}{\Gamma(2m+1)}} \\
 &= \frac{\sum_{m=0}^{\infty} t^{2m+1} \frac{(p^{1/p})^{2m+1+n} \Gamma(\frac{2m+2+n}{p})}{\Gamma(2m+1+n)}}{\sum_{m=0}^{\infty} t^{2m+2n'} \frac{(p^{1/p})^{2m+2n'} \Gamma(\frac{2m+2n'+1}{p})}{\Gamma(2m+2n'+1)}}.
 \end{aligned}$$

Pick  $n' = (n - 1)/2$  to obtain

$$\frac{\mathbb{E}[Y^n e^{tY}]}{\mathbb{E}[e^{tY}]} \leq \frac{t}{t^{n-1}} \times \frac{\sum_{m=0}^{\infty} t^{2m} \frac{(p^{1/p})^{2m+1+n} \Gamma(\frac{2m+2+n}{p})}{\Gamma(2m+1+n)}}{\sum_{m=0}^{\infty} t^{2m} \frac{(p^{1/p})^{2m+n-1} \Gamma(\frac{2m+n}{p})}{\Gamma(2m+n)}}.$$

Now, note that for each  $m \in \mathbb{N} \cup \{0\}$ ,

$$\frac{\frac{(p^{1/p})^{2m+1+n} \Gamma(\frac{2m+2+n}{p})}{\Gamma(2m+2)}}{\frac{(p^{1/p})^{2m+n-1} \Gamma(\frac{2m+n}{p})}{\Gamma(2m+n)}} = \frac{p^{2/p}}{2m+1+n} \frac{\Gamma(\frac{2m+2+n}{p})}{\Gamma(\frac{2m+1+n}{p})} \leq (2m+1+n)^{\frac{2}{p}-1} \leq 1,$$

where the second to last inequality is due to Wendel [41, Equation 7]. Thus, we have shown that  $\mathbb{E}[Y^n e^{tY}]/\mathbb{E}[e^{tY}] \leq t^{2-n}$ , which has at most linear growth for any  $n \in \mathbb{N}$ .

Lastly, we turn to the case when  $1 < p < 2$ . We simply demonstrate the case  $k = 1$ , the general result can be deduced using similarly. Again, we start with  $k = 1$ , and for general  $k \in \mathbb{N} \cup \{0\}$ , the result can be deduced using the same technique as in the case  $p > 2$ .

In view of (C.1) we have for  $t \in \mathbb{R}$ ,

$$\begin{aligned} \frac{d}{dt} \log M_{\gamma_p}(t) &= \frac{\mathbb{E}[Y e^{tY}]}{\mathbb{E}[e^{tY}]} \\ &= \frac{\sum_{m=0}^{\infty} t^{2m+1} \frac{\mathbb{E}[Y^{2m+2}]}{(2m+1)!}}{\sum_{m=0}^{\infty} t^{2m} \frac{\mathbb{E}[Y^{2m}]}{(2m)!}} \\ &= \frac{\sum_{m=0}^{n-1} t^{2m+1} \frac{(p^{1/p})^{2m+2} \Gamma(\frac{2m+3}{p})}{\Gamma(2m+2)} + \sum_{m=n}^{\infty} t^{2m+1} \frac{(p^{1/p})^{2m+2} \Gamma(\frac{2m+3}{p})}{\Gamma(2m+2)}}{\sum_{m=0}^{\infty} t^{2m} \frac{(p^{1/p})^{2m} \Gamma(\frac{2m+1}{p})}{\Gamma(2m+1)}} \\ &\leq \sum_{m=1}^{n-1} t^{2m+1} \frac{(p^{1/p})^{2m+2} \Gamma(\frac{2m+3}{p})}{\Gamma(2m+2)} + \frac{\sum_{m=n}^{\infty} t^{2m+1} \frac{(p^{1/p})^{2m+2} \Gamma(\frac{2m+3}{p})}{\Gamma(2m+2)}}{\sum_{m=0}^{\infty} t^{2m} \frac{(p^{1/p})^{2m} \Gamma(\frac{2m+1}{p})}{\Gamma(2m+1)}} \\ &= \sum_{m=1}^{n-1} t^{2m+1} \frac{(p^{1/p})^{2m+2} \Gamma(\frac{2m+3}{p})}{\Gamma(2m+2)} + t^{2n+1} \frac{\sum_{m=0}^{\infty} t^{2m} \frac{(p^{1/p})^{2m+2n+2} \Gamma(\frac{2m+2n+3}{p})}{\Gamma(2m+2n+2)}}{\sum_{m=0}^{\infty} t^{2m} \frac{(p^{1/p})^{2m} \Gamma(\frac{2m+1}{p})}{\Gamma(2m+1)}}, \end{aligned}$$

where the inequality follows from  $\mathbb{E}[e^{tY}] \geq 1$ , which is due to Jensen's inequality.

To conclude the proof of the lemma, it suffices to show that there exists  $n \in \mathbb{N}$  such that for all  $m \in \mathbb{N} \cup \{0\}$

$$\frac{\frac{(p^{1/p})^{2m+2n+2} \Gamma(\frac{2m+2n+3}{p})}{\Gamma(2m+2n+2)}}{\frac{(p^{1/p})^{2m} \Gamma(\frac{2m+1}{p})}{\Gamma(2m+1)}} = p^{(2n+2)/p} \frac{\Gamma((2m+2n+3)/p) \Gamma(2m+1)}{\Gamma((2m+1)/p) \Gamma(2m+2n+2)} \leq 1.$$

To this end, pick  $a, b \in \mathbb{N}$  such that the following inequalities hold:

$$\begin{aligned} (2m+1) \left(1 - \frac{1}{p}\right) - 1 &< a < (2m+1) \left(1 - \frac{1}{p}\right); \\ (2m+2n) \left(1 - \frac{1}{p}\right) - \frac{3}{p} &< b < (2m+2n) \left(1 - \frac{1}{p}\right) - \frac{3}{p} + 1. \end{aligned} \tag{C.2}$$

Then we have the inequality

$$2n - \frac{2n}{p} - 1 - \frac{2}{p} < b - a < 2n - \frac{2n}{p} + 1 - \frac{2}{p}. \tag{C.3}$$

Now we use the identity  $\Gamma(z+1) = z\Gamma(z)$  and the chosen  $a, b$  above to obtain

$$\begin{aligned} &p^{\frac{2n+2}{p}} \frac{\Gamma((2m+2n+3)/p) \Gamma(2m+1)}{\Gamma((2m+1)/p) \Gamma(2m+2n+2)} \\ &= p^{\frac{2n+2}{p}} \frac{\Gamma(2m+1) \left(\frac{2m+1}{p}\right) \cdots \left(\frac{2m+1}{p} + a\right)}{\Gamma\left(\frac{2m+1}{p} + a + 1\right)} \frac{\Gamma\left(\frac{2m+2n+3}{p} + b + 1\right)}{\Gamma(2m+2n+2) \left(\frac{2m+2n+3}{p}\right) \cdots \left(\frac{2m+2n+3}{p} + b\right)} \\ &\leq p^{2n+1} \frac{(2m+1) \cdots (2m+1+ap)}{(2m+2n+3) \cdots (2m+2n+3+bp)}, \end{aligned}$$

where the inequality follows from (C.2) and (C.3). We further see that

$$\begin{aligned}
 & p^{\frac{2n+2}{p}} \frac{\Gamma((2m+2n+3)/p) \Gamma(2m+1)}{\Gamma((2m+1)/p) \Gamma(2m+2n+2)} \\
 & \leq p^{2n+1} \frac{2m+1}{2m+2n+3} \cdots \frac{2m+1+ap}{2m+2n+3+ap} \frac{1}{(2m+2n+3+(a+1)p) \cdots (2m+2n+3+bp)} \\
 & \leq p^{2n+1} \frac{1}{(2n+3+(a+1)p) \cdots (2n+3+bp)} \\
 & \leq p^{2n+1} \left( \frac{1}{(2n+3+(a+1)p)} \right)^{b-a} \\
 & \leq p \left( \frac{p}{(2n+3+(a+1)p)^{1-\frac{1}{p}}} \right)^{2n} \left( \frac{1}{(2n+3+(a+1)p)} \right)^{1-\frac{2}{p}}
 \end{aligned}$$

which tends to zero as  $n$  tends to infinity, uniformly in  $m$ . This concludes the proof of the lemma.  $\square$

## D Geometric information in sharp large deviation estimates

Fix  $p \in (1, \infty)$  and  $n \in \mathbb{N}$ . We now demonstrate how sharp large deviation estimates encode geometric properties of the underlying high-dimensional measure. First observe from the estimate in (2.18) that the leading order term that depends on  $\theta$  is  $R_a^n(\theta^n)$ , which, in turn, depends on  $\theta$  only through  $\Psi_{p,\theta}^n(\lambda_a)$ , as evident from its definition in (5.8). From the definitions in (2.4), (2.9) and (5.7), we have

$$\Psi_{p,\theta}^n(\lambda_a) = \frac{1}{n} \sum_{j=1}^n \Lambda_p(\sqrt{n}\theta_j^n \lambda_{a,1}, \lambda_{a,2}), \quad (\text{D.1})$$

where we suppress the  $\theta^n$  dependence in  $\Psi_{p,\theta}^n$ . We first state a lemma regarding the properties of  $\Lambda_p$  in [16].

**Lemma D.1** ([16, Lemma 7.5]). *Let  $p \in (1, \infty)$  and  $t_2 < 1/p$ . The map  $\mathbb{R}_+ \ni t_1 \mapsto \Lambda_p(\sqrt{t_1}, t_2)$  is concave but not linear for  $p > 2$ , linear for  $p = 2$  and convex but not linear for  $p < 2$ .*

*Proof of Proposition 2.7.* From the definition of  $R_a^n(\theta^n)$  in (5.8), it suffices to understand the behavior of (D.1). Since by (2.9),  $\lambda_{a,2} < 1/p$ . We may apply Lemma D.1 in the following proof.

First, for  $p = 2$ , from (2.4) and (5.2), a simple calculation yields

$$\Lambda_2(\lambda_{a,1}, \lambda_{a,2}) = -\frac{1}{2} \log(1 - 2\lambda_{a,2}) + \frac{1}{2} \frac{\lambda_{a,1}^2}{1 - 2\lambda_{a,2}}.$$

Hence, by (D.1) and the last display,  $\Psi_{p,\theta}^n(\lambda_a)$  does not depend on  $\theta$  and thus is a constant.

Next, consider  $p > 2$ . By D.1,  $\Lambda_p(\sqrt{\cdot}, \lambda_{a,2})$  is concave but not linear. By the definition of  $\Lambda_p$  in (2.4) and the symmetry of the  $p$ -Gaussian distribution (2.3),  $\Lambda_p(\cdot, \lambda_{a,2})$  is an even function. Therefore, for  $\theta^n \in \mathbb{S}^{n-1}$ ,

$$\begin{aligned}
 \frac{1}{n} \sum_{j=1}^n \Lambda_p(\sqrt{n}\theta_j^n \lambda_{a,1}, \lambda_{a,2}) &= \frac{1}{n} \sum_{j=1}^n \Lambda_p\left(\sqrt{n(\theta_j^n)^2 (\lambda_{a,1})^2}, \lambda_{a,2}\right) \\
 &\leq \Lambda_p\left(\sqrt{\sum_{j=1}^n (\theta_j^n)^2 (\lambda_{a,1})^2}, \lambda_{a,2}\right)
 \end{aligned}$$

$$= \Lambda_p(\sqrt{n}\lambda_{a,1}, \lambda_{a,2}).$$

Moreover, since  $\Lambda_p(\sqrt{\cdot}, \lambda_{a,2})$  is not linear, the equality in the last display holds only when

$$(\theta_1^n)^2 = (\theta_2^n)^2 = \dots = (\theta_n^n)^2 = \frac{1}{n}.$$

Thus, we conclude that the maximum of  $\Psi_{p,\theta}^n(\lambda_a)$  is attained at  $(\pm 1, \pm 1, \dots, \pm 1)/\sqrt{n}$ .

On the other hand, to identify the minimizers of  $\theta^n \mapsto \Psi_{p,\theta}^n(\lambda_a)$ , note from Lemma D.1 and the fact that  $\Lambda_p(\cdot, \lambda_{a,2})$  is even, we can write  $\Psi_{p,\theta}^n(\lambda_a) = \mathcal{F}(\theta_1^n, \dots, \theta_n^n)$ , where  $\mathcal{F} = \mathcal{F}_a$  is defined to be

$$\mathcal{F}(t_1, t_2, \dots, t_n) := \frac{1}{n} \sum_{j=1}^n \Lambda_p\left(\sqrt{nt_j(\lambda_{a,1})^2}, \lambda_{a,2}\right). \tag{D.2}$$

for  $(t_1, \dots, t_n)$  lies in the compact domain

$$\mathcal{A} := \left\{ (t_1, t_2, \dots, t_n) \in \mathbb{R}_+^n : \sum_{j=1}^n t_j = 1 \right\}.$$

Since  $\mathcal{F}$  is strictly concave by Lemma D.1, the minimum of  $\mathcal{F}$  is obtained at the extreme points of  $\mathcal{A}$ , namely, the vectors,  $\pm e_j, j = 1, \dots, n$ . Thus, by (D.2), the minimum of  $\Psi_{p,\theta}^n(\lambda_a)$  is also attained at

$$\theta^n = \pm e_j, \quad \text{for } j = 1, \dots, n.$$

The case  $p < 2$  follows from the same argument on interchanging maxima and minima, and invoking now the convexity of  $t_1 \mapsto \Lambda_p(\sqrt{t_1}, t_2)$  from Lemma D.1.  $\square$

## E Proof of Proposition 5.6

*Proof.* Fix  $m, d, D, h^n$  and  $(f, x^*, \alpha, g^n)$  as in the proposition. By setting  $\tilde{f}(x) = f(x) - f(x^*)$ , without loss of generality, we assume  $f(x^*) = 0$ . For  $k \in \mathbb{N}$  and any multi-index  $\beta = (\beta_1, \dots, \beta_k) \in \mathbb{N}^k$ , we define  $f_\beta := \partial_{\beta_1, \dots, \beta_k}^{| \beta |} f(x^*)$ . Since  $D \subset \mathbb{R}^m \times \mathbb{R}_+^d$  and  $x^* = (0, 0, \dots, 0)$ ,  $x^*$  lies in the boundary of  $D$ . Moreover, since  $f$  is twice continuously differentiable in  $D$  and its minimum over  $\text{cl}(D)$  is attained at  $x^*$ , we have  $\nabla f(x^*) \cdot (a_1, \dots, a_m, 0, \dots, 0) = 0$  for all  $a_i \in \mathbb{R}, i = 1, \dots, m$ , which implies that  $f_i = 0$  for  $i = 1, \dots, m$ . Moreover, since  $f$  achieves its minimum uniquely at  $x^*$ ,  $f_{ii} > 0$  for  $i = 1, \dots, m$ . By Taylor's theorem, we may then write  $f$  as

$$f(x) = \sum_{i=1}^m \frac{f_{i,i}}{2} x_i^2 (1 + P_i(x)) + \sum_{i=m+1}^d f_i x_i (1 + P_i(x)), \quad \text{for } x \in D$$

where  $(P_i)_{i=1, \dots, m+d}$  are continuously differentiable real-valued functions on  $D$ . We will proceed by making several changes of variables. We start with the transformation  $T_1 : \mathbb{R}^m \times \mathbb{R}_+^d \rightarrow \mathbb{R}^m \times \mathbb{R}_+^d$  where  $u = T_1(x)$  is defined to be

$$u_i = x_i(1 + P_i(x))^{1/2}, \quad i = 1, \dots, m \quad \text{and} \quad u_i = x_i(1 + P_i(x)), \quad i = m + 1, \dots, m + d.$$

Note that the Jacobian  $\mathcal{J}_1$  of  $T_1$  is 1 for  $x \in D$ . Let  $u^* := T_1(x^*) = (0, \dots, 0)$ . Let  $D' := T_1(D)$  and note that  $D' \subset \mathbb{R}^m \times \mathbb{R}_+^d$  contains a neighborhood of the origin. For  $u \in D'$ , define

$$F(u) := f(T_1^{-1}(u)) = \sum_{i=1}^m \frac{f_{i,i}}{2} u_i^2 + \sum_{i=m+1}^{m+d} f_i u_i. \tag{E.1}$$

Next, let  $m^*$  be the maximum of  $F$  in  $D'$  or equivalently, of  $f$  in  $D$  and define  $G^n(u) := g^n(T_1^{-1}(u))$ . Using the change of variables  $T_1$ , we see that

$$I^n := \int_D h^n(x)dx = \int_{D'} G^n(u)e^{-nF(u)}du. \tag{E.2}$$

Since  $D' \subset \mathbb{R}^m \times \mathbb{R}_+^d$ , we see that  $D'$  is covered by the family of surfaces  $F(u) = t$  for  $t \in [0, m^*]$ , that is,  $D' \subset \cup_{t \in [0, m^*]} \{u \in \mathbb{R}^{k+1} : F(u) = t\}$ . Note also that  $D'$  is bounded and  $\nabla F$  is nonzero in  $D'$ . Using the method of resolution of multiple integrals [42, Theorem 9, Chapter V], we then have

$$I^n = \int_0^{m^*} \mathfrak{r}^n(t)e^{-nt}dt, \tag{E.3}$$

where

$$\mathfrak{r}^n(t) := \int_{\{F(u)=t\}} \frac{G^n(u)}{\sqrt{\sum_{i=1}^{k+1} F_{u_i}^2}}dA, \quad t \in [0, m^*], \tag{E.4}$$

with  $dA$  denoting the surface element of the surface  $F(u) = t$ .

To further simplify the integral in (E.3)–(E.4), we introduce an additional change of variables  $T_2 : \mathbb{R}^m \times \mathbb{R}_+^d \rightarrow \mathbb{R}_+ \times [\pi/2, \pi, 2]^m \times [0, \pi/2]^{d-1}$  by letting  $(\xi, \theta, \phi) = T_2(u)$  be such that

$$\begin{aligned} u_i &= \left(\frac{2\xi}{f_{i,i}}\right)^{1/2} \cos \theta_1 \cdots \cos \theta_{i-1} \sin \theta_i, & i = 1, \dots, m, \\ u_{m+i} &= \frac{\xi}{f_{m+i}} \cos^2 \theta_1 \cdots \cos^2 \theta_m \cos^2 \phi_1 \cdots \cos^2 \phi_{i-1} \sin^2 \phi_i, & i = 1, \dots, d-1, \\ u_{m+d} &= \frac{\xi}{f_{m+d}} \cos^2 \theta_1 \cdots \cos^2 \theta_m \cos^2 \phi_1 \cdots \cos^2 \phi_{d-2} \cos^2 \phi_{d-1}, \end{aligned} \tag{E.5}$$

for  $\theta_i \in [-\pi/2, \pi/2]$ ,  $i = 1, \dots, m$ ,  $\phi_i \in [0, \pi/2]$ ,  $i = 1, \dots, d-1$  and  $\xi \in [0, m^*]$ . Since for  $i = m+1, \dots, m+d$ ,  $u_i \in \mathbb{R}_+$ , the domain of  $\phi_i$  is restricted to  $[0, \pi/2]$ , and thus we may take the square in cosines and sines for  $i = m+1, \dots, m+d$ . Therefore,  $T_2$  is a modified version of polar coordinates and is well defined. From (E.1) and (E.5),  $F(u) = \xi$  for  $u \in \mathbb{R}^{m+d}$  and the Jacobian  $\mathcal{J}_2$  of  $T_2$  is

$$\begin{aligned} \mathcal{J}_2 &:= \frac{\partial(u_1, \dots, u_{m+d})}{\partial(\xi, \theta_1, \dots, \theta_m, \phi_1, \dots, \phi_{d-1})} \\ &= \frac{(2\xi)^{m/2+d-1}}{f_{m+1} \cdots f_{m+d} \sqrt{f_{1,1} \cdots f_{m,m}}} \prod_{i=1}^m \cos^{2d+m-1-i} \theta_i \prod_{i=1}^{d-1} \cos^{2d-1-2i} \phi_i \sin \phi_i. \end{aligned} \tag{E.6}$$

From the second change of variables in (E.5), we have

$$\frac{dA}{\|\nabla F\|} = \mathcal{J}_2 d\theta d\phi. \tag{E.7}$$

Recall  $g^n(x) = e^{r^n(x)}$  and  $G^n(u) = g^n(T_1^{-1}(u))$ . Let  $\hat{r}^n(\xi, \theta, \phi)$  denotes the transformation of  $r^n$  under  $T_2 \circ T_1$ . Then, (E.4), (E.6) and (E.7) imply that

$$\begin{aligned} \mathfrak{r}^n(t) &= \frac{(2t)^{m/2+d-1}}{f_{m+1} \cdots f_{m+d} \sqrt{f_{1,1} \cdots f_{m,m}}} \\ &\times \int_{\theta \in [-\pi/2, \pi/2]^m} \int_{\phi \in [0, \pi/2]^{d-1}} e^{\hat{r}^n(t, \theta, \phi)} \prod_{i=1}^m \cos^{2d+m-1-i} \theta_i \prod_{i=1}^{d-1} \cos^{2d-1-2i} \phi_i \sin \phi_i d\phi d\theta \end{aligned} \tag{E.8}$$

$$= \frac{(2t)^{m/2+d-1}}{f_{m+1} \cdots f_{m+d} \sqrt{f_{1,1} \cdots f_{m,m}}} e^{\tilde{r}^n(t)}, \tag{E.9}$$

where

$$\tilde{r}^n(t) = \log \int_{\theta \in [-\pi/2, \pi/2]^m} \int_{\phi \in [0, \pi/2]^{d-1}} e^{\tilde{r}^n(t, \theta, \phi)} \prod_{i=1}^m \cos^{2d+m-1-i} \theta_i \prod_{i=1}^{d-1} \cos^{2d-1-2i} \phi_i \sin \phi_i d\phi d\theta. \tag{E.10}$$

Since  $|r^n(x)| \leq Cn^\alpha \|x\|$  for  $n$  large and  $x$  in a neighborhood of the origin, there exist  $\varepsilon > 0$  and  $\tilde{r}^n(t)$  such that  $|\tilde{r}^n(t)| \leq Cn^\alpha t$  for  $n$  large and  $t \in (0, \varepsilon)$ .

From (E.3) and (E.9), we observe that

$$I^n = \frac{2^{m/2+d-1}}{f_{m+1} \cdots f_{m+d} \sqrt{f_{1,1} \cdots f_{m,m}}} \int_0^{m^*} t^{m/2+d-1} e^{-nt+\tilde{r}^n(t)} dt.$$

Now, applying [31, Chapter 9, Theorem 2.1] with  $p, r, q, \lambda, \mu,$  and  $\nu$  being  $t, \tilde{r}^n, t^{m/2+d-1}, m/2 + d, 1$  and  $1$ , we obtain

$$\begin{aligned} I^n &= \frac{2^{m/2+d-1}}{f_{m+1} \cdots f_{m+d} \sqrt{f_{1,1} \cdots f_{m,m}}} \prod_{i=1}^m \int_{-\pi/2}^{\pi/2} \cos^{2d+m-1-i} \theta d\theta \\ &\quad \times \prod_{i=1}^{d-1} \int_0^{2\pi} \cos^{2d-1-2i} \phi \sin \phi d\phi \frac{\Gamma(m/2 + d)}{n^{m/d+2}} \\ &= \frac{(2\pi)^{m/2} g^n(x^*)}{n^{d+m/2} f_{m+1} \cdots f_{m+d} \sqrt{f_{1,1} \cdots f_{m,m}}} (1 + o(1)). \end{aligned}$$

□

## F A uniform deviation estimate

We now establish Lemma 5.9. Key ingredients of the proof include the Gaussian concentration inequality and certain deviation estimates that are uniform with respect to a class of functions, much in the spirit of uniform Glivenko-Cantelli or Donsker classes.

*Proof of Lemma 5.9.* Fix  $\varepsilon > 0$ . Also, consider  $x \in \mathbb{R}$  and  $(t_1, t_2) \in \mathbb{D}$ . We will repeatedly use the fact that  $(xt_1, t_2)$  lies in  $\mathbb{D}$ . By the assumed differentiability properties of  $\mathcal{K}$ , an application of Taylor’s theorem shows that

$$\mathcal{K}(xt_1, t_2) = \mathcal{K}(0, t_2) + \partial_1 \mathcal{K}(\rho(xt_1, t_2), t_2) xt_1, \tag{F.1}$$

where  $\rho : \mathbb{D} \rightarrow \mathbb{R}$  is a function that satisfies

$$|\rho(y, t)| \leq |y| \quad \text{for } (y, t) \in \mathbb{D}. \tag{F.2}$$

$|\rho(xt_1, t_2)| \leq |xt_1|$ . By the polynomial growth assumption on the partial derivatives of  $\mathcal{K}$ , there exist  $q, \tilde{C} \in (0, \infty)$  such that

$$\sup_{t \in \mathbb{D}, \|t\| < \varepsilon} |\partial_1 \mathcal{K}(xt_1, t_2)| \leq \tilde{C}(1 + \varepsilon^q |x|^q), \tag{F.3}$$

$$\sup_{t \in \mathbb{D}, \|t\| < \varepsilon} |\partial_{1j} \mathcal{K}(xt_1, t_2)| \leq \tilde{C}(1 + \varepsilon^q |x|^q), \quad \text{for } j = 1, 2. \tag{F.4}$$

By (F.1), for  $t \in \mathbb{D}, \|t\| < \varepsilon$ , we see that

$$\sum_{j=1}^n \left( \mathcal{K} \left( \frac{\sqrt{n} Z_j}{\|Z^{(n)}\|} t_1, t_2 \right) - \mathbb{E} [\mathcal{K}(Z t_1, t_2)] \right) = t_1 (I_1^n(t) + I_2^n(t)), \tag{F.5}$$

where

$$I_1^n(t) := \sum_{j=1}^n \left( \partial_1 \mathcal{K} \left( \rho \left( \frac{\sqrt{n} Z_j}{\|Z^{(n)}\|} t_1, t_2 \right), t_2 \right) \frac{\sqrt{n} Z_j}{\|Z^{(n)}\|} - \partial_1 \mathcal{K}(\rho(Z_j t_1, t_2), t_2) Z_j \right), \quad (\text{F.6})$$

$$I_2^n(t) := \sum_{j=1}^n (\partial_1 \mathcal{K}(\rho(Z_j t_1, t_2), t_2) Z_j - \mathbb{E}[\partial_1 \mathcal{K}(\rho(Z t_1, t_2), t_2) Z]). \quad (\text{F.7})$$

The proof follows in several steps.

**Step 1.** We claim that for  $\alpha \in (1/2, 1)$ , there exist  $C_1 \in (0, \infty)$  and a random integer  $N_1$  such that  $\mathbb{P}'$ -almost surely,

$$\sup_{t \in \mathbb{D}, \|t\| < \varepsilon} |n^{-\alpha} I_1^n(t)| \leq C_1 \quad \text{for } n \geq N_1. \quad (\text{F.8})$$

**Proof of claim of Step 1.** Note from (F.1) that for  $x \in \mathbb{R}$ , and  $t \in \mathbb{D}$ ,  $\|t\| < \varepsilon$ ,

$$\frac{d}{dx} (\partial_1 \mathcal{K}(\rho(x t_1, t_2), t_2) x) = \begin{cases} \frac{d}{dx} \left( \frac{\mathcal{K}(x t_1, t_2) - \mathcal{K}(0, t_2)}{t_1} \right) = \partial_1 \mathcal{K}(x t_1, t_2), & \text{if } t_1 \neq 0, \\ \partial_1 \mathcal{K}(0, t_2), & \text{if } t_1 = 0. \end{cases}$$

Together with (F.3), this implies that for  $x \in \mathbb{R}$ ,

$$\sup_{t \in \mathbb{D}, \|t\| < \varepsilon} \left| \frac{d}{dx} (\partial_1 \mathcal{K}(\rho(x t_1, t_2), t_2) x) \right| = \sup_{t \in \mathbb{D}, \|t\| < \varepsilon} |\partial_1 \mathcal{K}(x t_1, t_2)| \leq \tilde{C}(1 + \varepsilon^q |x|^q).$$

Combining the last two displays, we see that

$$\begin{aligned} \sup_{t \in \mathbb{D}, \|t\| < \varepsilon} |n^{-\alpha} I_1^n(t)| &\leq n^{-\alpha} \sup_{t \in \mathbb{D}, \|t\| < \varepsilon} \sum_{j=1}^n \sup_{x \in \left[ \frac{\sqrt{n} Z_j}{\|Z^{(n)}\|}, Z_j \right]} |\partial_1 \mathcal{K}(x t_1, t_2)| \left| \frac{\sqrt{n} Z_j}{\|Z^{(n)}\|} - Z_j \right| \\ &\leq \left| \frac{(\sqrt{n} - \|Z^{(n)}\|) n^{1/2-\alpha}}{\|Z^{(n)}\| / \sqrt{n}} \right| \frac{\tilde{C}}{n} \sum_{j=1}^n \left( 1 + \varepsilon^q \max \left( \frac{\sqrt{n}}{\|Z^{(n)}\|}, 1 \right)^q |Z_j|^q \right) |Z_j|. \end{aligned} \quad (\text{F.9})$$

Since  $(Z_j)_{j \in \mathbb{N}}$  are independent, by the strong law of large numbers,  $\mathbb{P}'$ -almost surely, as  $n \rightarrow \infty$ ,

$$\frac{1}{n} \sum_{j=1}^n (1 + (\tilde{C}\varepsilon)^q |Z_j|^q) |Z_j| \rightarrow \mathbb{E} \left[ (1 + (\tilde{C}\varepsilon)^q |Z|^q) |Z| \right] \quad \text{and} \quad \frac{\|Z^{(n)}\|}{\sqrt{n}} \rightarrow 1. \quad (\text{F.10})$$

Furthermore, the Gaussian concentration inequality [38, Theorem 3.1.1], implies that there exists  $c \in (0, \infty)$  such that

$$\mathbb{P} \left( n^{1/2-\alpha} \left| \|Z^{(n)}\| - \sqrt{n} \right| > \tilde{C} \right) = \mathbb{P} \left( \left| \|Z^{(n)}\| - \sqrt{n} \right| > n^{\alpha-1/2} \tilde{C} \right) \leq 2e^{-c(\tilde{C})^2 n^{2\alpha-1}},$$

which is summable because  $\alpha > 1/2$ . Hence, by the Borel-Cantelli lemma and the second limit in (F.10), there exists a random integer  $N'_1 \in \mathbb{N}$  such that  $\mathbb{P}'$ -almost surely,

$$n^{1/2-\alpha} \left| \|Z^{(n)}\| - \sqrt{n} \right| \leq \tilde{C} \quad \text{and} \quad \max \left( \frac{\sqrt{n}}{\|Z^{(n)}\|}, 1 \right) \leq \tilde{C}, \quad n \geq N'_1. \quad (\text{F.11})$$

The claim of Step 1 then follows from (F.9), (F.10), and (F.11).

**Step 2.** We now establish a bound on  $I_2^n(t)$  defined in (F.7). Specifically we show that there exist  $C_2 \in (0, \infty)$  and a random integer  $N_2 \in \mathbb{N}$  such that  $\mathbb{P}'$ -almost surely, for  $n \geq N_2$ ,

$$\sup_{\|t\| < \varepsilon} |n^{-\alpha} I_2^n(t)| \leq C_2, \quad \text{for } n \geq N_2. \tag{F.12}$$

Before proving this bound, we first show how when combined with Step 1, this proves the lemma. Indeed, (F.5), (F.8) and (F.12) together show that there exist  $C_1, C_2 \in (0, \infty)$  such that almost surely, for  $t \in \mathbb{D}$ ,  $\|t\| \leq \varepsilon$  and  $n \geq \max\{N_1, N_2\}$ ,

$$\left| \sum_{j=1}^n \mathcal{K} \left( \frac{\sqrt{n} Z_j}{\|Z^{(n)}\|} t_1, t_2 \right) - \mathbb{E}[\mathcal{K}(Z t_1, t_2)] \right| \leq n^\alpha |t_1| (C_1 + C_2) \leq (C_1 + C_2) n^\alpha \|t\|.$$

This implies that (5.24) holds  $\mathbb{P}'$ -almost surely and concludes the proof of the lemma.

To complete the proof of the lemma, it only remains to prove the bound in Step 2.

**Proof of bound in Step 2.** To prove (F.12), we introduce a suitable truncation of  $(x, (t_1, t_2)) \mapsto \partial_1 \mathcal{K}(\rho(x t_1, t_2), t_2) x$ . To this end, recall the definition of  $q$  in (F.3). Since  $\alpha > 1/2$ , we may choose  $\beta > 0$  so that

$$2\alpha - 1 - 2\beta q > 0 \quad \text{and define } r_n := n^\beta. \tag{F.13}$$

Then for  $x \in \mathbb{R}$  and  $t \in \mathbb{D}$ ,  $\|t\| < \varepsilon$ , define

$$\mathbb{T}_n(x, t) := \begin{cases} \partial_1 \mathcal{K}(\rho(x t_1, t_2), t_2) x, & \text{if } |x| < r_n, \\ \partial_1 \mathcal{K}(\rho(\text{sgn}(x) r_n t_1, t_2), t_2) \text{sgn}(x) r_n, & \text{if } |x| \geq r_n, \end{cases} \tag{F.14}$$

where  $\text{sgn} : \mathbb{R} \rightarrow \{-1, 1\}$  is defined by  $\text{sgn}(x) = 1$  if  $x \geq 0$  and  $\text{sgn}(x) = -1$  if  $x < 0$ . We bound (F.12) above by the sum of three terms:

$$\sup_{\|t\| < \varepsilon} |n^{-\alpha} I_2^n(t)| \leq I_{21}^n + I_{22}^n + I_{23}^n, \tag{F.15}$$

where

$$I_{21}^n := \sup_{t \in \mathbb{D}, \|t\| < \varepsilon} \left| \frac{1}{n^\alpha} \sum_{j=1}^n (\mathbb{T}_n(Z_j, t) - \mathbb{E}[\mathbb{T}_n(Z, t)]) \right|, \tag{F.16}$$

$$I_{22}^n := \sup_{t \in \mathbb{D}, \|t\| < \varepsilon} \left| \frac{1}{n^\alpha} \sum_{j=1}^n (\partial_1 \mathcal{K}(\rho(Z_j t_1, t_2), t_2) Z_j - \mathbb{T}_n(Z_j, t)) \right|, \tag{F.17}$$

$$I_{23}^n := \sup_{t \in \mathbb{D}, \|t\| < \varepsilon} \left| \mathbb{E} \left[ \frac{1}{n^\alpha} \sum_{j=1}^n (\partial_1 \mathcal{K}(\rho(Z_j t_1, t_2), t_2) Z_j - \mathbb{T}_n(Z_j, t)) \right] \right|. \tag{F.18}$$

We now treat each of these terms individually.

**Step 2A.** For the first term  $I_{21}^n$  in (F.16), we start by proving that there exist  $C_3 < \infty$  and a random integer  $N_3$  such that  $\mathbb{P}'$ -almost surely,

$$I_{21}^n \leq C_3 + 2\varepsilon \quad \text{for } n \geq N_3. \tag{F.19}$$

**Proof of Step 2A bound.** The proof of (F.19) starts with the following claim.

**Claim.** For  $x \in \mathbb{R}$  and  $t \in \mathbb{D}$ ,  $\|t\| < \varepsilon$ ,  $t \mapsto \mathbb{T}_n(x, t)$  is Lipschitz continuous with constant  $\tilde{C}(1 + \varepsilon^q r_n^q) r_n^2$ .

*Proof of the Claim.* Define

$$x_n := \begin{cases} x, & \text{if } x < r_n, \\ \text{sgn}(x)r_n, & \text{if } x \geq r_n. \end{cases}$$

First, note that since  $\mathcal{K}$  is twice continuously differentiable, by (F.1) and (F.14), for  $t \in \mathbb{D}$ ,  $\|t\| < \varepsilon$  and  $t_1 \neq 0$ ,

$$\begin{aligned} \partial_{t_1} \mathbb{T}_n(x, t) &= \partial_{t_1} \left( \frac{\mathcal{K}(xt_1, t_2) - \mathcal{K}(0, t_2)}{t_1} \right) = \frac{xt_1 \partial_1 \mathcal{K}(xt_1, t_2) - \mathcal{K}(xt_1, t_2) + \mathcal{K}(0, t_2)}{t_1^2} \\ &= -\frac{1}{2} \xi^2 \partial_{11} \mathcal{K}(\xi, t_2), \end{aligned}$$

where the second equality follows from Taylor's theorem with  $\xi$  being a constant such that  $|\xi| \leq |x_n t_1| \leq \varepsilon r_n$ . Likewise, by (F.2),  $\rho(0, t_2) = 0$ , and for  $t \in \mathbb{D}$ ,  $\|t\| < \varepsilon$  and  $t_1 = 0$ ,

$$\begin{aligned} \partial_{t_1} \mathbb{T}_n(x, t)|_{t_1=0} &= \lim_{t_1 \rightarrow 0} \left( \frac{\partial_1 \mathcal{K}(\rho(xt_1, t_2), t_2) x - \partial_1 \mathcal{K}(0, t_2) x}{t_1} \right) \\ &= \lim_{t_1 \rightarrow 0} \frac{\mathcal{K}(xt_1, t_2) - \mathcal{K}(0, t_2) - \partial_1 \mathcal{K}(0, t_2) xt_1}{t_1^2} \\ &= \frac{1}{2} x^2 \partial_{11} \mathcal{K}(0, t_2). \end{aligned}$$

By (F.4) and the last two displays, for  $t \in \mathbb{D}$ ,  $\|t\| < \varepsilon$ , we have

$$|\partial_{t_1} \mathbb{T}_n(x, t)| \leq \tilde{C}(1 + \varepsilon^q r_n^q) r_n^2. \tag{F.20}$$

Similarly, for  $x \in \mathbb{R}$  and  $t \in \mathbb{D}$ ,  $\|t\| < \varepsilon$  by (F.1), (F.14) and Taylor's theorem, when  $t_1 \neq 0$ ,

$$\begin{aligned} \partial_{t_2} \mathbb{T}_n(x, t) &= \partial_{t_2} \left( \frac{\mathcal{K}(xt_1, t_2) - \mathcal{K}(0, t_2)}{t_1} \right) = \frac{\partial_2 \mathcal{K}(xt_1, t_2) - \partial_2 \mathcal{K}(0, t_2)}{t_1} \\ &= \xi \partial_{12} \mathcal{K}(\xi, t_2), \end{aligned}$$

where  $\xi$  is a constant such that  $|\xi| \leq |x_n t_1|$  and when  $t_1 = 0$ ,

$$\partial_{t_2} \mathbb{T}_n(x, t) = \partial_{t_2} (\partial_1 \mathcal{K}(0, t_2) x) = x \partial_{12} \mathcal{K}(0, t_2).$$

The last two displays and (F.4) imply that

$$|\partial_{t_2} \mathbb{T}_n(x, t)| \leq \tilde{C}(1 + \varepsilon^q r_n^q) r_n, \quad \text{for } x \in \mathbb{R} \text{ and } \|t\| \leq \varepsilon. \tag{F.21}$$

Thus, the claim follows from (F.20) and (F.21).  $\square$

We now continue with the proof of Step 2A. For  $n \in \mathbb{N}$ , let  $\delta_n$  and  $k_n$  be finite positive constants given by

$$\delta_n := \frac{\varepsilon}{\tilde{C}(1 + \varepsilon^q r_n^q) r_n^2 n^{1-\alpha}} \quad \text{and} \quad k_n := \left\lceil \left( \tilde{C}(1 + \varepsilon^q r_n^q) r_n^2 n^{1-\alpha} \right)^2 \right\rceil. \tag{F.22}$$

Given the claim, there exist  $(l_k)_{k=1, \dots, k_n} \subset \{t \in \mathbb{D} : \|t\| \leq \varepsilon\}$  such that  $\cup_{k=1}^{k_n} B_{\delta_n}(l_k) \supset \{t \in \mathbb{D} : \|t\| \leq \varepsilon\}$  and for  $x \in \mathbb{R}$ ,

$$|\mathbb{T}_n(x, u) - \mathbb{T}_n(x, v)| \leq \|u - v\| \tilde{C}(1 + \varepsilon^q r_n^q) r_n^2 \leq \frac{2\varepsilon}{n^{1-\alpha}} \quad \text{for } u, v \in B_{\delta_n}(l_k), \tag{F.23}$$

where the last inequality uses  $\|u - v\| \leq 2\delta_n$  and (F.22). Together with the expression for  $I_{21}^n$  in (F.16), this shows that

$$\begin{aligned}
 I_{21}^n &\leq \sup_{k=1, \dots, k_n} \sup_{t \in B_{\delta_n}(l_k)} \left| \frac{1}{n^\alpha} \sum_{j=1}^n (\mathbb{T}_n(Z_j, t) - \mathbb{E}[\mathbb{T}_n(Z, t)]) \right| \\
 &\leq \sup_{k=1, \dots, k_n} \left| \frac{1}{n^\alpha} \sum_{j=1}^n \left( \sup_{t \in B_{\delta_n}(l_k)} \mathbb{T}_n(Z_j, t) - \mathbb{E} \left[ \sup_{t \in B_{\delta_n}(l_k)} \mathbb{T}_n(Z_j, t) \right] \right) \right| \\
 &\quad + \sup_{k=1, \dots, k_n} n^{1-\alpha} \left| \mathbb{E} \left[ \sup_{t \in B_{\delta_n}(l_k)} \mathbb{T}_n(Z, t) - \inf_{t \in B_{\delta_n}(l_k)} \mathbb{T}_n(Z, t) \right] \right| \\
 &\quad + \sup_{k=1, \dots, k_n} n^{1-\alpha} \left| \sup_{t \in B_{\delta_n}(l_k)} \mathbb{T}_n(Z, t) - \inf_{t \in B_{\delta_n}(l_k)} \mathbb{T}_n(Z, t) \right| \\
 &\leq \sup_{k=1, \dots, k_n} \left| \frac{1}{n^\alpha} \sum_{j=1}^n \left( \sup_{t \in B_{\delta_n}(l_k)} \mathbb{T}_n(Z_j, t) - \mathbb{E} \left[ \sup_{t \in B_{\delta_n}(l_k)} \mathbb{T}_n(Z_j, t) \right] \right) \right| + 4\varepsilon, \quad (\text{F.24})
 \end{aligned}$$

where the last inequality follows from (F.23). In addition, note that, together, (F.3) and (F.14) imply

$$\left| \sup_{t \in B_{\delta_n}(l_k)} \mathbb{T}_n(Z_j, t) \right| \leq \tilde{C}(1 + \varepsilon^q r_n^q).$$

Hence, by the union bound and Hoeffding's inequality [38, Theorem 2.2.6], which is applicable since  $(Z_j)_{j \in \mathbb{N}}$  are i.i.d., for any  $n \in \mathbb{N}$  and  $C_3 \in (0, \infty)$  we have

$$\begin{aligned}
 &\mathbb{P} \left( \sup_{k=1, \dots, k_n} \left| \frac{1}{n^\alpha} \sum_{j=1}^n \left( \sup_{t \in B_{\delta_n}(l_k)} \mathbb{T}_n(Z_j, t) - \mathbb{E} \left[ \sup_{t \in B_{\delta_n}(l_k)} \mathbb{T}_n(Z_j, t) \right] \right) \right| > C_3 \right) \\
 &\leq \sum_{k=1}^{k_n} \mathbb{P} \left( \left| \frac{1}{n^\alpha} \sum_{j=1}^n \left( \sup_{t \in B_{\delta_n}(l_k)} \mathbb{T}_n(Z_j, t) - \mathbb{E} \left[ \sup_{t \in B_{\delta_n}(l_k)} \mathbb{T}_n(Z_j, t) \right] \right) \right| > C_3 \right) \\
 &\leq 2k_n \exp \left( -\frac{2C_3^2 n^{2\alpha}}{n\tilde{C}^2(1 + \varepsilon^q r_n^q)^2} \right),
 \end{aligned}$$

which is summable in  $n$  by (F.22) and (F.13). Step 2A then follows from the Borel-Cantelli lemma and (F.24).

**Step 2B.** Next, we deal with the quantity  $I_{22}^n$  in (F.17) and show that there exists  $C_4 < \infty$  and a random integer  $N_5$  such that  $\mathbb{P}'$ -almost surely, for  $n \geq N_5$ ,

$$I_{22}^n \leq C_4, \quad \text{and} \quad I_{23}^n \leq C_4. \quad (\text{F.25})$$

**Proof of Step 2B bounds.** Note that by (F.17), (F.14), (F.3) and (F.2),

$$I_{22}^n \leq \frac{1}{n^\alpha} \sum_{i=1}^n \left( |Z_j| \tilde{C}(1 + \varepsilon^q |Z_j|^q) + r_n \tilde{C}(1 + \varepsilon^q r_n^q) \right) 1_{\{|Z_j| > r_n\}}.$$

Hence, Markov's inequality and the fact that  $(Z_j)_{j \in \mathbb{N}}$  are i.i.d. imply that for any  $C_4 \in (0, \infty)$ ,

$$\mathbb{P}(I_{22}^n > C_4) \leq \mathbb{P} \left( \frac{1}{n^\alpha} \sum_{i=1}^n \left( |Z_j| \tilde{C}(1 + \varepsilon^q |Z_j|^q) + r_n \tilde{C}(1 + \varepsilon^q r_n^q) \right) 1_{\{|Z_j| > r_n\}} > C_4 \right)$$

## Sharp large deviations for random projections of $\ell_p^n$ balls

$$\leq \frac{1}{C_4} n^{1-\alpha} \mathbb{E} \left[ \left( |Z_j| \tilde{C}(1 + \varepsilon^q |Z_j|^q) + r_n \tilde{C}(1 + \varepsilon^q r_n^q) \right) 1_{\{|Z_j| > r_n\}} \right].$$

Now, for any  $k \in \mathbb{N} \cup \{0\}$ , the Laplace approximation (see e.g. [42, Chapter 2]) implies

$$\mathbb{E} \left[ |Z|^k 1_{\{|Z| > r_n\}} \right] = 2 \int_{\{x > r_n\}} x^k \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 2r_n^{k-\frac{1}{2}} e^{-\frac{r_n^2}{2}} (1 + o(1)).$$

Hence, there exist  $C'_4 \in (0, \infty)$  and  $N_4 \in \mathbb{N}$  such that for  $n \geq N_4$  and  $k \in \mathbb{N} \cup \{0\}$ ,

$$\mathbb{E} \left[ |Z|^k 1_{\{|Z| > r_n\}} \right] \leq C'_4 r_n^k e^{-\frac{r_n^2}{2}}.$$

The last three displays together yield the following bound on the tail probability of  $I_{22}^n$ :

$$\mathbb{P}(I_{22}^n > C_4) \leq \frac{C'_4}{C_4} n^{1-\alpha} \left( r_n \tilde{C}(1 + \varepsilon^q r_n^q) \right) e^{-r_n^2/2}.$$

Since this is summable in  $n$  due to the definition of  $r_n$  in (F.13), the first inequality in (F.25) follows from the Borel-Cantelli lemma.

Since (F.17) and (F.18) imply  $I_{23}^n \leq \mathbb{E}[I_{22}^n]$ , the second inequality in (F.25) follows from the first. This concludes Step 2B. Moreover, when combined with (F.15), Step 2A and Step B prove the claim of Step 2.  $\square$

## References

- [1] D. Alonso-Gutiérrez, J. Prochno, and C. Thäle, *Large deviations for high-dimensional random projections of  $\ell_p^n$ -balls*, *Advances in Applied Mathematics* **99** (2018), 1–35. MR3806754
- [2] C. Andriani and P. Baldi, *Sharp estimates of deviations of the sample mean in many dimensions*, *Annales de l'Institut Henri Poincaré (B) Probability and Statistics* **33** (1997), no. 3, 371–385. MR1457057
- [3] S. Artstein-Avidan, A. Giannopoulos, and V. D. Milman, *Asymptotic geometric analysis. Part I*, *Mathematical Surveys and Monographs*, vol. 202, American Mathematical Society, Providence, RI, 2015. MR3331351
- [4] R. R. Bahadur and R. Ranga Rao, *On deviations of the sample mean*, *Annals of Mathematical Statistics* **31** (1960), 1015–1027. MR0117775
- [5] P. Barbe and M. Broniatowski, *On sharp large deviations for sums of random vectors and multidimensional Laplace approximation*, *Theory of Probability and Its Applications* **49** (2005), no. 4, 561–588. MR2142565
- [6] N. Bleistein and R. Handelsman, *Asymptotic expansions of integrals*, Courier Corporation, 1986. MR0863284
- [7] S. G. Bobkov, *On concentration of distributions of random weighted sums*, *Annals of Probability* **31** (2003), 195–215. MR1959791
- [8] K. Borovkov, *On the convergence of projections of uniform distributions on balls*, *Theory of Probability and Its Applications* **35** (1991), no. 3, 546–550. MR1091211
- [9] A. Bovier and H. Mayer, *A conditional strong large deviation result and a functional central limit theorem for the rate function*, *Latin American Journal of Probability and Mathematical Statistics* **12** (2015), 533–550. MR3373797
- [10] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Bartlett Publishers, Boston, MA, 1993. MR1202429
- [11] P. Diaconis and D. Freedman, *Asymptotics of graphical projection pursuit*, *Annals of Statistics* **12** (1984), no. 3, 793–815. MR0751274
- [12] R. Durrett, *Probability: theory and examples*, *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, 2010. MR2722836
- [13] B. Fleury, *Between Paouris concentration inequality and variance conjecture*, *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, vol. 46, 2010, pp. 299–312. MR2667700

- [14] G. Folland, *Real analysis: modern techniques and their applications*, New York: Wiley, 1999. MR1681462
- [15] N. Gantert, S. S. Kim, and K. Ramanan, *Cramér’s theorem is atypical*, Advances in the Mathematical Sciences: Research from the 2015 Association for Women in Mathematics Symposium, Springer International Publishing, Cham, 2016, pp. 253–270. MR3654500
- [16] N. Gantert, S. S. Kim, and K. Ramanan, *Large deviations for random projections of  $\ell^p$  balls*, Annals of Probability **45** (2017), 4419–4476. MR3737915
- [17] O. Guédon and E. Milman, *Interpolating thin-shell and sharp large-deviation estimates for isotropic log-concave measures*, Geometric and Functional Analysis **21** (2011), no. 5, 1043. MR2846382
- [18] C. Joutard, *Multidimensional strong large deviation results*, Metrika **80** (2017), no. 6-8, 663–683. MR3716610
- [19] I. Kabluchko, J. Prochno, and C. Thäle, *High-dimensional limit theorems for random vectors in  $\ell_p^n$  balls. II*, 2021. MR3904638
- [20] Z. Kabluchko, J. Prochno, and C. Thäle, *High-dimensional limit theorems for random vectors in  $\ell_p^n$ -balls*, Communications in Contemporary Mathematics (2017), 1750092. MR3904638
- [21] T. Kaufmann, *Sharp asymptotics for  $q$ -norms of random vectors in high-dimensional  $\ell_p^n$ -balls*, Modern Stochastics: Theory and Applications **8** (2021), no. 2, 239–274. MR4279877
- [22] S. S. Kim, *Problems at the interface of probability and convex geometry: random projections and constrained processes*, Ph.D. thesis, Brown University, 2017.
- [23] S. S. Kim, Y.-T. Liao, and K. Ramanan, *An asymptotic thin shell condition and large deviations for multidimensional projections*, Advances in Applied Mathematics **134** (2022), 102306. MR4352552
- [24] S. S. Kim and K. Ramanan, *A conditional limit theorem for high-dimensional  $\ell_p$  spheres*, Journal of Applied Probability **55** (2018), 1060–1077. MR3899928
- [25] S. S. Kim and K. Ramanan, *Large deviation principles induced by the Stiefel manifold, and random multi-dimensional projections*, Electronic Journal of Probability, to appear, 2023.
- [26] B. Klartag, *A central limit theorem for convex sets*, Inventiones Mathematicae **168** (2007), 91–131. MR2285748
- [27] Y.-T. Liao, *Sharp large deviation estimates and their applications to asymptotic convex geometry*, Ph.D. thesis, Brown University, 2022.
- [28] E. Meckes, *Approximation of projections of random vectors*, Journal of Theoretical Probability **25** (2012), no. 2, 333–352. MR2914432
- [29] V. D. Milman and G. Schechtman (eds.), *Geometric aspects of functional analysis*, Israel Seminar, 2004. MR2089261
- [30] S. Nadarajah, *A generalized normal distribution*, Journal of Applied Statistics **32** (2005), no. 7, 685–694. MR2119411
- [31] F. Olver, *Asymptotics and special functions*, AK Peters/CRC Press, 1997. MR1429619
- [32] G. Paouris, *Small ball probability estimates for log-concave measures*, Transactions of the American Mathematical Society **364** (2012), no. 1, 287–308. MR2833584
- [33] J. Prochno, C. Thäle, and N. Turchi, *Geometry of  $\ell_p^n$ -balls: Classical results & recent developments*, High Dimensional Probability VIII, Progress in Probability, Birkhäuser, 2019. MR4181365
- [34] R. Rockafellar, *Convex analysis*, Princeton University Press, 1970. MR0274683
- [35] G. Schechtman and J. Zinn, *On the volume of the intersection of two  $\ell_p^n$  balls*, Proceedings of the American Mathematical Society **110** (1990), no. 1, 217–217. MR1015684
- [36] M. Schmuckenschläger, *CLT and the volume of intersections of  $\ell_p^n$ -balls*, Geometriae Dedicata **85** (2001), 189–195. MR1845607
- [37] V. N. Sudakov, *Typical distributions of linear functionals in finite-dimensional spaces of high dimension*, Doklady Akademii Nauk SSSR **243** (1978), no. 6, 1402–1405. MR0517198
- [38] R. Vershynin, *High-dimensional probability: an introduction with applications in data science*, vol. 47, Cambridge University Press, 2018. MR3837109

- [39] C. Villani, *Optimal transport: old and new*, vol. 338, Springer Science & Business Media, 2008. MR2459454
- [40] H. von Weizsäcker, *Sudakov's typical marginals, random linear functionals and a conditional central limit theorem*, *Probability Theory and Related Fields* **107** (1997), no. 3, 313–324. MR1440135
- [41] J. G. Wendel, *Note on the Gamma function*, *The American Mathematical Monthly* **55** (1948), no. 9, 563–564. MR0029448
- [42] R. Wong, *Asymptotic approximations of integrals*, Society for Industrial and Applied Mathematics, 2001. MR1851050
- [43] R. Zia, E. Redish, and S. McKay, *Making sense of the Legendre transform*, *American Journal of Physics* **77** (2009), no. 7, 614–622.