Sharp large deviation estimates and their applications to asymptotic convex geometry

by

Yin-Ting Liao

B.S., National Taiwan University, Taiwan, 2015

M.S., Brown University, Providence, RI, 2019

A dissertation submitted in partial fulfillment of the

requirements for the degree of Doctor of Philosophy

in the Division of Applied Mathematics at Brown University

PROVIDENCE, RHODE ISLAND

October 2022

This dissertation by Yin-Ting Liao is accepted in its present form
by the Division of Applied Mathematics as satisfying the
dissertation requirement for the degree of Doctor of Philosophy.

Date________________ ____

Kavita Ramanan, Ph.D., Advisor

Recommended to the Graduate Council

Date________________ ____

Paul Dupuis, Ph.D., Reader

Date________________ ____

Grigoris Paouris, Ph.D., Reader

Approved by the Graduate Council

Date________________ ____

Andrew G. Campbell, Dean of the Graduate School

Vita

Yin-Ting Liao received his B.S. in Mathematics and B.S. in Electrical Engineering from National Taiwan University, Taiwan in 2015. After working a year as a research assistant at Institute of Mathematics, Academia Sinica in Taiwan, he enrolled as a Ph.D. student in the Division of Applied Mathematics at Brown University in 2017. While at Brown, he has worked under the supervision of Prof. Kavita Ramanan. He was supported by a Government Scholarship to Study Abroad from Taiwan during 2020-2022, and was awarded the Sigma Xi Award in 2021 for excellence in research. His preprints and publications include [49, 55, 56, 58].

Acknowledgments

I would like to express my deep gratitude to my advisor Professor Kavita Ramanan. Throughout my graduate study, she has been sharing her passions in mathematics research and providing valuable guidance to become a mathematician.

I would also like to thank Professor Paul Dupuis and Professor Grigoris Paouris for serving on my committer. They have been sharing many insights in probability theory and convex geometry.

I am grateful for Academia Sinica and NCTS, where they provided me great resources during the time of Covid. In particular, I would like to thank Professor Chii-Ruei Hwang and Ju-Yi Yen for the encouragement and guidance, and also Te-Chun and I-Hsun for many helpful discussion. I am also supported by the National Science Foundation under grant DMS-1954351.

In addition, I would like to thank my roommates Vito and Charlie, for all the cooking, math, movie and gym sessions, Justin, Emily and Tingwei, for the fun times in the office, everyone in the research group of Professor Ramanan, for the fruitful discussions, friends in the Taiwanese association, all other friends in Division of Applied Math, and also Nick for his consistent support during my hardest time in finishing the thesis.

Finally, I thank my family for their support over the years.

Abstract of "Sharp large deviation estimates and their applications to asymptotic convex geometry", by Yin-Ting Liao, Ph.D., Brown University, October 2022

The study of high-dimensional measures is an integral part of probability, statistics and asymptotic convex geometry, where the object of interest is the normalized volume measures of a convex body. A common theme is to try to analyze these high-dimensional objects by looking at their lower-dimension projections. Classical theorems in probability theory on independent and identically distributed random variables, such as the law of large numbers and central limit theorem, provide information about scalar projections of high-dimensional product measures along a particular direction (i.e., the direction $(1, \ldots, 1)$). Although the study of (typical) projections of more general (non-product) high-dimensional measures dates back to Borel, only recently has a theory begun to emerge in the field of asymptotic convex geometry, which in particular identifies the role of certain geometric assumptions that lead to better behaved projections. In this thesis, we study large deviation properties for random projections of high-dimensional measures and their applications to asymptotic convex geometry. The thesis is divided into three parts as follows.

In the first part of the thesis, we study large deviation principles for random multidimensional projections, which are averaged over the randomness of the projections (the so-called annealed setting). Antilla et al. showed that if a random vector $X^n \in \mathbb{R}^n$ satisfies a *thin-shell* condition, which is a quantitative estimate on how the mass is concentrated around a thin sphere, then for most directions θ^n, fluctuations of most lower-dimension projections $\langle X^n, \theta^n \rangle$ are almost Gaussian. We introduce an *asymptotic thin-shell condition*, which describes the asymptotic rate of concentration of the mass around a thin sphere as the dimension grows to infinity, and show that this condition is sufficient for annealed random projections of the high-dimensional vector to satisfy a large deviation principle. We also verify the asymptotic thin-shell condition for sev-

eral classes of high-dimensional measures including normalized volume measures of a large class of Orlicz balls.

In the second part of the thesis, we obtain refined sharp large deviation estimates for annealed random projections under an additional asymptotic density condition and apply these large deviation estimates to study problems of interest in asymptotic convex geometry. The study of intersections of convex bodies was originated by a classical question raised by Milman, "What is the volume left in a unit ℓ_p^n ball after removing a t-multiple of ℓ_q^n ball?" To answer this question, Schechtman and Schmuckenschläger studied the limit of the ratio of $\left|B_p^n(1) \cap B_q^n(t)\right|$ and $\left|B_p^n(1)\right|$, where $B_p^n(t)$ is an ℓ_p^n ball with radius t. They showed that an interesting phase transition occurs in t: the ratio converges to either 0 or 1 in a certain *subcritical* and *supercritical* regime, respectively. The critical regime is more subtle and was finally resolved after a decade by Schechtman, who showed the ratio converges to $1/2$ as $n \to \infty$. The corresponding results for other classes of convex bodies were not as well understood. We extend these results to Orlicz balls and obtain again a phase transition by deriving a sharp estimate for the volume of the intersection of two Orlicz balls. Various other volumetric properties for Orlicz balls are also obtained.

In the last part of the thesis, we study quenched (i.e., conditioned on the random projection bases) large deviation estimates for random multidimensional projections, and investigate what geometric information they capture about the high-dimensional measures. LDPs for random projections of ℓ_p^n balls were obtained in Gantert et al. While they captured some geometric information, they could not distinguish between balls and spheres. Sharp large deviation estimates were originated obtained for sums of independent and identically distributed random variables $\{X_i\}_{i\in\mathbb{N}}$ by Bahadur and Ranga Rao. With the perspective that such a sum can be viewed as the projection of a product measure in the $(1, ..., 1)$ direction, we generalize these results in two ways. First, we look at projections along a random fixed direction $\theta^n \in \mathbb{S}^{n-1}$ sampled uniformly from the unique rotational invariant measure on $\mathbb{S}^{n-1}$. Next, we consider random vectors $X^n \in \mathbb{R}^n$ distributed according to some non-product measures. The main idea behind the derivation of the corresponding sharp large deviation estimates is to identify the correct change of measure for the rare event corresponding to the deviation and to obtain a quantitative central limit theorem under this change of measure. We also propose an importance sampling algorithm to numerically approximate the tail probability or the volume of spherical caps of ℓ_p^n balls. Numerical simulations show that the sharp large deviation estimate is a very good approximation even when the dimension of the vector is not particularly high.

Contents

Chapter One

Introduction

With the emergence of big data, there is an urgent need to develop new ideas and techniques for handling enormous amounts of data. As the dimension (i.e., number of features) of a dataset increases, the computational cost of analyzing the data grows exponentially, a phenomenon known as the curse of dimensionality. One way to overcome the curse of dimensionality is to use dimension reduction by capturing the main features of a high-dimensional dataset using a lower-dimensional space. A dimension reduction technique that is commonly used when there is no a priori knowledge of the dataset is Projection Pursuit [22, 37], which projects the data onto random directions and only chooses those in which the projected data is not close to Gaussian. However, such projection directions are scarce due to the universality result that most random projections of (a large class of) high-dimensional measures look Gaussian [51]. When there is an abundance of data, a potentially valuable alternative is to look at rare events or the tails of random projections in order to capture information about or to distinguish among different high-dimensional data or more general measures.

The study of high-dimensional probability distributions through their lower-dimensional projections, especially random projections, is a common theme in a wide range of areas, including geometric functional analysis [39], statistics and data analysis [22, 31], information retrieval [38, 69], machine learning [60] and asymptotic geometric analysis [8, 51]. In the latter case, the study of lower-dimensional projections is also interesting from a theoretical perspective as it relates to fundamental questions in asymptotic convex geometry. Dvoretzky's theorem states that every high-dimensional normed vector space has a lower-dimensional subspace that is "close" to a Euclidean space. Finite-dimensional Banach spaces are characterized by the level sets of their norm, which are (symmetric) convex sets. Therefore, Dvoretzky's theorem can be viewed as saying something about high-dimensional convex sets – specifically, that most of their (sufficiently) lower-dimensional sections are elliptic. From a probabilistic point of view, it is natural to also consider the normalized volume measure of a convex set, or more general high-dimensional measures, and consider their projections. A significant result is the central limit theorem for convex sets [51], which says that if a random variable X^n is distributed uniformly on an isotropic (i.e., having mean 0 and covariance matrix proportional to the identity matrix) convex body, then its one-dimensional scalar projections $\langle X^n, \theta^n \rangle$ along most directions θ^n on the unit $(n-1)$-dimensional sphere $\mathbb{S}^{n-1}$ in $\mathbb{R}^n$ is almost Gaussian.

In order to set this result in context, first consider a sequence of independent and identically distributed (i.i.d.) random variables $\{X_i\}_{i\in\mathbb{N}} \subset \mathbb{R}$ with mean 0 and variance 1, the classical central limit theorem says as $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \to Z \quad \text{weakly,}$$

where Z is a standard normal random variable. Beyond the central limit theorem fluctuation regime, large deviation principles capture the rate of decay of tail probabilities. A fundamental result in large deviation theory is Cramér's theorem, which provides an exponential rate of decay of the probability of a rare event. To be more precise, let $S^n := \frac{1}{n} \sum_{i=1}^{n} X_i$. Define $\Lambda(\lambda) = \mathbb{E}[e^{\lambda X_1}]$ and let $\mathbb{I}(a)$ be the Legendre transform of Λ, defined to be $\mathbb{I}(a) := \sup_{\lambda\in\mathbb{R}} \{a\lambda - \Lambda(\lambda)\}$. Then under suitable condition on the (marginal) distribution of X_1, for $a > \mathbb{E}[X_1]$, we have that as $n \to \infty$

$$\frac{1}{n} \log \mathbb{P}(S^n > a) \to -\mathbb{I}(a).$$

Both the central limit theorem and Cramér's theorem can be viewed in a geometric point of view. The empirical mean can be written as the inner product,

$$S^n = \frac{1}{n}\sum_{i=1}^{n} X_i^n = \frac{1}{\sqrt{n}}\left\langle X^n,\, \sqrt{n}\mathfrak{I}^n\right\rangle, \tag{1.1}$$

where $X^n := (X_1, \ldots, X_n)$ and $\mathfrak{I}^n := \frac{1}{\sqrt{n}}(1, 1, \cdots, 1) \in \mathbb{S}^{n-1}$. Hence, a natural question is to understand projections along directions, possibly random, other than the typical direction $\mathfrak{I}$. Moreover, one can go beyond product measures and consider more general measures.

As mentioned earlier, a significant result in this direction is the so-called central limit theorem (CLT) for convex sets, which roughly says that most k-dimensional projections (equivalently, marginals) of an n-dimensional isotropic convex body are close to Gaussian in the total variation distance, when n is sufficiently large and k is of a smaller order than n^α for some universal constant $\alpha \in (0, 1)$. Although foreshadowed by results of Sudakov [77], Diaconis and Freedman [22] and von Weizsäcker [80], the conjecture that most marginals are Gaussian was precisely formulated by Anttila, Ball and Perissinaki in [8] (see also [18]). Specifically, they showed that if the Euclidean norm of a symmetric high-dimensional random vector $X^{(n)}$ satisfies a certain concentration estimate referred to as the "thin shell" condition, then "most" of its marginals are approximately Gaussian.

Definition 1.0.1. Given $\varepsilon > 0$, a random vector $X \in \mathbb{R}^n$ is said to satisfy the ε thin shell condition if for some $m \in (0, \infty)$,

$$\mathbb{P}\left(\left|\frac{\|X\|}{\sqrt{n}} - m\right| < \varepsilon\right) \leq \varepsilon.$$

They also verified this condition for random vectors uniformly distributed on a certain class of convex sets whose modulus of convexity and diameter satisfy certain assumptions. Subsequently, the thin shell condition was verified for various classes of convex bodies by several authors, with a breakthrough verification due to Klartag [51] for any isotropic log-concave distribution, which, in particular, includes the uniform distribution on an isotropic convex body (see also [32] for a simplified proof). The result of [8] was further extended to general high-dimensional measures by Meckes [61], who showed that whenever an n-dimensional random vector satisfies a quantitative version of the thin shell condition, then most k-dimensional marginals are close to Gaussian (in the bounded-Lipschitz distance) if $k < 2\log n/\log\log n$, and that the latter cutoff for k is in some sense the best possible.

While the central limit theorem for convex sets is a beautiful universality result, it implies that most lower-dimensional projections do not provide much information about the high-dimensional measure. In contrast, large deviation principles (LDPs) are typically non-universal and distribution-dependent, and only recently has work begun to emerge in the study of LDPs for high-dimensional convex bodies. Since we are considering random projections of random vectors, the randomness come from two different sources. We consider two different scenarios, quenched (conditioned on the random projection bases) and annealed (averaging over all the directions) random projections. A first result on random projections of ℓ_p^n balls for $p \in (1, \infty]$ was studied in [34], where they obtained the large deviations for both annealed and quenched random projections. They also showed in [33] the atypicality of Cramér's theorem which considered quenched random projections of product measures. (See also Section 2.1.2 for discussion on other related works.)

In Chapter 2, we consider the large deviation principle for annealed random projections. In the thesis [48], Kim and Ramanan proposed the asymptotic thin shell condition which extends the previous results [34] to more general measures beyond ℓ_p^n balls and product measures, and more general projections, beyond one-dimensional projections A sequence of random vectors is said to satisfy the asymptotic thin shell condition if the sequence of scaled norms satisfies an LDP at speed n. However, for $p \in (1,2)$, ℓ_p^n balls do not satisfy the asymptotic thin shell condition, where the sequence of scaled norms satisfies an LDP at speed $n^{p/2}$, [43]. We extend the asymptotic thin shell condition in [48] to include these cases. Under this extension, we are faced with the challenge to establish LDPs of products or more general functions of multiple random vectors where different random vectors satisfy LDPs at different speeds, and deal with the subtleties involved to resolve these competing speeds to determine the speed of the LDP of the function of interest. We obtain the large deviation principles under the asymptotic thin shell condition in three regimes of dimension k_n of random projections – constant $k_n \equiv k$, sublinear $1 \ll k_n \ll n$ and linear $k_n \sim n$. We provide examples satisfying the asymptotic thin shell condition, including ℓ_p^n balls for $p \in (1,\infty]$, superquadratic Orlicz balls and a class of Gibbs measure.

While the LDP only provides asymptotic log equivalents of the probability (i.e., only exponential decay), in Chapter 3, we obtain refined large deviation estimates for annealed random projections which include a correction "or prefactor" that refines the exponential decay one obtains from the LDP. Such refined large deviation estimates for product measures were first given in Bahadur and Ranga Rao [10], 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 $S^n = \frac{1}{n}X_i$. Under suitable assumptions on the (marginal) distribution of X_1 it was shown in [10] that

$$\mathbb{P}\left(S^n \geq a\right) = \frac{e^{-n\mathbb{I}(a)}}{\sigma_a\tau_a\sqrt{2\pi n}}\left(1 + o(1)\right). \tag{1.2}$$

where $\mathbb{I}$ is the Legendre transform of Λ, the logarithmic moment generating function of X_1, and $\tau_a > 0$ and $\sigma_a > 0$ are suitable constants specified below. Key ingredients of the proof in [10] include first identifying a "tilted" measure under which the rare event on the left-hand side of (3.1) 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 $\otimes^n \widetilde{\mathbb{P}}_a$ where $\widetilde{\mathbb{P}}_a$ is given by

$$\frac{d\widetilde{\mathbb{P}}_a}{d\mathbb{P}} := e^{\tau_a X_1 - \Lambda(\tau_a)},$$

with τ_a being the unique positive constant such that under $\widetilde{\mathbb{P}}_a$, X_1 has mean a. The constant σ_a^2 in (3.1) is the variance of X_1 under $\widetilde{\mathbb{P}}_a$. 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 $\widetilde{\mathbb{P}}_a$. The techniques of sharp large deviation turn out to be useful in asymptotic convex geometry. The difficulty here involves considering non-independent summands and identifying correct tilted measures. We also provide a generalization of a one-dimensional Laplace approximation to resolve asymptotic integral with different speeds in the exponent. We apply the technique to obtain various quantities of interests including sharp volume estimates of Orlicz balls and volume ratio of intersections of Orlicz balls.

In Chapter 4, we turn to quenched random projections. As in Chapter 3, we consider the sharp estimates for ℓ_p^n balls. We also include the simpler case of product measures to illustrate the main idea. Unlike for annealed random projection, we lose the symmetry in the quenched case, that is, when conditioning on a direction sequence and need to take care of additional technicalities. The obtained sharp large deviation estimates provide additional geometric information that was not observed in the large deviation principles. Along the way, we also provide an importance sampling algorithm to approximate the volume of ℓ_p^n spherical caps and establish a central limit theorem for functionals direction sequences which might be of independent interest.

1.1 Preliminary and background results on large deviation principles

We set some initial notation and definitions, with a particular emphasis on large deviations terminology. First, for $a, b \in \mathbb{R}$, we will use $a \vee b$ and $a \wedge b$ to denote $\max(a, b)$ and $\min(a, b)$, respectively. Next, for $p \in [1, \infty]$, let $\|\cdot\|_p$ denote the ℓ_p^n norm (with some abuse of notation, we use common notation for the ℓ_p^n norm on $\mathbb{R}^n$ for any $n \in \mathbb{N}$). We use the notation $\mathcal{N}(\mu, \sigma^2)$ to denote a normal random variable with mean μ and variance σ^2.

Given a topological space $\mathcal{X}$ with Borel σ-algebra $\mathcal{B}$, let $\mathcal{P}(\mathcal{X})$ denote the space of probability measures on $\mathcal{X}$. By default, we impose the topology of weak convergence on $\mathcal{P}(\mathcal{X})$: recall that a sequence $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(\mathcal{X})$ is said to converge weakly to $\mu \in \mathcal{P}(\mathcal{X})$, also denoted as $\mu \Rightarrow \mu$, if and only if for every bounded continuous function f on $\mathcal{X}$, $\int f d\mu_n \to \int f d\mu$ as $n \to \infty$. On occasion, when $\mathcal{X} = \mathbb{R}^d$, we will consider subsets of probability measures that have certain finite moments. For $q \geq 1$ and $d \in \mathbb{N}$, define

$$\mathcal{P}_q(\mathbb{R}^d) := \left\{ \nu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |x|^q \nu(dx) < \infty \right\}.$$

Then, a sequence of probability measure $\{\nu_n\}_{n\in\mathbb{N}} \subset \mathcal{P}_q(\mathbb{R}^d)$ converges to $\nu \in \mathcal{P}_q(\mathbb{R}^d)$ with respect to the *q-Wasserstein topology* if we have weak convergence $\nu_n \Rightarrow \nu$ and convergence of q-th moments $\int_{\mathbb{R}^d} |x|^q \nu_n(dx) \to \int_{\mathbb{R}^d} |x|^q \nu(dx)$. In fact, as elaborated in

[79, Sec. 6], the q-Wasserstein topology can be metrized by a distance function called the q-Wasserstein metric, which we denote $\mathcal{W}_q$. Next, for $q > 0$, let

$$M_q(\nu) := \int_{\mathbb{R}^d} |x|^q \nu(dx), \quad \nu \in \mathcal{P}(\mathbb{R}^d), \tag{1.3}$$

denote the q-th moment map. In our analysis, we will frequently consider the following subset: for $j \in \mathbb{N}$, define $K_{2,j} \subset \mathcal{P}(\mathbb{R}^d)$ as

$$K_{2,j} := \left\{ \nu \in \mathcal{P}(\mathbb{R}^d) : M_2(\nu) \leq j \right\}. \tag{1.4}$$

Lemma 1.1.1. *Fix $j \in \mathbb{N}$. For any $q \in [1, 2)$, the set $K_{2,j} \subset \mathcal{P}_2(\mathbb{R}^d)$ is compact with respect to the q-Wasserstein topology. In addition, $K_{2,j}$ is convex and non-empty.*

Proof. The proof is a simple modification of the proof of the $j = 1$ case given in [50, Lemma 3]. □

We refer to [21] for general background on large deviations theory. In particular, we recall the definition:

Definition 1.1.2. Let $\mathcal{X}$ be a regular topological space with Borel σ-algebra $\mathcal{B}$. A sequence $\{P_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(\mathcal{X})$ is said to satisfy a *large deviation principle (LDP)* at *speed* s_n with rate function $I : \mathcal{X} \to [0, \infty]$ if I is lower-semicontinuous and for all $\Gamma \in \mathcal{B}$,

$$-\inf_{x\in\Gamma^\circ} I(x) \leq \liminf_{n\to\infty} \tfrac{1}{s_n} \log P_n(\Gamma) \leq \limsup_{n\to\infty} \tfrac{1}{s_n} \log P_n(\Gamma) \leq -\inf_{x\in\bar{\Gamma}} I(x),$$

where Γ° and $\bar{\Gamma}$ are the interior and closure of Γ, respectively. We say I is a *good rate function (GRF)* if, in addition, it has compact level sets. Analogously, a sequence of $\mathcal{X}$-valued random variables $\{\eta_n\}_{n\in\mathbb{N}}$ is said to satisfy an LDP at speed s_n and with a rate function I if the sequence of their laws $\{\mathbb{P} \circ \eta_n\}_{n\in\mathbb{N}}$ satisfies an LDP at the same speed and with the same rate function.

Remark 1.1.3. Let $\{P_n\}_{n\in\mathbb{N}}$ be a sequence of probability measures satisfying an LDP at speed s_n with GRF I that has a unique minimizer m. It is immediate from the definition of the LDP that $I(m) = 0$ and that then, for any speed t_n such that $t_n/s_n \to 0$, the sequence $\{P_n\}_{n\in\mathbb{N}}$ also satisfies an LDP at speed t_n, but with a degenerate rate function χ_m, which is defined to be 0 at m and $+\infty$ elsewhere.

Remark 1.1.4. Let $\{P_n\}_{n\in\mathbb{N}}$ be a sequence of probability measures satisfying an LDP at speed s_n with GRF I and suppose b_n is another sequence such that $s_n/b_n \to \lambda \in (0, \infty)$. Then it is immediate from the definition that $\{P_n\}_{n\in\mathbb{N}}$ also satisfies an LDP at speed b_n with GRF λI.

As a useful tool, we recall the following definition.

Definition 1.1.5. Let $\mathcal{X}$ be a metric space with distance d, equipped with its Borel σ-algebra. Two sequences of $\mathcal{X}$-valued random variables $\{\eta_n\}_{n\in\mathbb{N}}$ and $\{\tilde{\eta}_n\}_{n\in\mathbb{N}}$ are *exponentially equivalent* at speed s_n if for all $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{s_n} \log \mathbb{P}(d(\eta_n, \tilde{\eta}_n) > \delta) = -\infty. \tag{1.5}$$

Remark 1.1.6. The notion of exponential equivalence is valuable because if an LDP holds for $\{\eta_n\}_{n\in\mathbb{N}}$, then an LDP holds for an exponentially equivalent sequence $\{\tilde{\eta}_n\}_{n\in\mathbb{N}}$ with the same GRF (see, e.g., Theorem 4.2.13 of [21]).

Remark 1.1.7. Let $\{a_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ that converges to $a \in \mathbb{R}$ as $n \to \infty$. Suppose $\{U_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ at speed b_n with a GRF I. Then the sequences $\{a_n U_n\}_{n\in\mathbb{N}}$ and $\{aU_n\}_{n\in\mathbb{N}}$ are exponentially equivalent at speed b_n. Indeed, for any $M \in (0, \infty)$, since $\{U_n\}_{n\in\mathbb{N}}$ satisfies an LDP with a GRF, there exists $K \in (0, \infty)$ such that

$\limsup_{n\to\infty} \frac{1}{b_n} \log \mathbb{P}(|U_n| \geq K) < -M$. Given $\delta, M \in (0,\infty)$, pick $\varepsilon > 0$ such that $\delta/\varepsilon > K$. Then

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{b_n} \log \mathbb{P}\left(|a_n U_n - aU_n| > \delta\right) &\leq \lim_{n\to\infty} \frac{1}{b_n} \log \left(\mathbb{P}\left(|a_n - a| > \varepsilon\right) + \mathbb{P}\left(|U_n| > \delta/\varepsilon\right)\right) \\ &\leq \lim_{n\to\infty} \frac{1}{b_n} \log \mathbb{P}\left(|U_n| > K\right) \\ &< -M.\end{aligned}$$

Since M is arbitrary, we obtain the desired exponential equivalence by sending $M \to \infty$.

For some of our LDPs, the resulting rate functions will be expressed in terms of the following quantities. For $\nu \in \mathcal{P}(\mathbb{R})$, define the *entropy* of ν as

$$h(\nu) := -\int_{\mathbb{R}} \log\left(\tfrac{d\nu}{dx}\right) d\nu, \tag{1.6}$$

for ν with density (with respect to Lebesgue measure dx), and $h(\nu) := -\infty$ otherwise. Furthermore, for $\nu, \mu \in \mathcal{P}(\mathbb{R})$, define the *relative entropy* of ν with respect to μ as

$$H(\nu|\mu) := \int_{\mathbb{R}} \log\left(\tfrac{d\nu}{d\mu}\right) d\nu \tag{1.7}$$

if ν is absolutely continuous with respect to μ, and $H(\nu|\mu) := +\infty$ otherwise. Given a function $\Lambda : \mathbb{R} \to \mathbb{R}$, let Λ^* be its Legendre-Fenchel transform defined by

$$\Lambda^*(x) = \sup_{t\in\mathbb{R}}\{xt - \Lambda(t)\}, \quad x \in \mathbb{R}. \tag{1.8}$$

The following contraction principle for LDPs is used multiple times throughout the paper.

Lemma 1.1.8 (Contraction principle)**.** [21, Theorem 4.2.1] *Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces* *and $f : \mathcal{X} \to \mathcal{Y}$ a continuous function. Suppose $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of $\mathcal{X}$-valued random variables and satisfies an LDP in $\mathcal{X}$ at speed s_n with GRF $\mathcal{I}_X$. Then $\{f(X_n)\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{Y}$ at speed s_n with GRF $\mathcal{I}_Y$ defined by*

$$\mathcal{I}_Y(y) := \inf\{\mathcal{I}_X(x) : x \in \mathcal{X}, f(x) = y\}.$$

Finally, we state a simple lemma that is used multiple times in our proofs.

Lemma 1.1.9. *Suppose $\{U_n\}_{n\in\mathbb{N}}$, $\{V_n\}_{n\in\mathbb{N}}$ and $\{W_n\}_{n\in\mathbb{N}}$ satisfy LDPs in $\mathbb{R}$ at speed α_n, β_n and γ_n with GRFs J_U, J_V and J_W, respectively. Let $\alpha_n = \beta_n \ll \gamma_n$, (i.e., $\beta_n/\gamma_n \to 0$ as $n \to \infty$). Assume $\{U_n\}_{n\in\mathbb{N}}$ is independent of $\{V_n\}_{n\in\mathbb{N}}$ and J_W has a unique minimizer m. Then $\{(U_n, V_n, W_n)\}_{n\in\mathbb{N}}$ is exponentially equivalent to $\{(U_n, V_n, m)\}_{n\in\mathbb{N}}$ and satisfies an LDP at speed α_n with GRF $J : \mathbb{R}^3 \to [0,\infty]$ defined by*

$$J(u, v, w) := \begin{cases} J_U(u) + J_V(v), & w = m, \\ +\infty, & \text{otherwise}. \end{cases} \tag{1.9}$$

Moreover, if $m \neq 0$, then $\{V_n W_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed α_n with GRF $v \mapsto J_V(v/m)$.

Proof. Define $Y_n := (U_n, V_n, W_n)$ and $\tilde{Y}_n := (U_n, V_n, m)$. By the independence of $\{U_n\}_{n\in\mathbb{N}}$ and $\{V_n\}_{n\in\mathbb{N}}$, $\{\tilde{Y}_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed α_n with GRF J defined in (1.9) (this is easily deduced from the definition of the LDP, but the fact that the rate function for (U_n, V_n) is $J_U(u) + J_V(v)$ follows from [21, Exercise 4.2.7] with $\mathcal{X} = \mathbb{R}$, $\mathcal{Y} = \mathbb{R}^2$ and F being the identity mapping; the stated rate function for (U_n, V_n, m) follows as an immediate consequence). Now, for $\epsilon > 0$, we have

$$\lim_{n\to\infty} \frac{1}{\alpha_n} \log \mathbb{P}\left(\left\|Y_n - \tilde{Y}_n\right\|_2 > \epsilon\right) = \lim_{n\to\infty} \frac{1}{\alpha_n} \log \mathbb{P}(|W_n - m| > \epsilon) = -\infty,$$

where the last equality follows because J_W has a unique minimizer at m and $\{W_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $\gamma_n \gg \alpha_n$, as in the observation of Remark 1.1.3. Thus, $\{Y_n\}_{n\in\mathbb{N}}$ is exponentially equivalent to $\{\tilde{Y}_n\}_{n\in\mathbb{N}}$ at speed α_n. By Remark 1.1.6, this implies $\{Y_n\}_{n\in\mathbb{N}}$ satisfies an LDP with speed α_n with the same GRF as $\{\tilde{Y}_n\}_{n\in\mathbb{N}}$. This proves the first assertion of the lemma.

The second assertion follows by applying the contraction principle to the mapping $F : \mathbb{R}^3 \to \mathbb{R}$ defined by $F(u, v, w) = vw$. $\square$

Chapter Two

Annealed large deviation estimates for random projections

2.1 Introduction

In this chapter, we introduce a large deviation analogue of the results in [8]. Whereas the latter work shows that fluctuations of (most) random projections of high-dimensional vectors that satisfy a *thin shell condition* can be characterized (as almost Gaussian), our work characterizes tail behavior (at the level of annealed LDPs) for projections and their associated norms onto (possibly growing) random subspaces of high-dimensional random vectors that satisfy an *asymptotic thin shell condition*. In particular, our work goes beyond the more studied specific setting of measures on ℓ_p^n balls or spheres, and also univariate LDPs, although we also obtain new results in this setting (see Remark 2.6.5). Specifically, for any sequence of random vectors $\{X^{(n)}\}_{n\in\mathbb{N}}$ whose scaled Euclidean norms satisfy an LDP (see Assumption 2.1.3 and its specific case Assumption 2.1.3*), we characterize the tail behavior of the corresponding sequence of orthogonal projections of $X^{(n)}$ onto a random k_n-dimensional basis, $k_n \leq n$, drawn from the Haar measure on the Stiefel manifold $\mathbb{V}_{n,k}$ of orthonormal k_n-frames in $\mathbb{R}^n$, as the dimension n goes to infinity with $k_n/n \to \lambda \in [0,1]$. Assumption 2.1.3 (or rather, a slight strengthening of it) can be viewed as an "asymptotic thin shell" condition since it implies that for all sufficiently large n, the random vector $X^{(n)}$ satisfies the thin shell condition (see the discussion at the end of Section 2.1.1). Note, however, that for growing subspaces, in contrast to CLT results where approximate Gaussian marginals holds only for $k_n < \frac{2\log n}{\log\log n}$ [61] (or $k_n \sim n^\alpha$ [51] if one assumes additional regularity of $X^{(n)}$ such as logconcavity), the annealed LDP results indicate three crucial regimes for $\{k_n\}_{n\in\mathbb{N}}$, constant, sublinear and linear (see also [3] for ℓ_p^n balls).

A summary of our main results is as follows (the precise definition of LDPs, rate functions, and the Stiefel manifold are given in Section 1.1 and Section 2.1.1):

1. *LDPs in the constant regime (Theorem 2.3.3):* Given Assumption 2.1.3*, or a modification of it stated as Assumption 2.3.1, in the setting where $k_n = k$ for every n, we establish an (annealed) LDP for the sequence of k-dimensional random projections of $X^{(n)}$.

2. *LDPs in the sublinear and linear regimes (Theorems 2.4.1 and 2.5.1):* Given Assumption 2.1.3, in the setting where $\{k_n\}_{n\in\mathbb{N}}$ satisfies $k_n \to \infty$, we establish (annealed) LDPs for the sequence of empirical measures of the coordinates of the k_n-dimensional projections of $X^{(n)}$. This LDP is established with respect to the q-Wasserstein topology for $q < 2$, which is stronger than the weak topology. The rate function is shown to have a different form in the sublinear ($k_n/n \to 0$) and linear ($k_n/n \to \lambda \in (0,1]$) regimes.

3. *LDPs for norms of random projections (Corollary 2.3.4 and Theorems 2.4.5, 2.4.4 and 2.5.2):* We establish LDPs for sequences of ℓ_q norms of the multi-dimensional random projections in all regimes, with two different scalings considered in the sublinear regime.

4. *Illustrative examples (Section 2.6):* To show that our theory unites disparate examples under a common framework, recovering, and in some cases extending, existing results for ℓ_p^n balls, while also covering new examples, we verify our assumptions for many sequences $\{X^{(n)}\}_{n\in\mathbb{N}}$ of interest, including product measures, the uniform measure on certain scaled ℓ_p^n balls and generalized Orlicz balls, and Gibbs measures (see Remark 2.1.4). Along the way, we obtain several results of potentially independent interest. Specifically, in Theorem 2.6.4 we show that when $p \in [1,2)$, the LDP for Euclidean norms of multi-dimensional projections of ℓ_p^n balls exhibits an interesting phase transition depending on whether the ratio $k_n/n^{2p/(2+p)}$ is asymptotically finite or infinite, thus disproving a conjecture in [3] (see Remark 2.6.5 for more details). In addition, we also obtain the asymptotic

logarithmic volume of Orlicz balls, as elaborated in Remark 2.6.11. (See also recent extensions due to [42] and [56]).

The formulation of the correct form of the asymptotic thin shell condition that would also allow one to consider ℓ_p^n balls, when $p \in [1,2)$, is somewhat subtle, and involves a suitable rescaling argument. Verification of the asymptotic thin shell condition for ℓ_p^n balls, for all $p \geq 1$, makes use of a probabilistic representation for the uniform measure on ℓ_p^n balls (see [71, 75] and Section 2.6.1 for details). However, such probabilistic representations are not available for Orlicz balls, and so we develop a new approach in which the tail probability is expressed as a volume ratio (see (3.2.14) in the proof of Proposition 2.6.7), and the asymptotics of this volume ratio is determined using a tilted measure with respect to Lebesgue measure. Subsequent work that uses similar ideas to obtain the volumetric properties and the thin shell concentration for random vectors in Orlicz balls includes [2, 42]. More recently, detailed estimates of intersections and differences of Orlicz balls have been obtained in [58]. We leave for future work the identification of more sequences $\{X^{(n)}\}_{n\in\mathbb{N}}$ of random vectors that satisfy the asymptotic thin shell (and related) conditions. Furthermore, although here we focused on Euclidean (and more general ℓ_q) norms of the random projections, since they are of special relevance in asymptotic convex geometry, our results could potentially be used to also investigate other symmetric functionals of lower-dimensional projections that are of interest, such as the volume or barycenter of the projected body, or other functionals of interest in high-dimensional statistics.

2.1.1 Projection Regimes and Assumptions

For each $n \in \mathbb{N}$, consider a random vector $X^{(n)}$ that takes values in $\mathbb{R}^n$. For $k \in \mathbb{N}$, let I_k denote the $k\times k$ identity matrix, and for $n > k$, let $\mathbb{V}_{n,k} = \{A \in \mathbb{R}^{n\times k} : A^T A = I_k\}$ denote the Stiefel manifold of orthornomal k-frames in $\mathbb{R}^n$. Here, the constraint $A^T A = I_k$ ensures that the columns of A are orthonormal. As is well known, $\mathbb{V}_{n,k}$ is a compact subset of $\mathbb{R}^{n\times k}$, being the pre-image of the closed set $\{I_k\}$ under the continuous map $A \mapsto A^T A$ that is also bounded (by $\sqrt{k}$ with respect to the Frobenius norm). We are interested in the orthogonal projection of $X^{(n)}$ onto a random k_n-dimensional subspace, where $1 \leq k_n < n$. To this end, fix $n \in \mathbb{N}$, $1 \leq k_n \leq n$, and let

$$\mathbf{A}_{n,k_n} = [\mathbf{A}_{n,k_n}(i, j)]_{i=1,\ldots,n;\, j=1,\ldots,k_n}$$

be an $n \times k_n$ random matrix drawn from the Haar measure on the Stiefel manifold $\mathbb{V}_{n,k_n}$ (i.e., the unique probability measure on $\mathbb{V}_{n,k_n}$ that is invariant under orthogonal transformations). Note that the random matrix $\mathbf{A}^T_{n,k_n}$ linearly projects a vector from n to k_n dimensions. We assume that for each $n \in \mathbb{N}$, $\mathbf{A}_{n,k_n}$ is independent of $X^{(n)}$, and for simplicity, we assume that the sequences $\{X^{(n)}\}_{n\in\mathbb{N}}$ and $\{\mathbf{A}_{n,k_n}\}_{n\in\mathbb{N}}$ are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, although dependencies across n are immaterial for the questions we address.

We aim to analyze the large deviation behavior of the coordinates of random projections $\mathbf{A}^T_{n,k_n} X^{(n)}$ of $X^{(n)}$ in three regimes, constant, linear and sublinear, depending on how the dimension k_n of the projected vector changes with n:

Definition 2.1.1. Given a sequence $\{k_n\}_{n\in\mathbb{N}} \subset \mathbb{N}$, we say:

1. $\{k_n\}_{n\in\mathbb{N}}$ is *constant* at k, for some $k \in \mathbb{N}$, denoted $k_n \equiv k$, if $k_n = k$ for all $n \in \mathbb{N}$;

2. $\{k_n\}_{n\in\mathbb{N}}$ *grows sublinearly*, denoted $1 \ll k_n \ll n$, if $k_n \to \infty$ but $k_n/n \to 0$;

3. $\{k_n\}_{n\in\mathbb{N}}$ *grows linearly* with rate λ, for some $\lambda \in (0,1]$, denoted $k_n \sim \lambda n$, if $k_n/n \to \lambda$.

When $\{k_n\}_{n\in\mathbb{N}}$ is constant at some $k \in \mathbb{N}$, then one can investigate when the sequence of vectors $\{\mathbf{A}^T_{n,k_n} X^{(n)} = \mathbf{A}^T_{n,k} X^{(n)}\}_{n\in\mathbb{N}}$ satisfies an LDP in the space $\mathbb{R}^k$. In contrast, when k_n increases to infinity, in order to even pose the question of existence of an LDP, one must first embed the sequence $\{\mathbf{A}^T_{n,k_n} X^{(n)}\}_{n\in\mathbb{N}}$ of random vectors of different dimensions into a common topological space. Thus, we prove an LDP for the sequence $\{L^n\}_{n\in\mathbb{N}}$ of empirical measures of the coordinates of the k_n-dimensional random vectors $\mathbf{A}^T_{n,k_n} X^{(n)}$:

$$L^n := \frac{1}{k_n} \sum_{j=1}^{k_n} \delta_{(\mathbf{A}^T_{n,k_n} X^{(n)})_j}, \quad n \in \mathbb{N}. \tag{2.1}$$

Remark 2.1.2. Note that the law of $\mathbf{A}_{n,k_n}$ is invariant under permutation of its k_n columns, so the k_n coordinates of $\mathbf{A}^T_{n,k_n} X^{(n)}$ are exchangeable. Thus, the empirical measure L^n in (2.1) encodes the essential distributional properties of the coordinates of the projection, and hence, serves as a natural choice for a common infinite-dimensional embedding of the k_n coordinates of $\mathbf{A}^T_{n,k_n} X^{(n)}$, for all $n \in \mathbb{N}$.

We now present our main condition on the sequence $\{X^{(n)}\}_{n\in\mathbb{N}}$.

Assumption 2.1.3. The sequence of scaled norms $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF $J_X : \mathbb{R} \to [0,\infty]$.

When a special case of Assumption 2.1.3 holds with speed $s_n = n$, we say that Assumption 2.1.3* holds, i.e. the sequence of scaled norms $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n with GRF $J_X : \mathbb{R} \to [0,\infty]$.

Remark 2.1.4. The special case of Assumption 2.1.3* is important because, as shown in Section 2.6, it is satisfied by several important sequences of measures, including those whose elements are taken from a large family of product measures (see Proposition 2.2.7), the uniform measure on an ℓ_p^n ball of radius $n^{1/p}$, with $p \geq 2$ (see Proposition 2.2.8) or, in fact, a more general class of measures that includes the uniform measure on an Orlicz ball defined via a superquadratic function (see Proposition 2.6.7), and a general class of Gibbs measures with superquadratic potential and interaction functions (see Proposition 2.2.11). However, Assumption 2.1.3* no longer holds when $X^{(n)}$ is uniformly chosen from an ℓ_p^n ball of radius $n^{1/p}$ with $p \in [1, 2)$. Indeed, this can be deduced from the sharp large deviation upper bounds obtained in [68] and [35]. In addition, Theorem 1.3 of [43] shows that $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP with speed $s_n = n^{p/2}$. These results motivate the more general condition stated in Assumption 2.1.3.

Remark 2.1.5. The general form of Assumption 2.1.3 is also of interest because of its connection to the Kannan-Lovász-Simonovits (KLS) conjecture formulated in [46] (also see [1] for a nice exposition), which is one of the major open problems in convex geometry. Indeed, Theorem A in [4] states that the KLS conjecture is false if there exists a sequence of isotropic and log-concave random vectors $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfying Assumption 2.1.3 either with $s_n \ll \sqrt{n}$ and nontrivial J_X or with $s_n = \sqrt{n}$ and $\inf_{t>t_0}\{\inf_{x>t} J_X(x)/t\} = 0$ for some $t_0 \in (1, \infty)$. It is shown in [4] that when $X^{(n)}$ is uniformly distributed on the ℓ_1^n ball of radius n, then $\{X^{(n)}\}$ satisfies Assumption 2.1.3 with $s_n = \sqrt{n}$, but the condition on the rate function J_X is not satisfied (which is consistent with the fact that the KLS conjecture is widely believed to be true). In view of this observation, it would be interesting to extend our verification of Assumption 2.1.3 to more general measures such as Orlicz balls (respectively, Gibbs measures) with superlinear functions (respectively, potentials), extending the corresponding results obtained in this article for superquadratic functions (respectively, potentials).

When Assumption 2.1.3 is satisfied with a GRF J_X that has a unique minimum at $m \in \mathbb{R}_+$, go then for all sufficiently large n, the random variable $X^{(n)}$ satisfies the thin shell condition of [8, Equation (1)]. To be more precise, note that the observation that the minimum of J_X, which is equal to zero, is achieved uniquely at m, together with the fact that $s_\ell \to \infty$, implies that for every $c > 0$, there exists $\{\delta_\ell\}_{\ell\in\mathbb{N}}$ with $\delta_\ell \downarrow 0$ such that

$$\sqrt{s_\ell}\inf\{J_X(x) : |x - m| \geq \delta_\ell, x \in \mathbb{R}_+\} \geq c.$$

Setting $\varepsilon_\ell := \max(\delta_\ell, 2e^{-c\sqrt{s_\ell}})$, it then follows from Assumption 2.1.3 and the definition of the LDP (see Definition 1.1.2) that $\varepsilon_\ell \downarrow 0$ as $\ell \to \infty$, and for every $\ell \in \mathbb{N}$, there exists $N_\ell \in \mathbb{N}$ such that

$$\mathbb{P}\left(\left|\frac{\left\|X^{(n)}\right\|_2}{\sqrt{n}} - m\right| \geq \varepsilon_\ell\right) \leq \varepsilon_\ell, \qquad \text{for all } n \geq N_\ell. \tag{2.2}$$

In particular, this implies the following weak limit:

$$\frac{\|X^{(n)}\|_2}{\sqrt{n}} \xrightarrow{\mathbb{P}} m \in \mathbb{R}_+.$$

Thus, we refer to the strengthening of Assumption 2.1.3 with J_X having a unique minimum as the *asymptotic thin shell condition*. Note that the thin shell condition is usually stated for isotropic random vectors $X^{(n)}$ and with $m = 1$, but since we do not restrict our consideration to only isostropic random vectors $X^{(n)}$, we phrased the condition above for arbitrary $m > 0$. Just as the thin shell condition yields a central limit theorem in the sense that the projections of $X^{(n)}$ can be shown to be close to Gaussian (see, e.g., [8, 51]), our results, summarized in the next three sections, show that the asymptotic thin shell condition implies that the empirical measures $\{L^n\}_{n\in\mathbb{N}}$ of the coordinates of the projections of $X^{(n)}$ satisfy an LDP in the sublinear and linear regimes (with the weaker Assumption 2.1.3 sufficing in the latter regime).

Remark 2.1.6. There exist sequences $\{X^{(n)}\}_{n\in\mathbb{N}}$ for which the corresponding random projections satisfy an LDP, but the thin shell condition fails to hold. For example, this can happen when Assumption 2.1.3 holds with a rate function J_X that has multiple minima, as can happen for certain Gibbs measures (see Section 2.2.1) with non-convex potentials F and G. As another example, let $X^{(n)}$ be distributed according to a mixture of two Gaussian distributions in $\mathbb{R}^n$ both with mean 0 and covariance matrices I_n and $2I_n$, respectively. In other words, the density of $X^{(n)}$ takes the form $f_{X^{(n)}} = \frac{1}{2}\phi_{I^n} + \frac{1}{2}\phi_{2I^n}$ for a positive definite matrix C, where ϕ_C denotes the Gaussian density with mean 0 and covariance matrix C. Then $\{X^{(n)}\}_{n\in\mathbb{N}}$ does not satisfy the thin shell condition since half its mass is concentrated around the thin shell of radius $\sqrt{n}$ and the other half around the thin shell of radius $\sqrt{2n}$. However, it is easy to show that the sequence $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption 2.1.3* (e.g., since the sequence of norms of each of the Gaussian distributions in the mixture satisfies an LDP trivially, the LDP for the sequence of norms of the mixtures can be deduced from results in [23]).

2.1.2 Prior works

First steps towards studying LDPs of sequences of one-dimensional projections were taken in [33], [48] and [34], where it was shown that such LDPs capture geometric information about the convex body. However, LDPs for random projections have been first largely restricted to the setting of high-dimensional product measures [33] or the uniform measure on the (suitably renormalized) unit ball or sphere in the space ℓ_p^n for some $p \in [1,\infty)$ [3, 4, 34, 43, 44, 58], and secondly, limited to univariate LDPs (i.e., in $\mathbb{R}$) involving either the projection of a high-dimensional random vector onto a random one-dimensional subspace [34] or (annealed) LDPs of the Euclidean norm of an orthogonal projection onto a k_n-dimensional subspace, with k_n possibly tending to infinity [3, 4]. Indeed, the first paper to consider LDPs for norms of projections of scaled ℓ_p^n balls onto growing subspaces was [3], with further results obtained in subsequent papers (see, e.g., [4, 43, 44]). (Strictly speaking, in [3] the authors analyze norms of random vectors uniformly distributed on scaled ℓ_p^n balls projected onto k-dimensional subspaces indexed by the Grassmannian, whereas we analyze norms of random projections of more general random n-dimensional vectors (including those uniformly distributed on scaled ℓ_p^n balls) projected onto k-dimensional orthogonal bases indexed by the Stiefel manifold. In the annealed setting considered here, the Grassmanian and Stiefel manifold perspectives can be seen to be equivalent due to the exchangeability of the coordinates of the projections, but in quenched settings, as considered in [55], it is more appropriate to consider the Stiefel manifold.) The only prior example of a multivariate LDP that we know in this context is for the particular case of projections of a random vector sampled from a scaled ℓ_p^n ball onto the first k canonical directions [12, Theorem 3.4]. Further, in all cases, the analysis for non-product measures has focused on ℓ_p^n balls, where the analysis is greatly facilitated by a convenient probabilistic representation of the uniform measure on the ℓ_p^n ball (see [75, Lemma 1] or [71]), or a slightly more general class of measures supported on the ℓ_p^n ball that admit a similar probabilistic representation (see [13] or Section 2.2 of [4]). Such a representation has also been exploited in recent work [58] that obtains refined or sharp large deviations estimates for (quenched) random projections of ℓ_p^n balls and spheres.

2.2 Results under Assumption 2.1.3*

When Assumption 2.1.3* is satisfied, various LDPs in different regimes had been obtained in [48]. We collect the results here for completeness and provide the more general results in the latter sections.

Theorem 2.2.1 (constant, $k_n \equiv k$)**.** *Suppose $\{k_n\}_{n\in\mathbb{N}}$ is constant at $k \in \mathbb{N}$, and Assumption 2.1.3* holds with GRF J_X. Then $\{n^{-1/2}\mathbf{A}_{n,k}^T X^{(n)}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^k$ at speed n, with GRF $I_{\mathbf{A}X,k} : \mathbb{R}^k \to [0,\infty]$ defined by*

$$I_{\mathbf{A}X,k}(x) := \inf_{0<c<1} \left\{ J_X\left(\frac{\|x\|_2}{c}\right) - \tfrac{1}{2}\log\left(1-c^2\right)\right\}. \tag{2.3}$$

Theorem 2.2.2 (sublinear, $1 \ll k_n \ll n$)**.** *Suppose $\{k_n\}_{n\in\mathbb{N}}$ grows sublinearly and Assumption 2.1.3* holds with GRF J_X. Also, suppose that J_X has a unique minimum at $m > 0$. Let H be the relative entropy functional defined in (1.7). Then, for every $q \in [1,2)$, $\{L^n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}_q(\mathbb{R})$ at speed k_n, with GRF $\mathbb{I}_{L,k_n} : \mathcal{P}_q(\mathbb{R}) \to [0,\infty]$, defined by*

$$\mathbb{I}_{L,k_n}(\mu) := H(\mu|\gamma_m).$$

Recall the definition of the second moment map $M_2(\cdot)$ from (1.3).

Definition 2.2.3. For $\lambda \in (0,1]$, define $\mathcal{H}_\lambda : \mathcal{P}(\mathbb{R}) \to [0,\infty]$ as

$$\mathcal{H}_\lambda(\nu) := \begin{cases} -\lambda\, h(\nu) + \frac{\lambda}{2}\log(2\pi e) + \frac{1-\lambda}{2}\log\left(\frac{1-\lambda}{1-\lambda M_2(\nu)}\right), & \text{if } M_2(\nu) \le 1/\lambda, \\ +\infty, & \text{otherwise,} \end{cases} \tag{2.4}$$

where we adopt the convention that $0\log 0 = 0$ and hence, $0\log(0/0) = 0$.

Note that if $\lambda M_2(\nu) = 1$, then $\mathcal{H}_\lambda(\nu) = \infty$ when $\lambda < 1$, but (due to our convention) $\mathcal{H}_\lambda(\nu) = -h(\nu) + \frac{1}{2}\log(2\pi e)$ when $\lambda = 1$.

Remark 2.2.4. Since h is strictly concave and M_2 is linear, from the definition in (2.4), $\mathcal{H}_\lambda$ is strictly convex. A direct verification shows that $\mathcal{H}_\lambda(\gamma_1) = 0$, and the strict convexity of $\mathcal{H}_\lambda$ shows that γ_1 is the unique minimizer of $\mathcal{H}_\lambda$.

Theorem 2.2.5 (linear, $k_n \sim \lambda n$)**.** *Fix $q \in [1,2)$. Suppose $\{k_n\}_{n\in\mathbb{N}}$ grows linearly with rate $\lambda \in (0,1]$ and Assumption 2.1.3* holds with GRF J_X. Then $\{L^n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}_q(\mathbb{R})$ at speed n with GRF $\mathbb{I}_{L,\lambda} : \mathcal{P}_q(\mathbb{R}) \to [0,\infty]$, where for $\mu \in \mathcal{P}_q(\mathbb{R})$,*

$$\mathbb{I}_{L,\lambda}(\mu) = \inf_{c > \sqrt{\lambda M_2(\mu)}} \left\{ J_X(c) - \tfrac{1-\lambda}{2} \log\left(1 - \tfrac{\lambda M_2(\mu)}{c^2}\right) + \lambda \log(c) \right\} \tag{2.5}$$

$$- \lambda h(\mu) + \tfrac{\lambda}{2}\log(2\pi e) + \tfrac{1-\lambda}{2}\log(1-\lambda), \tag{2.6}$$

where we use the convention that $0 \log 0 = 0$.

For any $q \in (0,\infty)$ and $n \in \mathbb{N}$, define

$$Y^n_{q,k} := \|\mathbf{A}^T_{n,k} X^{(n)}\|_q. \tag{2.7}$$

Moreover, set $\mathcal{M}_q$ to be the q-th absolute moment of a standard Gaussian random variable,

$$\mathcal{M}_q := \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} |x|^q \exp(-x^2/2) dx = \frac{2^{q/2}}{\sqrt{\pi}} \Gamma\left(\frac{q+1}{2}\right). \tag{2.8}$$

Theorem 2.2.6. *Suppose $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption 2.1.3*, and $k_n \sim \lambda n$ for some $\lambda \in (0,1]$. Then, for $q \in [1,2]$ the sequence $\{n^{-1/q} Y^n_{q,k_n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF $\mathbb{J}^{\mathrm{lin}}_{q,\lambda}$, where for $x \in \mathbb{R}_+$,*

$$\mathbb{J}^{\mathrm{lin}}_{q,\lambda}(x) := \inf_{\nu \in \mathcal{P}(\mathbb{R}), c \in \mathbb{R}_+} \left\{ \mathcal{H}_\lambda(\nu) + J_X(c) : \lambda M_q(\nu) = \left(\frac{x}{c}\right)^q \right\}, \tag{2.9}$$

with M_q the q-th moment map as in (1.3) and $\mathcal{M}_q$ the q-th absolute moment of a standard Gaussian random variable.

2.2.1 Examples

In the thesis [48], the author gives some examples that satisfy Assumption 2.1.3*. We will provide yet another important class of examples, Orlicz balls, in Section 2.6.2.

Product measures

Lemma 2.2.7 (i.i.d. case)**.** *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. real-valued random variables, and let $X^{(n)} := (X_1, \ldots, X_n)$. Suppose that we have*

$$\Lambda(t) := \log \mathbb{E}[e^{tX_1^2}] < \infty$$

for t in some open ball of non-zero radius about 0, and let Λ^* be the Legendre transform of Λ defined in (1.8). Then, $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption 2.1.3 with $J_X(x) := \Lambda^*(x^2)$ for $x \in \mathbb{R}_+$ and $J_X(x) := +\infty$ otherwise. Moreover, $m = \sqrt{\mathbb{E}[|X_1|^2]}$ is the unique minimizer of J_X.*

ℓ_p^n balls when $p \in [2, \infty)$

Consider random vectors uniformly distributed on scaled ℓ_p^n balls. More precisely, for $n \in \mathbb{N}$ and $p > 0$, define $X^{(n,p)}$ to be a random vector uniformly distributed on $n^{1/p}\mathbb{B}_p^n$, where $\mathbb{B}_p^n$ denotes the unit ℓ_p^n-ball:

$$\mathbb{B}_p^n := \left\{ x \in \mathbb{R}^n : \sum_{k=1}^n |x_k|^p \le 1 \right\}.$$

We see that when $p \in [2,\infty)$, the sequence $\{X^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies Assumption 2.1.3*. However, for $p \in [1,2)$, we show in Section 2.6.1.2 that the sequence $\{X^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies instead the more general Assumption 2.1.3.

Proposition 2.2.8. *For $n \in \mathbb{N}$ and $p \in [2,\infty)$, $\{X^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies Assumption 2.1.3* with $s_n = n$ and GRF $J_{X,p}$ where*

$$J_{X,2}(x) := \begin{cases} -\log x & \text{if } x \in (0,1], \\ +\infty & \text{otherwise,} \end{cases}$$

and for $p > 2$,

$$J_{X,p}(x) := \begin{cases} \inf_{y \geq x}[\frac{1}{2}\log\frac{y^2}{x^2} + F_p^*(y)] & \text{if } x \in (0,1), \\ +\infty & \text{otherwise.} \end{cases}$$

Moreover, $J_{X,p}$ has a unique minimizer at $m := m(p)$ defined in (2.20).

The last result can also be deduced from Proposition 2.6.7 on Orlicz balls in Section 2.6.2 by choosing $V(x) = |x|^p$. Proposition 2.2.8 and Theorem 2.2.1 together yield the following corollary.

Corollary 2.2.9. *Fix $p \in [2,\infty)$, let k_n be constant at $k \in \mathbb{N}$ and let $J_{X,p}$ be defined as in Proposition 2.2.8. Then $\{n^{-1/2}\mathbf{A}_{n,k}^T X^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^k$ at speed n with GRF $I_{\mathbf{A}X^{(p)},k} : \mathbb{R}^k \to [0,\infty]$ defined by*

$$I_{\mathbf{A}X^{(p)},k}(x) := \inf_{0<c<1}\left\{J_{X,p}\left(\frac{\|x\|_2}{c}\right) - \tfrac{1}{2}\log\left(1-c^2\right)\right\}, \quad x \in \mathbb{R}^k.$$

The particular case when $k_n \equiv 1$, for all range of $p \in [1,\infty)$ was studied in [34]; see Theorem 2.2 therein.

Gibbs measures

We now consider the case when the random vector $X^{(n)}$ is drawn from a *Gibbs measure* on configurations of n interacting particles. To be precise, let $F : \mathbb{R} \to (-\infty, \infty]$ be a "confining" potential, $G : \mathbb{R} \times \mathbb{R} \to (-\infty, \infty]$ an "interaction" potential, and for $n \in \mathbb{N}$, define a Hamiltonian $\mathbf{H}_n : \mathbb{R}^n \to (-\infty, \infty]$ given by

$$\mathbf{H}_n(x) := \frac{1}{n} \sum_{i=1}^{n} F(x_i) + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} G(x_i, x_j), \quad x \in \mathbb{R}^n.$$

As in [24], we assume that there exists a non-atomic, σ-finite measure ℓ on $\mathbb{R}$ which, along with the potentials F and G, satisfies the following conditions:

1. F and G are lower semicontinuous on the respective sets on which they are finite;
2. there exists $a \in [0, 1)$ and $c \in \mathbb{R}$ such that F satisfies $\int_{\mathbb{R}} e^{-(1-a)F(x)} \ell(dx) < \infty$, $\inf_x F(x) > c$, and $\inf_{(x,y)\in\mathbb{R}\times\mathbb{R}} [G(x, y) + a(F(x) + F(y))] > c$;
3. there exists a Borel measurable set $A \subset \mathbb{R}$ with $\ell(A) > 0$ such that
$$\int_{A\times A} [F(x) + F(y) + G(x, y)]\, \ell(dx)\ell(dy) < \infty;$$
4. for all $\lambda \in \mathbb{R}$, we have
$$\int_{\mathbb{R}\times\mathbb{R}} \exp\left[\lambda(x^2 + y^2) - F(x) - F(y) - G(x, y)\right] \ell(dx)\ell(dy) < \infty.$$

Under the above conditions, it is straightforward to verify that for $n \in \mathbb{N}$, $Z_n := \int_{\mathbb{R}^n} e^{-n\mathbf{H}_n(x)} \ell^{\otimes n}(dx)$ is finite and so we can define $P_n \in \mathcal{P}(\mathbb{R})$ as follows:

$$P_n(dx) := \frac{1}{Z_n} e^{-n\mathbf{H}_n(x)} \ell^{\otimes n}(dx), \quad x \in \mathbb{R}^n. \tag{2.10}$$

Further, let $Q_n \in \mathcal{P}(\mathbb{R})$ be the empirical measure of the coordinates of $X^{(n)}$, when the latter is drawn from P_n; more precisely, set $Q_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i^{(n)}}$, which is a random $\mathcal{P}(\mathbb{R})$-valued element.

Theorem 2.2.10 (Theorem 2.7 and Lemma 2.6 of [24])**.** *Under the conditions stated above, $\{Q_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}_2(\mathbb{R})$ equipped with the 2-Wasserstein topology, at speed n, with GRF $\mathcal{I}_* : \mathcal{P}_2(\mathbb{R}) \to [0,\infty]$ defined by*

$$\begin{aligned} \mathcal{I}_*(\mu) &:= \mathcal{I}(\mu) - \inf_{\mu\in\mathcal{P}_2(\mathbb{R})} \mathcal{I}(\mu), \\ \mathcal{I}(\mu) &:= H(\mu|\ell) + \frac{1}{2}\int_{\mathbb{R}\times\mathbb{R}} G(x,y)\mu(dx)\mu(dy) + \int_{\mathbb{R}} F(x)\mu(dx), \end{aligned} \tag{2.11}$$

with H being the relative entropy functional defined in (1.7).

Proposition 2.2.11 (Gibbs measures)**.** *Suppose F, G and ℓ satisfy the conditions stated above, and for $n \in \mathbb{N}$, suppose $X^{(n)}$ is drawn from P_n of (2.10). Then $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption 2.1.3* with GRF*

$$J_X(x) := \inf\left\{\mathcal{I}_*(\mu) : \mu \in \mathcal{P}_2(\mathbb{R}), x = \sqrt{M_2(\mu)}\right\}, \quad x \geq 0,$$

with $\mathcal{I}_*$ as defined in (2.11).

Remark 2.2.12. In a similar fashion, the large deviation results in the recent work of [59, Theorem 2.8] can be combined with the contraction principle to show that Assumption 2.1.3* is also satisfied by sequences of Gibbs measures associated with a class of interaction potentials $G : \mathbb{R}^k \to (-\infty,\infty]$ that capture k-tuple interactions, for fixed $k \in \mathbb{N}$.

2.3 Results in the constant regime

To cover more general sequences $\{X^{(n)}\}_{n\in\mathbb{N}}$ like the uniform measure on an ℓ_p^n ball with $p \in [1,2)$, we propose the following modification of Assumption 2.1.3*.

Assumption 2.3.1. There exists a positive sequence $\{s_n\}_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} s_n = \lim_{n\to\infty} n/s_n = \infty$ such that the sequence of scaled norms $\{\sqrt{s_n}\|X^{(n)}\|_2/n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ at speed s_n with GRF $J_X : \mathbb{R} \to [0,\infty]$.

Remark 2.3.2. It is easy to see, by a simple rescaling argument, that Assumption 2.1.3* is equivalent to a modified version of Assumption 2.3.1, in which one requires that $\{\sqrt{s_n}\|X^{(n)}\|_2/n)\}_{n\in\mathbb{N}}$ satisfies an LDP at a speed s_n that satisfies $s_n/n \to r$, with $r \in (0,\infty)$ rather than $r = 0$. Indeed, this modified version of Assumption 2.3.1 would hold with GRF $J_X^{(r)}$ if and only if Assumption 2.1.3* is satisfied with GRF $J_X(x) := rJ_X^{(r)}(\sqrt{r}x)$.

Theorem 2.3.3 (constant, $k_n \equiv k$)**.** *Suppose $\{k_n\}_{n\in\mathbb{N}}$ is constant at $k \in \mathbb{N}$, and that Assumption 2.3.1 holds, with sequence $\{s_n\}_{n\in\mathbb{N}}$ and GRF J_X. Then $\{n^{-1/2}\mathbf{A}_{n,k}^T X^{(n)}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^k$ at speed s_n, with GRF $I_{\mathbf{A}X,k} : \mathbb{R}^k \to [0,\infty]$ defined by*

$$I_{\mathbf{A}X,k}(x) := \inf_{c>0}\left\{J_X\left(\frac{\|x\|_2}{c}\right) + \frac{c^2}{2}\right\}. \tag{2.12}$$

The proof of Theorem 2.3.3 is given in Section 2.7.2, where the role of Assumption 2.3.1 in the proof is discussed in detail. The rate function takes the form in (2.12) because rare events for $n^{-1/2}\mathbf{A}_{n,k}^T X^{(n)}$ occur through a combination of the "radial" component represented by J_X, and the "angular" component represented by $c \mapsto c^2/2$.

As an immediate corollary of Theorem 2.3.3, we have the following LDP for the corresponding scaled ℓ_q^k norms (or in the case $q \in (0,1)$, quasi-norms) of random projections.

Corollary 2.3.4 (constant, $k_n \equiv k$)**.** *Suppose $\{k_n\}_{n\in\mathbb{N}}$, $\{X^{(n)}\}_{n\in\mathbb{N}}$, $\{s_n\}_{n\in\mathbb{N}}$ and $I_{\mathbf{A}X,k}$ are as in Theorem 2.3.3. Then $\{n^{-1/2}Y^n_{q,k}\}_{n\in\mathbb{N}}$, defined in (2.7), satisfies an LDP at speed s_n with GRF*

$$\mathbb{J}_q^{\mathrm{con}}(x) := \inf_{z\in\mathbb{R}^k}\left\{I_{\mathbf{A}X,k}(z) : \|z\|_q = x\right\}, \quad x \in \mathbb{R}_+. \tag{2.13}$$

Proof. This is an immediate consequence of the LDP for $\{n^{-1/2}\mathbf{A}^T_{n,k}X^{(n)}\}_{n\in\mathbb{N}}$ in Theorem 2.3.3 and the contraction principle (Lemma 1.1.8) applied to the continuous mapping $\mathbb{R}^k \ni x \mapsto \|x\|_q \in \mathbb{R}$. □

2.4 Results in the sublinear regime

Recall that if, instead of being constant, the sequence k_n tends to infinity as $n \to \infty$, then our goal is to establish an LDP for the sequence of empirical measures $\{L^n\}_{n\in\mathbb{N}}$ of (2.1). We start in this section by analyzing the sublinear regime. Section 2.4.1 summarizes our LDP results for the sequences of empirical measures $\{L^n\}_{n\in\mathbb{N}}$ and Euclidean norms of the randomly projected vectors. Section 2.4.2 contains additional results on LDPs for a different scaling of Euclidean norms as well as more general q-norms, with $q \in [1,2)$, of projected vectors that is suitable for the sublinear regime.

2.4.1 LDPs for the empirical measures and norms of projected vectors

In what follows, we will write γ_σ to denote the Gaussian measure on $\mathbb{R}$ with mean 0 and variance σ^2; that is, for $\sigma > 0$, let

$$\mathcal{P}(\mathbb{R}) \ni \gamma_\sigma \sim \mathcal{N}(0,\sigma^2). \tag{2.14}$$

Theorem 2.4.1 (sublinear, $1 \ll k_n \ll n$)**.** *Suppose $\{k_n\}_{n\in\mathbb{N}}$ grows sublinearly and Assumption 2.1.3 holds with associated speed s_n and GRF J_X. Also, suppose that J_X has a unique minimum at $m > 0$. Let H be the relative entropy functional defined in (1.7). Then, for every $q \in [1, 2)$,*

1. *If $s_n \gg k_n$, $\{L^n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}_q(\mathbb{R})$ at speed k_n, with GRF $\mathbb{I}_{L,k_n} : \mathcal{P}_q(\mathbb{R}) \to [0, \infty]$, defined by*
$$\mathbb{I}_{L,k_n}(\mu) := H(\mu|\gamma_m).$$

2. *If $s_n = k_n$, $\{L^n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}_q(\mathbb{R})$ at speed k_n, with GRF $\mathbb{I}_{L,k_n} : \mathcal{P}_q(\mathbb{R}) \to [0, \infty]$, defined by*
$$\mathbb{I}_{L,k_n}(\mu) := \inf_{c>0} \{H(\mu|\gamma_c) + J_X(c)\}.$$

3. *If $s_n \ll k_n$, $\{L^n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}_q(\mathbb{R})$ at speed s_n, with GRF $\mathbb{I}_{L,k_n} : \mathcal{P}_q(\mathbb{R}) \to [0, \infty]$, defined by*
$$\mathbb{I}_{L,k_n}(\mu) := \begin{cases} J_X(c), & \mu = \gamma_c, \\ +\infty, & \textit{otherwise}. \end{cases}$$

As in Theorem 2.3.3, the rate functions above can again be decomposed into a radial component, represented by J_X (as a consequence of Assumption 2.1.3), and "an angular" component $\mathbf{A}_{n,k_n}$, which is captured by the relative entropy term. Depending on the relative rate of growth of k_n and s_n, different parts dominate the rate function, with both terms being present only when $s_n = k_n$.

Next, as in the constant regime (see Corollary 2.3.4), we also establish LDPs for the sequence of scaled Euclidean norms of the random projections. To deal with cases when Assumption 2.1.3* is not satisfied, it is not sufficient to consider Assumption 2.3.1 as in the constant regime. Instead, we will need to introduce the following refinement of Assumption 2.1.3.

Assumption 2.4.2. There exist $r \in [0, \infty]$, a GRF $J_X^{(r)} : \mathbb{R} \to [0, \infty]$, and a positive sequence $\{s_n\}_{n\in\mathbb{N}}$ satisfying $s_n \to \infty$, $s_n/n \to 0$ and $s_n/k_n \to r$ as $n \to \infty$, such that

1. if $r \in [0, \infty)$, then $\{\sqrt{k_n}\|X^{(n)}\|_2/n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF $J_X^{(r)}$;

2. if $r = \infty$, then $\{\sqrt{s_n}\|X^{(n)}\|_2/n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF $J_X^{(\infty)}$.

The following simple observation is analagous to the one made in Remark 2.3.2.

Remark 2.4.3. It is easy to see that Assumption 2.4.2 holds with $r \in (0, \infty)$, $\{s_n\}_{n\in\mathbb{N}}$ and GRF $J_X^{(r)}$ if and only if it also holds with $r' = 1$, $\{s'_n := k_n\}_{n\in\mathbb{N}}$, and GRF $J_X^{(1)}(x) := rJ_X^{(r)}(\sqrt{r}x)$, $x \in [0, \infty)$. Therefore, in essence, one need only consider the cases $r \in \{0, 1, \infty\}$ in Assumption 2.4.2.

Theorem 2.4.4. *Suppose k_n grows sublinearly, and recall the definition of $Y_{q,k}^n$ given in (2.7).*

1. *If Assumption 2.1.3* holds with GRF J_X, then $\{n^{-1/2}Y_{2,k_n}^n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ at speed n with GRF $\mathbb{J}_2^{\mathrm{sub}} : \mathbb{R} \to [0, \infty]$, defined by*

$$\mathbb{J}_2^{\mathrm{sub}}(x) := \inf_{c\in(0,1)} \left\{ -\frac{1}{2}\log(1 - c^2) + J_X\left(\frac{x}{c}\right) \right\}.$$

2. *If Assumption 2.4.2 holds with $r \in [0, \infty]$, $\{s_n\}_{n\in\mathbb{N}}$ and GRF $J_X^{(r)}$, then $\{n^{-1/2}Y_{2,k_n}^n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$, at speed s_n when $r \in \{0, \infty\}$ and at speed k_n when $r \in (0, \infty)$, with GRF $\mathbb{J}_2^{\mathrm{sub}} : \mathbb{R} \to [0, \infty]$, where*

$$\mathbb{J}_2^{\mathrm{sub}}(x) := \begin{cases} J_X^{(0)}(x), & \text{if } r = 0, \\ \inf_{c>0}\left\{\frac{c^2-1}{2} - \log c + rJ_X^{(r)}\left(\frac{\sqrt{r}x}{c}\right)\right\}, & \text{if } r \in (0, \infty), \\ \inf_{c>0}\left\{\frac{c^2}{2} + J_X^{(\infty)}\left(\frac{x}{c}\right)\right\}, & \text{if } r = \infty. \end{cases}$$

Note that there is a transition in the form of the LDP depending on the relative growth rates of $\{s_n\}_{n\in\mathbb{N}}$ and $\{k_n\}_{n\in\mathbb{N}}$. The implications of these results for the special case when $X^{(n)}$ is the uniform measure on an ℓ_p^n ball, and their relation to the work of [3], is discussed in Section 2.6.1.2; see Theorem 2.6.4 and Remark 2.6.5.

2.4.2 LDP for an alternative scaling of q-norms of projections

In this section, we show that we can also establish LDPs for a different scaling of the norm, and in this case we consider not just the Euclidean norm, but q-norms for $q \in [1, 2]$. More precisely, we consider the sequence $\{k_n^{-1/q} Y^n_{q,k_n}\}_{n\in\mathbb{N}}$. For $q \in [1, 2)$ and $t \in \mathbb{R}$ or $q = 2$ and $t < 1/2$, define

$$\Lambda_q(t) := \log \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(t\,|x|^q - \frac{1}{2}x^2\right) dx, \quad t \in \mathbb{R}, \tag{2.15}$$

and let Λ_q^* be the Legendre transform of Λ_q.

Theorem 2.4.5. *Fix $q \in [1, 2]$, suppose $1 \ll k_n \ll n$ and Assumption 2.1.3 holds with speed s_n and GRF J_X, which additionally has a unique minimum at $m > 0$. Also, recall the definition of $Y^n_{q,k}$ given in (2.7). Then $\{k_n^{-1/q} Y^n_{q,k_n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $s_n \wedge k_n$ with GRF $\widehat{\mathbb{J}}_q^{\mathrm{sub}} : \mathbb{R} \to [0, \infty]$ defined by*

$$\widehat{\mathbb{J}}_q^{\mathrm{sub}}(x) := \begin{cases} \Lambda_q^*\left(x^q/m^q\right), & \text{if } s_n \gg k_n, \\ \inf_{c>0}\left\{\Lambda_q^*(c^q) + J_X\left(x/c\right)\right\}, & \text{if } s_n = k_n, \\ J_X(x/\mathcal{M}_q^{1/q}), & \text{if } s_n \ll k_n, \end{cases} \tag{2.16}$$

for $x \geq 0$, and $\widehat{\mathbb{J}}_q^{\mathrm{sub}}(x) := +\infty$ for $x < 0$.

The proof of Theorem 2.4.5 is given in Section 2.8. When $q \in [1,2)$, the LDP for $\{k_n^{-1/q} Y^n_{q,k_n}\}_{n\in\mathbb{N}}$ is an immediate consequence of Theorem 2.4.1 and the contraction principle. For $q = 2$, the contraction principle no longer applies since the moment map $M_2(\cdot)$ is not continuous in $\mathcal{P}_q(\mathbb{R})$ for $q < 2$. We take a different approach to the proof for all cases $q \in [1,2]$, by looking directly at the norm, instead of using the LDP of empirical measures. In the special case when $q = 2$ and $X^{(n)}$ is uniformly distributed on the scaled ℓ_p^n ball of radius $n^{1/p}$, with $p \in [2,\infty]$ (or admits a slightly more general representation), the LDP for $\{\widetilde{Y}^n_{2,k_n}\}_{n\in\mathbb{N}}$ was also obtained in [4, Theorem B].

2.5 Results in the linear regime

Theorem 2.5.1 (linear, $k_n \sim \lambda n$)**.** *Fix $q \in [1,2)$. Suppose $\{k_n\}_{n\in\mathbb{N}}$ grows linearly with rate $\lambda \in (0,1]$ and Assumption 2.1.3 holds with sequence $\{s_n\}_{n\in\mathbb{N}}$ and GRF J_X. Then $\{L^n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}_q(\mathbb{R})$ at speed s_n with GRF $\mathbb{I}_{L,\lambda} : \mathcal{P}_q(\mathbb{R}) \to [0,\infty]$, where*

1. *If $s_n = n$, then for $\mu \in \mathcal{P}_q(\mathbb{R})$,*

$$\mathbb{I}_{L,\lambda}(\mu) = \begin{aligned} &\inf_{c > \sqrt{\lambda M_2(\mu)}} \left\{ J_X(c) - \tfrac{1-\lambda}{2} \log\left(1 - \tfrac{\lambda M_2(\mu)}{c^2}\right) + \lambda \log(c) \right\} \\ &- \lambda h(\mu) + \tfrac{\lambda}{2} \log(2\pi e) + \tfrac{1-\lambda}{2} \log(1-\lambda), \end{aligned} \tag{2.17}$$

where we use the convention that $0 \log 0 = 0$.

2. *If $s_n \ll n$, then for $\mu \in \mathcal{P}_q(\mathbb{R})$*

$$\mathbb{I}_{L,\lambda}(\mu) := \begin{cases} J_X(c), & \mu = \gamma_c \\ +\infty, & \textit{otherwise.} \end{cases}$$

The LDP of the sequence of ℓ_q norms of the randomly projected vectors is given in the following theorem.

Theorem 2.5.2. *Suppose $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption 2.1.3, and $k_n \sim \lambda n$ for some $\lambda \in (0,1]$. Then, for $q \in [1,2]$ the sequence $\{n^{-1/q} Y^n_{q,k_n}\}_{n\in\mathbb{N}}$ defined in (2.7) satisfies an LDP at speed s_n with GRF $\mathbb{J}^{\text{lin}}_{q,\lambda}$, where for $x \in \mathbb{R}_+$,*

$$\mathbb{J}^{\text{lin}}_{q,\lambda}(x) := \begin{cases} \inf_{\nu\in\mathcal{P}(\mathbb{R}), c\in\mathbb{R}_+} \left\{ \mathcal{H}_\lambda(\nu) + J_X(c) : \lambda M_q(\nu) = \left(\frac{x}{c}\right)^q \right\}, & \text{if } s_n = n, \\ J_X\left(\frac{x}{(\lambda \mathcal{M}_q)^{1/q}}\right), & \text{if } s_n \ll n, \end{cases} \tag{2.18}$$

with M_q the q-th moment map as in (1.3) and $\mathcal{M}_q$ the q-th absolute moment of a standard Gaussian random variable defined in (2.8).

The proof of Theorem 2.5.2 is deferred to Section 2.9.2. It is shown there that when $q \in [1,2)$, the result is an immediate consequence of Theorem 2.5.1 and an application of the contraction principle. However, even though the rate function still takes an analogous form when $q = 2$, the contraction principle is no longer applicable in that setting because the LDP of Theorem 2.5.1 only holds in the q-Wasserstein topology for $q \in [1,2)$. Despite this apparent gap, using a different argument in Section 2.9.2 we show that the result nevertheless holds. In the process, we provide an alternative representation for the rate function (2.18) for all $q \in [1,2]$ (see Proposition 2.9.1). This is a manifestation of a somewhat nuanced technical issue, which is elaborated upon in Remark 2.9.4.

2.6 Examples satisfying the main assumptions

In this section, we present several examples of sequences of random vectors $\{X^{(n)}\}_{n\in\mathbb{N}}$ that satisfy the asymptotic thin-shell assumptions.

2.6.1 Scaled ℓ_p^n balls for $p \in [1, 2)$

We consider random vectors uniformly distributed on scaled ℓ_p^n balls. More precisely, for $n \in \mathbb{N}$ and $p > 0$, define $X^{(n,p)}$ to be a random vector uniformly distributed on $n^{1/p}\mathbb{B}_p^n$, where $\mathbb{B}_p^n$ denotes the unit ℓ_p^n-ball:

$$\mathbb{B}_p^n := \left\{ x \in \mathbb{R}^n : \sum_{k=1}^n |x_k|^p \leq 1 \right\}.$$

2.6.1.1 Preliminaries

For $p \in [1, \infty)$, let f_p be the density of the p-generalized normal distribution:

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

where Γ denotes the Gamma function, and let $\xi^{(p)}$ denote a p-generalized normal random variable, namely one with density f_p. We provide here a useful tail estimate: for $t > 0$,

$$2\exp\left(-\frac{1}{p}t^{p/2} - \frac{p-1}{2}\log t\right) \geq \mathbb{P}(|\xi^{(p)}|^2 > t) \geq \frac{t^{p/2}}{t^{p/2}+1}\exp\left(-\frac{1}{p}t^{p/2} - \frac{p-1}{2}\log t\right). \quad (2.19)$$

This estimate was proved in [3, Lemma 5.3] for $p \in [1,2)$, but can easily be extended to include $p = 2$. Now, let

$$m(p) := \left(\frac{p^{2/p}\Gamma\left(1+\frac{3}{p}\right)}{3\Gamma\left(1+\frac{1}{p}\right)} \right)^{1/2}. \tag{2.20}$$

Also, let U be a uniform random variable on $[0,1]$ and $\{\xi_i^{(p)}\}_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables with density f_p independent of U. For $n \in \mathbb{N}$ and $p \in [1,\infty)$, denote $\xi^{(n,p)} = (\xi_1^{(p)}, \ldots, \xi_n^{(p)})$. In this section, we take advantage of the following useful representation of $X^{(n,p)}$ obtained in [75, Lemma 1]:

$$X^{(n,p)} \overset{(d)}{=} n^{1/p} U^{1/n} \frac{\xi^{(n,p)}}{\left\|\xi^{(n,p)}\right\|_p}. \tag{2.21}$$

In view of the representation (2.21), we first establish a property of p-generalized normal random variables which we strengthen from the result in [65].

Lemma 2.6.1. *Let $p \in [1,2]$, let $\{k_n\}_{n\in\mathbb{N}}$ satisfy $k_n \to \infty$ as $n \to \infty$. Then, given any $\{b_n\}_{n\in\mathbb{N}}$ such that $k_n/b_n \to 0$ as $n \to \infty$, the sequence $\{b_n^{-1}\sum_{i=1}^{k_n}(\xi_i^{(p)})^2\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $b_n^{p/2}$ and GRF $\mathcal{J}_{\xi,p} : \mathbb{R} \to [0,\infty]$ defined by*

$$\mathcal{J}_{\xi,p}(t) := \begin{cases} \frac{t^{p/2}}{p}, & t \geq 0, \\ +\infty, & t < 0. \end{cases}$$

Proof. Fix $p \in [1,2)$, $t > 0$, and denote $\xi := \xi_i^{(p)}$. Define $m := \mathbb{E}[\xi_1^2]$ and for $n \in \mathbb{N}$ define $S_n := \sum_{i=1}^{k_n}(\xi_i^2 - m)$. Since $k_n/b_n \to 0$ as $n \to \infty$, by Theorem 3 in [65], with n, x_n therein and ε replaced by k_n, $b_n t$ and $1 - p/2$, we see that

$$\mathbb{P}\left(S_n > b_n t\right) = k_n \mathbb{P}\left(\xi_1^2 - m > b_n t\right)(1 + o(1)).$$

Hence, by (2.19) and the convergence $k_n/b_n \to 0$ as $n \to \infty$,

$$\lim_{n\to\infty} \frac{1}{b_n^{p/2}} \log \mathbb{P}\left(\frac{1}{b_n} S_n > t\right) = -\mathcal{J}_{\xi,p}(t). \tag{2.22}$$

For $p = 2$, the sum $\sum_{i=1}^{k_n} \xi_i^2$ is distributed as a chi-squared distribution with k_n degrees of freedom. Hence for $t > 0$, we have the following tail probability estimate:

$$\mathbb{P}(S_n > b_n t) = \frac{1}{2^{k_n/2}\Gamma(k_n/2)} (b_n t + m)^{k_n/2-1} e^{-(b_n t+m)/2}.$$

Since $k_n/b_n \to 0$ as $n \to \infty$, we again have

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{b_n} \log \mathbb{P}\left(\frac{1}{b_n} S_n > t\right) \\
&= -\frac{t}{2} + \lim_{n\to\infty} \left[\frac{k_n/2-1}{b_n} \log(b_n t + m) - \frac{k_n}{2b_n} \log 2 - \frac{k_n/2+1/2}{b_n} \log(k_n/2) - \frac{k_n/2}{b_n}\right] \\
&= -\frac{t}{2} + \lim_{n\to\infty} \left[\frac{k_n}{2b_n} \log \frac{b_n t + m}{k_n} - \frac{\log(b_n t + m)}{b_n} - \frac{k_n}{2b_n} - \frac{\log k_n}{2b_n}\right] \\
&= -\frac{t}{2},
\end{aligned} \tag{2.23}$$

where the second equality follows from Stirling's approximation.

Let $T_n := b_n^{-1} \sum_{i=1}^{k_n} \xi_i^2 = S_n + mk_n/b_n$. We now combine the two estimates (2.22) and (2.23) to show that $\{T_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $b_n^{p/2}$ with GRF $\mathcal{J}_{\xi,p}$. Fix a closed set $F \subset \mathbb{R}$. If $0 \in F$, then $\inf_{x\in F} \mathcal{J}_{\xi,p}(x) = 0$, and the large deviation upper bound is automatic since $\mathbb{P}(T_n \in F) \leq 1$. Suppose $0 \notin F$. Let $b := \inf\{\beta > 0 : \beta \in F\}$. Then for $b > \tau > 0$, by the positivity of T_n, (2.22) or (2.23) and the monotonicity of $\mathcal{J}_{\xi,p}$,

$$\limsup_{n\to\infty} \frac{1}{b_n^{p/2}} \log \mathbb{P}(T_n \in F) \leq \limsup_{n\to\infty} \frac{1}{b_n^{p/2}} \log \mathbb{P}(T_n \in [b, \infty))$$

$$\leq \limsup_{n\to\infty} \frac{1}{b_n^{p/2}} \log \mathbb{P}(T_n \in [b-\tau, \infty))$$
$$= -\mathcal{J}_{\xi,p}(b-\tau).$$

Letting $\tau \to 0$ and appealing to the continuity and monotonicity of $\mathcal{J}_{\xi,p}$ we obtain

$$\limsup_{n\to\infty} \frac{1}{b_n^{p/2}} \log \mathbb{P}(T_n \in F) = -\inf_{x\in F} \mathcal{J}_{\xi,p}(x),$$

which completes the proof of the upper bound.

Next, fix an open set $U \subset \mathbb{R}$. If $0 \in U$, then $\inf_{x\in U} \mathcal{J}_{\xi,p}(x) = 0$. Since U is open, there exists $\varepsilon > 0$ such that $(-\varepsilon, \varepsilon) \subset U$. Since $k_n/b_n \to 0$ as $n \to \infty$, the strong law of large numbers implies $\lim_{n\to\infty} \mathbb{P}(T_n \in (-\varepsilon, \varepsilon)) = 1$. Hence, the large deviation lower bound follows. Next suppose $0 \notin U$. For $\delta > 0$, there exists $\beta \in U$ such that $\mathcal{J}_{\xi,p}(\beta) < \inf_{x\in U} \mathcal{J}_{\xi,p}(x) + \delta$. Pick $\varepsilon > 0$ such that $(\beta - \varepsilon, \beta + \varepsilon) \subset U$. If $\beta < 0$, then the large deviation lower bound is trivial. Suppose $\beta > 0$. Then for $\varepsilon > \tau > 0$,

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{b_n^{p/2}} \log \mathbb{P}(T_n \in U) &\geq \liminf_{n\to\infty} \frac{1}{b_n^{p/2}} \log \mathbb{P}(S_n \in (\beta-\varepsilon, \beta+\varepsilon-\tau)) \\
&= \liminf_{n\to\infty} \frac{1}{b_n^{p/2}} \log \left[\mathbb{P}(S_n \in (\beta-\varepsilon, \infty)) - \mathbb{P}(S_n \in [\beta+\varepsilon-\tau, \infty))\right] \\
&= \liminf_{n\to\infty} \frac{1}{b_n^{p/2}} \log \mathbb{P}(S_n \in (\beta-\varepsilon, \infty)) \\
&= -\mathcal{J}_{\xi,p}(\beta-\varepsilon) \\
&\geq -\inf_{x\in U} \mathcal{J}_{\xi,p}(x) - \delta,
\end{aligned}$$

where the third equality follows by the monotonicity of $\mathcal{J}_{\xi,p}$ and the last inequality by the continuity of $\mathcal{J}_{\xi,p}$ on choosing ε sufficiently small. The large deviation lower bound then follows on sending δ to 0. This completes the proof of the lemma. $\square$

2.6.1.2 Verification of Assumptions 2.3.1 and 2.4.2 when $p \in [1,2)$

When $p \in [1,2)$, it should be noted that annealed LDPs associated with the random vectors $\{X^{(n,p)}\}_{n\in\mathbb{N}}$ defined in Proposition 2.2.8 occur at different speeds than in the case $p > 2$; see Theorem 2.3 of [34]. In particular, from Theorem 1.3 of [43] it follows that in this case the sequence of scaled norms $\{\|X^{(n,p)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $n^{p/2}$. Thus, in this case $\{X^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies Assumption 2.1.3 with $s_n = n^{p/2}$ and unique minimizer $m = m(p)$ defined in (2.20). In particular, Assumption 2.1.3* is not satisfied. We show that nevertheless, Assumption 2.3.1 does hold. In what follows, let U, $\{\xi_i^{(p)}\}_{i\in\mathbb{N}}$, and $\xi^{(n,p)}$ be as in the representation of $X^{(n,p)}$ in (2.21), and note that then

$$\frac{\sqrt{s_n}\left\|X^{(n,p)}\right\|_2}{n} \overset{(d)}{=} \frac{U^{1/n}}{\left\|\xi^{(n,p)}\right\|_p/n^{1/p}} \frac{\sqrt{s_n}\left\|\xi^{(n,p)}\right\|_2}{n}, \quad n \in \mathbb{N}. \tag{2.24}$$

Proposition 2.6.2. *For $p \in [1,2)$, $\{X^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies Assumption 2.3.1 with speed $s_n := n^{2p/(2+p)}$ and GRF*

$$J_{X,p}(x) := \begin{cases} \dfrac{x^p}{p}, & x \geq 0, \\ +\infty, & x < 0. \end{cases} \tag{2.25}$$

Proof. Fix $p \in [1,2)$. It follows from [34, Lemma 3.3] that $\{U^{1/n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n, and from Cramér's theorem and the contraction principle that $\{\|\xi^{(n,p)}\|_p/n^{1/p}\}_{n\in\mathbb{N}}$ also satisfies an LDP at speed n. Moreover, it is easy to see that both sequences converge almost surely to 1. Since, in addition, $\xi^{(n,p)}$ and U are independent, appealing again to the contraction principle it follows that their ratio $W_n := U^{1/n}n^{1/p}/\|\xi^{(n,p)}\|_p$ also satisfies an LDP at speed n. Furthermore, since $W_n \to 1$ almost surely as $n \to \infty$, the LDP rate function for $\{W_n\}_{n\in\mathbb{N}}$ has the unique minimizer 1. Given $s_n \ll n$ and the representation in (2.24), an application of (the last assertion of) Lemma 1.1.9, with W_n as above, $m = 1$,

$\gamma_n = n$, $V_n := \sqrt{s_n}\|\xi^{(n,p)}\|_2/n$, and $\beta_n = s_n$ shows that to establish the proposition it suffices to prove that $\{V_n\}_{n\in\mathbb{N}}$ satisfies an LDP with speed s_n and GRF $J_{X,p}$. To this end, using the relation $\frac{\sqrt{s_n}\|\xi^{(n,p)}\|_2}{n} = \left(\frac{s_n}{n^2}\sum_{i=1}^n(\xi_i^{(p)})^2\right)^{1/2}$, the contraction principle, Lemma 2.6.1 with $k_n = n$ and $b_n = n^2/s_n$ therein, and the property that $s_n \ll n$, we can conclude that $\{\sqrt{s_n}\|\xi^{(n,p)}\|_2/n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $n^p s_n^{-p/2}$ with GRF $J_{X,p}$ given in (2.25). Finally, since $s_n = n^{2p/(2+p)}$, we obtain $n^p s_n^{-p/2} = s_n$. □

Next, we turn to verifying Assumption 2.4.2. In the following lemma, we show that for $1 \le p < 2$, different conditions in Assumption 2.4.2 are satisfied according to the growth speed of k_n.

Proposition 2.6.3. *When $p \in [1,2)$, let $J_{X,p}$ be defined as in (2.25). Then $\{X^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies Assumption 2.4.2 with GRF $J_X^{(r)} = J_{X,p}$ for all $r \in [0,\infty]$, where s_n is defined by*

$$s_n := \begin{cases} n^p k_n^{-p/2}, & \textit{if} \quad k_n \gg n^{2p/(2+p)}, \\ n^{2p/(2+p)}, & \textit{if} \quad k_n = n^{2p/(2+p)} \quad \textit{or} \quad k_n \ll n^{2p/(2+p)}. \end{cases}$$

Proof. Let s_n be as defined in the proposition. As in Assumption 2.4.2, let $r \in [0,\infty]$ be the limit of s_n/k_n as $n \to \infty$. It is easy to see that $r = 0$ if $k_n \gg n^{2p/(2+p)}$, $r = 1$ if $k_n = n^{2p/(2+p)}$, and $r = \infty$ when $k_n \ll n^{2p/(2+p)}$. Now, set $c_n := \sqrt{n/k_n}$ if $r \in [0,\infty)$ and $c_n := \sqrt{n/s_n}$ if $r = \infty$ and note that $c_n \to \infty$ as $n \to \infty$. By (2.21), we have the following representation

$$\frac{\left\|X^{(n,p)}\right\|_2}{c_n\sqrt{n}} \overset{(d)}{=} \frac{U^{1/n}}{\left\|\xi^{(n,p)}\right\|_p/n^{1/p}}\frac{\left\|\xi^{(n,p)}\right\|_2}{c_n\sqrt{n}}, \quad n \in \mathbb{N}.$$

We first claim that $\{\|\xi^{(n,p)}\|_2/(c_n\sqrt{n})\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $s_n \ll n$. Indeed, by Lemma 2.6.1 with $k_n = n$ and $b_n = c_n^2 n$ there, the fact that $b_n \to \infty$ as $n \to \infty$ and the contraction principle with the mapping $x \mapsto \sqrt{x}$, $\left\{\|\xi^{(n,p)}\|_2/(c_n\sqrt{n})\right\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $n^{p/2}c_n^p$ with GRF $J_{X,p}$ given in (2.25). By [34, Lemma 3.3], Cramér's theorem and the contraction principle, the independence of $\xi^{(n,p)}$ and U, and the contraction principle, we see that $W_n := U^{1/n} n^{1/p}/\|\xi^{(n,p)}\|_p$ satisfies an LDP at speed n and further, it also converges almost surely to 1. The lemma then follows upon applying Lemma 1.1.9 with $V_n := \|\xi^{(n,p)}\|_2/(c_n\sqrt{n})$ and $W_n := U^{1/n} n^{1/p}/\|\xi^{(n,p)}\|_p$, $m = 1$, $\gamma_n = n$ and $\beta_n := s_n \ll n$. □

When combined with the results of Section 2.4.1, the last two propositions imply the following LDPs for norms of projections in the constant and sublinear regimes. Here, we focus on the norm $Y^{(n,p)}_{2,k_n}$, which is defined as in (2.7) but with $X^{(n)}$ replaced with $X^{(n,p)}$. The corresponding results for the differently scaled norms $\{k_n^{-1/2} Y^{(n,p)}_{2,k_n}\}_{n\in\mathbb{N}}$ were obtained in Theorem B of [4].

Theorem 2.6.4. *Fix $p \in [1, 2)$.*

1. *Suppose k_n is constant at $k \in \mathbb{N}$. Then $\{n^{-1/2}\mathbf{A}^T_{n,k} X^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^k$ at speed $s_n = n^{2p/(2+p)}$ with GRF $I_{\mathbf{A}X^{(p)},k} : \mathbb{R}^k \to [0,\infty]$ defined by*

$$I_{\mathbf{A}X^{(p)},k}(x) := \frac{p+2}{2p}\,\|x\|_2^{2p/(p+2)}, \quad x \in \mathbb{R}^k.$$

Moreover, $\{n^{-1/2} Y^{(n,p)}_{2,k}\}_{n\in\mathbb{N}}$ satisfies an LDP with GRF

$$\mathbb{J}_{Y^{(p)}_{2,k}}(x) := \begin{cases} \frac{p+2}{2p} x^{2p/(p+2)}, & x \geq 0, \\ +\infty, & x < 0. \end{cases} \tag{2.26}$$

2. *Suppose k_n grows sublinearly.*

   (a) *If $k_n \ll n^{2p/(2+p)}$. Then $\{n^{-1/2} Y^{(n,p)}_{2,k}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ at speed $n^{2p/(2+p)}$ with GRF $\mathbb{J}_{Y^{(p)}_{2,k_n}} : \mathbb{R} \to [0,\infty]$, which is equal to $\mathbb{J}_{Y^{(p)}_{2,k}}$ defined in (2.26).*

(b) *If* $k_n = n^{2p/(2+p)}$. *Then* $\{n^{-1/2}Y_{2,k}^{(n,p)}\}_{n\in\mathbb{N}}$ *satisfies an LDP in* $\mathbb{R}$ *at speed* $n^{2p/(2+p)}$ *with GRF* $\mathbb{J}_{Y^{(p)}_{2,k_n}} : \mathbb{R} \to [0,\infty]$ *defined by*

$$\mathbb{J}_{Y^{(p)}_{2,k_n}}(x) := \begin{cases} \frac{p+2}{2p}\frac{x^p}{\bar{c}(x)^p} - \log(\bar{c}(x)), & x \geq 0, \\ +\infty, & x < 0. \end{cases}$$

where $\bar{c}(x) \in [1 + x^{p/(p+2)}, \infty)$ *is the unique positive solution to* $c^{p+2} - c^p - x^p = 0$.

(c) *If* $k_n \gg n^{2p/(2+p)}$. *Then* $\{n^{-1/2}Y_{2,k}^{(n,p)}\}_{n\in\mathbb{N}}$ *satisfies an LDP in* $\mathbb{R}$ *at speed* $n^p k_n^{-p/2}$ *with GRF* $\mathbb{J}_{Y^{(p)}_{2,k_n}} : \mathbb{R} \to [0,\infty]$ *defined by*

$$\mathbb{J}_{Y^{(p)}_{2,k_n}}(x) := \begin{cases} \frac{x^p}{p}, & x \geq 0, \\ +\infty, & x < 0, \end{cases}$$

which is equal to $J_{X,p}$ *in* (2.25) *of Proposition* 2.6.2.

Proof. In the constant regime, Proposition 2.6.2 shows that $\{X^{(n,p)}\}$ satisfies Assumption 2.3.1 with rate function $J_{X,p}$ taking the explicit form given in (2.25). Subsituting this into Theorem 2.3.3, we see that $\{n^{-1/2}\mathbf{A}_{n,k}^T X^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^k$ at speed $s_n = n^{2p/(2+p)}$ with GRF

$$I_{\mathbf{A}X^{(p)},k}(x) = \inf_{c>0}\left\{\frac{\|x\|_2^p}{pc^p} + \frac{c^2}{2}\right\} = \frac{p+2}{2p}\|x\|_2^{2p/(p+2)}, \quad x \in \mathbb{R}^k. \tag{2.27}$$

The LDP for the norm follows from the contraction principle applied to $x \mapsto \|x\|_2$.

In the sublinear regime, Proposition 2.6.3 shows that $\{X^{(n,p)}\}$ satisfies Assumption 2.4.2 with GRF $J_{X,p}$ (for all values of $r \in [0,\infty]$) and at speed $s_n = n^{2p/(2+p)}$ in case (a), which corresponds to $r = \infty$, and in case (b), which corresponds to $r = 1$, and speed

$s_n = n^p k_n^{-p/2}$ in case (c), which corresponds to $r = 0$. Together with Theorem 2.4.4(ii), this immediately implies case (c) and case (a) follows from the second equality in (2.27) above. On the other hand, if case (b) holds, Theorem 2.4.4(ii) and the expression for $J_{X,p}$ in (2.25) show that for $x \geq 0$,

$$\begin{aligned}\mathbb{J}_{Y^{(p)}_{2,k_n}}(x) &:= \inf_{c>0}\left\{\frac{c^2-1}{2} - \log c + \frac{x^p}{pc^p}\right\} \\ &= \frac{p+2}{2p}\frac{x^p}{\bar{c}(x)^p} - \log(\bar{c}(x)),\end{aligned}$$

where $\bar{c}(x)$ is the unique positive solution to $c^{p+2} - c^p - x^p = 0$, whose existence and uniqueness is guaranteed by Descartes' rule of signs (see e.g. [5]). Since $x \geq 0$, the unique positive solution satisfies $\bar{c}(x) \geq 1$. Furthermore, $\bar{c}(x) = (\bar{c}(x)^p + x^p)^{1/(p+2)} \geq (1 + x^p)^{1/(p+2)} \geq 1 + x^{p/(p+2)}$. □

In a similar fashion, applying Theorem 2.4.1 and Theorem 2.5.1, one can also deduce LDPs for the empirical measures of the coordinates of the sequences of projections of $X^{(n)} = X^{(n,p)}$ in the sublinear and linear regimes. Furthermore, in the linear regime, similar LDP results for the norms of these projections can be deduced from Theorem 2.5.2 on noting that $\{X^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies Assumption 2.1.3. But, we do not state these results since they were already obtained in Theorem 1.2 of [3].

Remark 2.6.5. Theorem 2.6.4 shows that when $p \in [1, 2)$, the speed of the LDP for Euclidean norms of k_n-dimensional projections of ℓ_p^n balls exhibits an interesting phase transition depending on whether the ratio $k_n/n^{2p/(2+p)}$ is asymptotically finite or infinite, and the form of the rate function also differs depending on whether the limit is zero, strictly positive or infinite. Indeed, note that $x^p \geq x^p/\bar{c}(x)^p = \bar{c}(x)^2 - 1 \geq x^{2p/(p+2)}$. Hence, a comparison of cases (b) and (c) of the last theorem above shows that the faster the growth of the speed k_n, the faster the growth of the rate function $\mathbb{J}_{Y^{(p)}_{2,k_n}}(x)$, as $x \to \infty$. This disproves a conjecture in [3] (see the statement after Theorem 1.2 therein), which states that the LDP (that is, speed and rate function) of these Euclidean norms should be the same in the constant and sublinear cases. This behavior is in contrast to the case of ℓ_p balls, with $p > 2$, where the rate functions of the (Euclidean) norms of the projections in the sublinear and constant cases coincide (see Proposition 2.2.8).

2.6.2 Orlicz and generalized Orlicz balls

We now consider the class of *Orlicz and generalized Orlicz balls*. These form a natural class of examples because, like the uniform measure on any convex set, the uniform measure on a generalized Orlicz ball is logconcave, and much like ℓ_p^n balls, the coordinates of a uniformly distributed random vector satisfy a certain "subindependence" property known as "negative association" [70], and (under suitable conditions) the Orlicz ball satisfies the KLS conjecture [54].

Definition 2.6.6. We say V is an *Orlicz function* if $V : \mathbb{R} \to \mathbb{R}_+ \cup \{\infty\}$ is convex and satisfies $V(0) = 0$ and $V(x) = V(-x)$ for $x \in \mathbb{R}$. Let the domain of V be $D_V := \{x \in \mathbb{R} : V(x) < \infty\}$. We say V is *superquadratic* if

$$V(x)/x^2 \to \infty, \qquad \text{as } x \to \infty. \tag{2.28}$$

In particular, 2-convex Orlicz functions are superquadratic in the sense of Definition 2.6.6, and this includes the case $V(x) = |x|^p$, with $p > 2$. Fix a superquadratic Orlicz function V, and denote the associated symmetric Orlicz ball by

$$\mathbb{B}_V^n := \left\{ x \in \mathbb{R}^n : \sum_{i=1}^n V(x_i) \le n \right\}. \tag{2.29}$$

We also discuss so-called *generalized* Orlicz balls, associated with a sequence of superquadratic Orlicz functions $V_i : \mathbb{R} \to \mathbb{R}_+$, $i \in \mathbb{N}$. These generalized Orlicz balls, which are induced by a norm defined by functions that vary with dimension, are also known in the literature as *Musielak-Orlicz balls* [63]. For $n \in \mathbb{N}$, define the set

$$\mathbb{B}^n_{V_1,\dots,V_n} := \left\{ x \in \mathbb{R}^n : \sum_{i=1}^n V_i(x_i) \le n \right\}. \tag{2.30}$$

In contrast with the usual (symmetric) Orlicz ball, the generalized Orlicz ball $B^n_{V_1,\dots,V_n}$ need not be invariant to permutations of the coordinates.

We verify Assumption 2.1.3* for symmetric and generalized Orlicz balls in Sections 2.6.2.1 and 2.6.2.2, respectively. As mentioned above, the special case $\mathbb{B}^n_V = n^{1/p}\mathbb{B}^n_p$, with $p > 2$, was analyzed in Proposition 2.2.8. However, unlike $\mathbb{B}^n_p$, whose study is facilitated by the probabilistic representation (2.21), there is no analogous representation for general $\mathbb{B}^n_V$, making its analysis significantly more complicated. Instead, our proof shows that one can represent $\mathbb{B}^n_V$ in terms of an exponential Gibbs measure; this idea has been subsequently further exploited in [42]. Conditions like 2-convexity (and 2-concavity) often arise in the local theory of Banach spaces. In our case this condition arises naturally because we are also considering 2-norms in this section. However, it would be interesting to investigate whether an analogous analysis can be carried out for Orlicz balls defined via superlinear but subquadratic Orlicz functions V (it is worth noting that we have been able to cover the special case of ℓ^n_p balls with $p \in [1, 2)$) or whether there is a more fundamental obstruction. In the following, by a slight abuse of notation, we use $|A|$ to denote the volume of a measurable set $A \subset \mathbb{R}^n$ and use dx to denote the integral over Lebesgue measure in $\mathbb{R}$ or $\mathbb{R}^n$.

2.6.2.1 Symmetric Orlicz balls

In analogy with (1.3), for $\nu \in \mathcal{P}(\mathbb{R})$, define $M_V : \mathcal{P}(\mathbb{R}) \to \mathbb{R}_+$ as

$$M_V(\nu) := \int_{\mathbb{R}} V(x)\nu(dx), \tag{2.31}$$

for ν such that the integral in the last display is finite. Also, for $b > 0$, define $\mu_{V,b} \in \mathcal{P}(\mathbb{R})$ as follows: for $x \in D_V$, where D_V is the domain of V as specified in Definition 2.6.6,

$$\mu_{V,b}(dx) := \frac{1}{Z_{V,b}} e^{-b\,V(x)} dx, \tag{2.32}$$

with $Z_{V,b}$ being the normalizing constant $Z_{V,b} := \int_{D_V} e^{-bV(x)}dx$. Moreover, define

$$\mathcal{J}(u, v) := \sup_{s<0, t\in\mathbb{R}} \left\{ su + tv - \log\left(\int_{D_V} e^{sV(x)+tx^2} dx \right) \right\} \quad \text{for} \quad u, v \in \mathbb{R}_+, \tag{2.33}$$

We now state the main result of this section.

Proposition 2.6.7 (Annealed example, Orlicz)**.** *Suppose V is a superquadratic Orlicz function and for $n \in \mathbb{N}$, let $X^{(n)} \sim \mathrm{Unif}(\mathbb{B}^n_V)$. Then $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption 2.1.3* with $J_X = J_{X,V}$, where*

$$J_{X,V}(z) := \mathcal{J}(1, z^2) - \sup_{s<0} \left\{ s - \log\left(\int_{D_V} e^{sV(x)} dx \right) \right\}, \quad z \in \mathbb{R}_+. \tag{2.34}$$

Moreover, there exists a unique $b^* > 0$ such that $M_V(\mu_{V,b^*}) = 1$ and $J_{X,V}$ has a unique minimizer $m := \sqrt{M_2(\mu_{V,b^*})}$, with M_2 as defined in (1.3).

The proof of Proposition 2.6.7, which is given at the end of the section, relies on certain properties of the function $\mathcal{J}$ specified in (2.33), which we state in the lemma below, whose proof is relegated to Section 2.11. For $s < 0$, $t \in \mathbb{R}$, define $\nu_{s,t} \in \mathcal{P}(\mathbb{R})$ as

$$\nu_{s,t}(dx) := \frac{1}{Z_{s,t}} e^{sV(x)+tx^2} dx, \quad x \in D_V,$$

where $Z_{s,t}$ is the normalizing constant $Z_{s,t} := \int_{D_V} e^{sV(x)+tx^2} dx$, which is finite since V is superquadratic and $s < 0$.

Lemma 2.6.8. *Suppose V is a superquadratic Orlicz function and let $\mathcal{J}$ be defined in (2.33).*

1. *For $u, v \in \mathbb{R}_+$ and (u, v) lying in the interior of the domain of $\mathcal{J}$, there exists a unique $(s, t) \in \mathbb{R}_- \times \mathbb{R}$ such that the supremum in the definition of (2.33) of $J(u, v)$ is attained. Moreover, this (s, t) satisfies $M_V(\nu_{s,t}) \le u$ and $M_2(\nu_{s,t}) = v$.*

2. *There exists a unique constant $b^* > 0$ such that $M_V(\mu_{V,b^*}) = 1$. Moreover, $v \mapsto \mathcal{J}(1, v)$ is a convex function with minimizer $m = M_2(\mu_{V,b^*})$, and furthermore,*

$$\min_{x \in \mathbb{R}_+} \mathcal{J}(1, x) = \mathcal{J}(1, m) = \sup_{s<0} \left\{ s - \log\left(\int_{D_V} e^{sV(x)} dx \right) \right\}. \tag{2.35}$$

3. *For $v > m$, the supremum in the definition (2.33) of $\mathcal{J}(1, v)$ is attained at $(s, t) \in \mathbb{R}_- \times \mathbb{R}_+$ while for $0 < v < m$, the supremum is attained at $(s, t) \in \mathbb{R}_- \times \mathbb{R}_-$.*

4. *For $u, v \in \mathbb{R}_+$ such that $\mathcal{J}(u, v) < \infty$, $\partial_u \mathcal{J}(u, v) \le 0$.*

Remark 2.6.9. Let $V, \tilde{V}$ be Orlicz functions, and suppose now that the definition of $\mathcal{J}$ in (2.33) is replaced with

$$\tilde{\mathcal{J}}(u, v) := \sup_{s<0, t\in\mathbb{R}} \left\{ su + tv - \log\left(\int_{D_V} e^{sV(x)+t\tilde{V}(x)} dx \right) \right\} \quad \text{for} \quad u, v \in \mathbb{R}_+. \tag{2.36}$$

Then, as can be seen from the proof in Section 2.11, the conclusions of Lemma 2.6.8 continues to hold, with $M_{\tilde{V}}$ and $\tilde{\mathcal{J}}$ in place of M_2 and $\mathcal{J}$, respectively, as long as

$V(x)/\tilde{V}(x) \to \infty$ as $x \to \infty$.

Proof of Proposition 2.6.7. Fix a superquadratic Orlicz function V (see Definition 2.6.6). For a measurable set $A \subset \mathbb{R}$, define $\mathbb{B}^n_{2,V}[A] := \mathbb{B}^n_V \cap \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 \in nA\}$, and note that

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n (X_i^n)^2 \in A\right) = \frac{\left|\mathbb{B}^n_{2,V}[A]\right|}{\left|\mathbb{B}^n_V\right|}, \tag{2.37}$$

which expresses a tail probability in terms of the ratio of volumes of two convex bodies. We now proceed in three steps to characterize the asymptotics of these volumes in order to estimate the tail probabilities.

Step 1. For any closed set $F \subset \mathbb{R}_+$, we will establish an upper bound on the numerator in (3.2.14),

$$\limsup_{n\to\infty} \frac{1}{n} \log \left|\mathbb{B}^n_{2,V}[F]\right| \leq -\inf_{x\in F} \mathcal{J}(1,x). \tag{2.38}$$

With $b_* > 0$ as defined in property 2 of Lemma 2.6.8, and M_2 as defined in (1.3), set $m := M_2(\mu_{V,b^*})$. Next, set $\alpha_+ := \min\{x \in [m, +\infty) \cap F\}$ and $\alpha_- := \max\{x \in [0, m] \cap F\}$. Assume $\alpha_- < m < \alpha_+$. Then

$$\begin{aligned}\left|\mathbb{B}^n_{2,V}[F]\right| &\leq \left|\mathbb{B}^n_V \cap \left\{x \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 \geq n\alpha_+\right\}\right| + \left|\mathbb{B}^n_V \cap \left\{x \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 \leq n\alpha_-\right\}\right| \\ &= \left|\mathbb{B}^n_{2,V}[[\alpha_+, \infty)]\right| + \left|\mathbb{B}^n_{2,V}[[0, \alpha_-]]\right|. \end{aligned} \tag{2.39}$$

Fix $s < 0$ and $t > 0$, and note that then

$$\mathbb{B}^n_{2,V}[[\alpha_+, \infty)] = \left\{x \in \mathbb{R}^n : \exp\left(s\sum_{i=1}^n V(x_i)\right) \geq e^{ns}, \exp\left(t\sum_{i=1}^n x_i^2\right) \geq e^{nt\alpha_+}\right\},$$

and therefore, it follows that

$$\begin{aligned}\left|\mathbb{B}^n_{2,V}[[\alpha_+,\infty)]\right| &\leq \int_{\mathbb{B}^n_{2,V}[[\alpha_+,\infty)]} \exp\left(s\sum_{i=1}^n V(x_i) - ns + t\sum_{i=1}^n x_i^2 - nt\alpha_+\right)dx \\ &\leq e^{-ns-nt\alpha_+}\int_{(D_V)^n} \exp\left(s\sum_{i=1}^n V(x_i) + t\sum_{i=1}^n x_i^2\right)dx.\end{aligned}$$

Hence, for every $s < 0$ and $t > 0$, we see that

$$\limsup_{n\to\infty}\frac{1}{n}\log\left|\mathbb{B}^n_{2,V}[[\alpha_+,\infty)]\right| \leq -s - t\alpha_+ + \log\left(\int_{D_V} e^{sV(x)+tx^2}dx\right).$$

Taking the infimum over $s < 0$ and $t > 0$, we obtain

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\left|\mathbb{B}^n_{2,V}[[\alpha_+,\infty)]\right| &\leq \inf_{s<0,t>0}\left\{-s - t\alpha_+ + \log\left(\int_{D_V} e^{sV(x)+tx^2}dx\right)\right\} \\ &= -\sup_{s<0,t>0}\left\{s + t\alpha_+ - \log\left(\int_{D_V} e^{sV(x)+tx^2}dx\right)\right\} \\ &= -\mathcal{J}(1,\alpha_+),\end{aligned}$$

where, since $\alpha_+ > m$, the last equality follows by property 3 of Lemma 2.6.8.

Similarly, once again fixing $s < 0$ but now also choosing $t < 0$, we have

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\left|\mathbb{B}^n_{2,V}[[0,\alpha_-)]\right| &\leq -\sup_{s<0,t<0}\left\{s + t\alpha_- - \log\left(\int_{D_V} e^{sV(x)+tx^2}dx\right)\right\} \\ &= -\mathcal{J}(1,\alpha_-),\end{aligned}$$

where, noting that $\alpha_- < m$, the last inequality once again follows by property 3 of Lemma 2.6.8. Together with (2.39), the last two displays show that

$$\limsup_{n\to\infty}\frac{1}{n}\log|B^n_{2,V}[F]| \leq -\min\left(\mathcal{J}(1,\alpha_+),\mathcal{J}(1,\alpha_-)\right) = -\inf_{x\in F}\mathcal{J}(1,x),$$

where the last equality follows from the convexity of $x \to \mathcal{J}(1,x)$ and the fact that m is the unique minimizer of $\mathcal{J}(1,\cdot)$, as established in property 2 of Lemma 2.6.8, and the definitions of α_- and α_+. This proves the claim (2.38) in this case.

On the other hand, if $\alpha_+ = m$ or $\alpha_- = m$, then by property 2 of Lemma 2.6.8, $\inf_{x\in F} \mathcal{J}(1,x) = \mathcal{J}(1,m)$. We first obtain an estimate of the volume of the Orlicz ball $\mathbb{B}_V^n$. For $s < 0$,

$$\begin{aligned}\frac{1}{n}\log\left|\mathbb{B}_V^n\right| &\leq \frac{1}{n}\log\int_{\sum_{i=1}^n V(x_i)\leq n}\exp\left(s\sum_{i=1}^n V(x_i) - ns\right)dx\\ &\leq -s + \log\left(\int_{D_V} e^{sV(x)}dx\right).\end{aligned}$$

The claim (2.38) then follows in this case as well since

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\left|\mathbb{B}_{2,V}^n[F]\right| &\leq \limsup_{n\to\infty}\frac{1}{n}\log\left|\mathbb{B}_V^n\right|\\ &\leq -\sup_{s<0}\left\{s - \log\left(\int_{D_V} e^{sV(x)}dx\right)\right\}\\ &= -\min_{x\in\mathbb{R}_+}\mathcal{J}(1,x)\\ &\leq -\min_{x\in F}\mathcal{J}(1,x),\end{aligned}$$

where the second inequality follows on taking the infimum over $s < 0$ in the last display and the equality on the third line follows from (2.35).

Step 2. For any open set $U \subset \mathbb{R}_+$, we will show the lower bound

$$\liminf_{n\to\infty}\frac{1}{n}\log\left|\mathbb{B}_{2,V}^n[U]\right| \geq -\inf_{x\in U}\mathcal{J}(1,x). \tag{2.40}$$

Since $\inf_{x\in U}\mathcal{J}(1,x) = \inf_{x\in U\cap D_{\mathcal{J}(1,\cdot)}}\mathcal{J}(1,x)$, with the latter being infinity if the intersection is empty, it suffices to only consider U that lies in the interior of the domain of $\mathcal{J}(1,\cdot)$. By property 4 of Lemma 2.6.8, $\partial_u \mathcal{J}(u,v) \le 0$, and so for each $v \in \mathbb{R}_+$, $u \mapsto \mathcal{J}(u,v)$ is decreasing. By the convexity, and hence continuity, of $(u,v) \mapsto \mathcal{J}(u,v)$ in $\mathbb{R}_+ \times \mathbb{R}_+$, for every $\varepsilon > 0$, there exist $0 < y < 1$ and $z \in U$ such that

$$\inf_{x\in U} \mathcal{J}(1,x) > \mathcal{J}(y,z) - \varepsilon. \tag{2.41}$$

Pick $\delta > 0$ small such that $(z-\delta, z+\delta) \subset U$. Let (s_y, t_z) be the value that attains the supremum in the expression (2.33) for $\mathcal{J}(y,z)$, which exists by property 1 of Lemma 2.6.8 since U lies in the interior of the domain of $\mathcal{J}(1,\cdot)$. Define

$$A_\delta^n := \left\{ x \in \mathbb{R}^n : 0 \le \frac{1}{n}\sum_{i=1}^n V(x_i) < 1, z-\delta < \frac{1}{n}\sum_{i=1}^n x_i^2 < z+\delta \right\}$$

and $\bar{Z}_{y,z} := Z_{s_y,t_z} = \int_{D_V} e^{s_y V(x) + t_z x^2} dx$, which is finite by the definition of (s_y, t_z), the definition of $\mathcal{J}(y,z)$ in (2.33) and the finiteness of $\mathcal{J}(y,z)$. Suppose $t_z < 0$. We then obtain the following estimate for the lower bound:

$$\begin{aligned}
\left|\mathbb{B}_{2,V}^n[U]\right| &\ge \int_{A_\delta^n} dx \\
&= \int_{A_\delta^n} \left(\bar{Z}_{y,z}\right)^n e^{-s_y \sum_{i=1}^n V(x_i) - t_z x_i^2} \prod_{i=1}^n \frac{1}{\bar{Z}_{y,z}} e^{s_y V(x_i) + t_z x_i^2} dx \\
&\ge \exp\left(n(\log \bar{Z}_{y,z} - s_y y - t_z(z-\delta))\right) \int_{A_\delta^n} \prod_{i=1}^n \frac{1}{\bar{Z}_{y,z}} e^{s_y V(x_i) + t_z x_i^2} dx.
\end{aligned}$$

The case $t_z > 0$ can be handled analogously, simply by replacing $z - \delta$ with $z + \delta$ on the right-hand side.

Let $\{\Xi_i\}_{i=1}^\infty$ be a sequence of i.i.d continuous random variables with density $\frac{1}{\bar{Z}_{y,z}} e^{s_y V(x) + t_z x^2}$. Since (s_y, t_z) attains the supremum in (2.33) of $\mathcal{J}(y,z)$, by property 1 of Lemma 2.6.8, $\mathbb{E}[V(\Xi_i)] \leq y < 1$ and $\mathbb{E}[\Xi_i^2] = z$. Then the integral in the last display can be rewritten as

$$\int_{A_\delta^n} \prod_{i=1}^n \frac{1}{\bar{Z}_{y,z}} e^{s_y V(x_i) + t_z x_i^2} dx = \mathbb{P}\left(0 \leq \frac{1}{n}\sum_{i=1}^n V(\Xi_i) < 1, z - \delta < \frac{1}{n}\sum_{i=1}^n \Xi_i^2 < z + \delta\right),$$

which converges to 1 by the weak law of large numbers and finiteness of the respective moments. Thus, combining the last two displays with the expression for $\mathcal{J}(y,z)$ in (2.33) and the definition of (s_y, t_z), we obtain

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log \left|\mathbb{B}_{2,V}^n[U]\right| &\geq \log \bar{Z}_{y,z} - s_y(y-\delta) - t_z(z-\delta) \\
&= -\mathcal{J}(y,z) + s_y\delta + t_z\delta \\
&\geq -\inf_{x\in U} \mathcal{J}(1,x) + \varepsilon + s_y\delta + t_z\delta,
\end{aligned}$$

where the last inequality invokes (2.41). Since ε and δ are arbitrary, this implies the desired lower bound in (2.40).

Step 3. We claim that the denominator of (3.2.14) satisfies

$$\lim_{n\to\infty} \frac{1}{n} \log \left|\mathbb{B}_V^n\right| = -\sup_{s<0} \left\{ s - \log\left(\int_{D_V} e^{sV(x)} dx\right)\right\}. \tag{2.42}$$

Since $\mathbb{B}_V^n = \mathbb{B}_{2,V}^n[\mathbb{R}_+]$, applying (2.38) with $F = \mathbb{R}_+$ and (2.40) with $U = (0,\infty)$, we obtain

$$\lim_{n\to\infty} \frac{1}{n} \log |\mathbb{B}_V^n| = -\inf_{x\in\mathbb{R}_+} \mathcal{J}(1,x).$$

The claim then follows from (2.35) of Lemma 2.6.8.

Finally, the combination of (2.42), (2.38) and (2.40) with (3.2.14) yields an LDP for

$\{\frac{1}{n}\sum_{i=1}^{n}(X_i^n)^2\}_{n\in\mathbb{N}}$ at speed n with rate function

$$\mathcal{J}(1,x) - \sup_{s<0}\left\{s - \log\left(\int_{D_V} e^{sV(x)}dx\right)\right\}.$$

The proposition then follows by an application of the contraction principle to the mapping $x \mapsto \sqrt{x}$. □

Remark 2.6.10. The relation (2.42) provides the asymptotic logarithmic volume of an Orlicz ball. A careful inspection of the argument shows that this limit result does not require the super-quadratic restriction on the Orlicz function; it is the LDP for the norms that uses this restriction. Indeed, let V and $\tilde{V}$ be Orlicz functions such that $V(x)/\tilde{V}(x) \to \infty$ as $x \to \infty$, let $\tilde{\mathcal{J}}$ be defined as in (2.36). and for any measurable set $A \subset \mathbb{R}$ define $\mathbb{B}_{\tilde{V}}^n[A] := \{x \in \mathbb{R}^n : \sum_{i=1}^n \tilde{V}(x_i) \in nA\}$. Then, we see from Step 1 in the proof of Proposition 2.6.7 that the following more general inequality holds:

$$\limsup_{n\to\infty}\frac{1}{n}\log\left|\mathbb{B}_V^n[A]\cap\mathbb{B}_{\tilde{V}}^n[B]\right| \leq -\inf_{x\in A, y\in B}\tilde{\mathcal{J}}(x,y), \qquad \text{for closed sets } A, B \subset \mathbb{R}_+.$$

Remark 2.6.11. After the first version of this paper was posted on arXiv, the asymptotic logarithmic volume of an Orlicz ball was also obtained in [42], where they also established further refined estimates using ideas from sharp large deviations estimates (see, e.g., [10,58]).

2.6.2.2 Generalized Orlicz balls

Suppose that $\mathbb{B}^n_{V_1,\dots,V_n}$ is close to a symmetric Orlicz ball in the following sense: there exists an Orlicz function $\bar{V} : \mathbb{R}_+ \to \mathbb{R}_+$ such that as $n \to \infty$,

$$\frac{1}{n} \log \frac{\left|\mathbb{B}^n_{\bar{V}} \,\Delta\, \mathbb{B}^n_{V_1,\dots,V_n}\right|}{\left|\mathbb{B}^n_{\bar{V}} \cup \mathbb{B}^n_{V_1,\dots,V_n}\right|} \to -\infty, \tag{2.43}$$

where $\mathbb{B}^n_{\bar{V}}$ is the Orlicz ball as defined in (2.29), and for any sets $A, B \subset \mathbb{R}^n$, $A \Delta B = [A \setminus B] \cup [B \setminus A]$ represents the symmetric difference of A and B.

Lemma 2.6.12 (Generalized Orlicz)**.** *Fix Orlicz functions $\{V_i\}_{i\in\mathbb{N}}$ and $\bar{V}$ that satisfy (2.43), and are uniformly superquadratic in the sense that there exists a constant $c_V \in (0, \infty)$ such that for $x > c_V$, $V_i(x) > x^2$ for $i \in \mathbb{N}$ and $\bar{V}(x) > x^2$. Suppose $X^{(n)} \sim \mathrm{Unif}(\mathbb{B}^n_{V_1,\dots,V_n})$. Then, $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption 2.1.3* with $J_{X,\bar{V}}$, defined as in (2.34), but with V replaced with $\bar{V}$.*

Proof. Let $\bar{X}^{(n)} \sim \mathrm{Unif}(\mathbb{B}^n_{\bar{V}})$, $U^{(n)} \sim \mathrm{Unif}(\mathbb{B}^n_{\bar{V}} \cup \mathbb{B}^n_{V_1,\dots,V_n})$ and $X^{(n)} \sim \mathrm{Unif}(\mathbb{B}^n_{V_1,\dots,V_n})$ be independent random vectors defined on a common probability space. Define

$$\begin{aligned} W^{(n)} &:= U^{(n)} \mathbf{1}_{\{U^{(n)} \in \mathbb{B}^n_{V_1,\dots,V_n}\}} + X^{(n)} \mathbf{1}_{\{U^{(n)} \notin \mathbb{B}^n_{V_1,\dots,V_n}\}}, \\ \bar{W}^{(n)} &:= U^{(n)} \mathbf{1}_{\{U^{(n)} \in \mathbb{B}^n_{\bar{V}}\}} + \bar{X}^{(n)} \mathbf{1}_{\{U^{(n)} \notin \mathbb{B}^n_{\bar{V}}\}}. \end{aligned}$$

First note that conditioned on $\{U^{(n)} \in \mathbb{B}^n_{V_1,\dots,V_n}\}$, $U^{(n)}$ has the same distribution as $X^{(n)}$. By definition, $X^{(n)}$ is supported on $\mathbb{B}^n_{V_1,\dots,V_n}$. For a measurable set $A \subset \mathbb{B}^n_{V_1,\dots,V_n}$, we have

$$\begin{aligned} \mathbb{P}\left(W^{(n)} \in A\right) = \mathbb{P}&\left(W^{(n)} \in A \middle| U^{(n)} \in \mathbb{B}^n_{V_1,\dots,V_n}\right) \mathbb{P}\left(U^{(n)} \in \mathbb{B}^n_{V_1,\dots,V_n}\right) \\ &+ \mathbb{P}\left(W^{(n)} \in A \middle| U^{(n)} \notin \mathbb{B}^n_{V_1,\dots,V_n}\right) \mathbb{P}\left(U^{(n)} \notin \mathbb{B}^n_{V_1,\dots,V_n}\right) \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{P}\left(U^{(n)} \in A \middle| U^{(n)} \in \mathbb{B}^n_{V_1,\dots,V_n}\right) \mathbb{P}\left(U^{(n)} \in \mathbb{B}^n_{V_1,\dots,V_n}\right) \\
&\qquad + \mathbb{P}\left(X^{(n)} \in A \middle| U^{(n)} \notin \mathbb{B}^n_{V_1,\dots,V_n}\right) \mathbb{P}\left(U^{(n)} \notin \mathbb{B}^n_{V_1,\dots,V_n}\right) \\
&= \mathbb{P}\left(X^{(n)} \in A\right) \mathbb{P}\left(U^{(n)} \in \mathbb{B}^n_{V_1,\dots,V_n}\right) + \mathbb{P}\left(X^{(n)} \in A\right) \mathbb{P}\left(U^{(n)} \notin \mathbb{B}^n_{V_1,\dots,V_n}\right) \\
&= \mathbb{P}\left(X^{(n)} \in A\right).
\end{aligned}
$$

Hence, $W^{(n)} \overset{(\mathrm{d})}{=} X^{(n)}$. A similar argument can be used to show that $\bar{W}^{(n)} \overset{(\mathrm{d})}{=} \bar{X}^{(n)}$, thus providing a useful coupling between the uniform measures on $\mathbb{B}^n_{V_1,\dots,V_n}$ and $\mathbb{B}^n_{\bar{V}}$. Due to the contraction principle applied to $x \mapsto \sqrt{x}$, Proposition 2.6.7 and Remark 1.1.6, it suffices to show that $\{\|W^{(n)}\|_2^2/n\}_{n\in\mathbb{N}}$ is exponentially equivalent (see Definition 1.1.5) to $\{\|\bar{W}^{(n)}\|_2^2/n\}_{n\in\mathbb{N}}$.

To this end, let $\kappa := \sqrt{c_V^2 + 1}$, where c_V is the positive finite constant stated in the lemma, and note that $\mathbb{B}^n_{V_1,\dots,V_n} \subset \kappa\sqrt{n}\mathbb{B}^n_2$ and $\mathbb{B}^n_{\bar{V}} \subset \kappa\sqrt{n}\mathbb{B}^n_2$ since for $x \in \mathbb{B}^n_{V_1,\dots,V_n}$, $\sum_{i=1}^n x_i^2 \le \sum_{i=1}^n (c_V^2 + V_i(x_i)) \le (c_V^2+1)n = \kappa^2 n$, where the last inequality uses (2.30), and for $x \in \mathbb{B}^n_{\bar{V}}$, $\sum_{i=1}^n x_i^2 \le \sum_{i=1}^n (c_V^2 + \bar{V}(x_i)) \le (c_V^2+1)n = \kappa^2 n$, where the last inequality uses (2.29). Thus,

$$
\left| \tfrac{\|\bar{W}^{(n)}\|_2^2}{n} - \tfrac{\|W^{(n)}\|_2^2}{n} \right| = \frac{1}{n} \left| \|\bar{W}^{(n)}\|_2^2 - \|W^{(n)}\|_2^2 \right| \le \kappa \mathbf{1}_{\{U^{(n)} \notin (\mathbb{B}^n_{\bar{V}} \cap \mathbb{B}^n_{V_1,\dots,V_n})\}}.
$$

Therefore, for every $\delta > 0$,

$$
\begin{aligned}
\frac{1}{n}\log\mathbb{P}\left(\left|\|\bar{W}^{(n)}\|_2^2/n - \|W^{(n)}\|_2^2/n\right| > \delta\right) &\le \frac{1}{n}\log\mathbb{P}\left(\kappa\mathbf{1}_{\{U^{(n)} \notin (\mathbb{B}^n_{\bar{V}} \cap \mathbb{B}^n_{V_1,\dots,V_n})\}} > 0\right) \\
&= \frac{1}{n}\log\mathbb{P}\left(U^{(n)} \in (\mathbb{B}^n_{\bar{V}} \,\Delta\, \mathbb{B}^n_{V_1,\dots,V_n})\right) \\
&= \frac{1}{n}\log\frac{\left|\mathbb{B}^n_{\bar{V}} \,\Delta\, \mathbb{B}^n_{V_1,\dots,V_n}\right|}{\left|\mathbb{B}^n_{\bar{V}} \cup \mathbb{B}^n_{V_1,\dots,V_n}\right|},
\end{aligned}
$$

which converges to $-\infty$ as $n \to \infty$ by (2.43). This establishes the desired exponential equivalence and thus completes the proof. □

2.7 Proofs of LDPs for the random projections

In this section, we prove the main large deviation results, namely, Theorems 2.3.3, 2.4.1 and 2.5.1, which deal with the constant, sublinear and linear regimes, respectively. At many points, we will refer to certain properties of the top row of $\mathbf{A}_{n,k_n}$, which we first establish in Section 2.7.1. Sections 2.7.2, 2.7.3 and 2.7.4 consider the regimes of $\{k_n\}$ constant, sublinear and linear, respectively. Throughout, let $\zeta_1, \zeta_2, \ldots,$ denote a sequence of i.i.d. standard Gaussian random variables, let $\zeta^{(n)} := (\zeta_1, \ldots, \zeta_n) \in \mathbb{R}^n$, let $e_1 := (1, 0, \ldots, 0) \in \mathbb{R}^n$.

2.7.1 The top row of $\mathbf{A}_{n,k_n}$

Lemma 2.7.1. *Fix $k, n \in \mathbb{N}$ such that $k \le n$. Then the following relation holds:*

$$\mathbf{A}_{n,k}(1, \cdot\,) = (\mathbf{A}_{n,k}(1,1), \ldots, \mathbf{A}_{n,k}(1,k)) \stackrel{(\mathrm{d})}{=} \frac{(\zeta_1, \ldots, \zeta_k)}{\|\zeta^{(n)}\|_2}. \tag{2.44}$$

Proof. Let $\mathbf{O}_n$ be a random $n \times n$ orthogonal matrix (i.e., sampled from the normalized Haar measure on the group of $n \times n$ orthogonal matrices). Let $I_{n,k}$ be the $n \times k$ matrix of ones on the diagonal and zeros elsewhere. Note that

$$\mathbf{A}_{n,k}^T \stackrel{(\mathrm{d})}{=} I_{n,k}^T \mathbf{O}_n,$$

which implies that $\mathbf{A}_{n,k}(1, \cdot\,)$ is equal in distribution to the vector of the first k elements in the top row of $\mathbf{O}_n$. The marginal distribution of the top row of $\mathbf{O}_n$ is the uniform measure on the unit sphere of $\mathbb{R}^n$, which establishes the identity (2.44), due to the classical fact that $\zeta^{(n)}/\|\zeta^{(n)}\|_2$ is uniformly distributed on the unit sphere in $\mathbb{R}^n$ (see, e.g., [62]). ☐

Lemma 2.7.2. *Fix $n \in \mathbb{N}$ and $k \leq n$. Suppose $X^{(n)}$ is an n-dimensional random vector independent of $\mathbf{A}_{n,k}$. Then the following relation holds:*

$$\left(\mathbf{A}_{n,k}^T \tfrac{X^{(n)}}{\|X^{(n)}\|_2}, \|X^{(n)}\|_2\right) \stackrel{(\mathrm{d})}{=} \left(\mathbf{A}_{n,k}^T e_1, \|X^{(n)}\|_2\right).$$

Proof. As in Lemma 6.3 of [34], which considered the case $k = 1$, this result is easily deduced from the fact that the distribution of $\mathbf{A}_{n,k}$ is invariant under orthogonal transformations and independent of $X^{(n)}$. In particular, fix $n \in \mathbb{N}$ and given any $n \times n$ orthogonal matrix O_n and $x \in \mathbb{R}^n$, we have

$$\left(\mathbf{A}_{n,k}^T \frac{x}{\|x\|_2}, \|x\|_2\right) \stackrel{(\mathrm{d})}{=} \left(\mathbf{A}_{n,k}^T O_n^{-1} \frac{x}{\|x\|_2}, \|x\|_2\right).$$

Now, given $e_1 \in S^{n-1}$, for any $y \in S^{n-1}$ let $\mathcal{R}(y) \in \mathbb{V}_{n,n}$ be the unique orthogonal matrix O_n such that $O_n^{-1} y = e_1$. It is easy to see that $\mathcal{R} : S^{n-1} \mapsto \mathbb{V}_{n,n}$ is a measurable map. Then, for any $x \in \mathbb{R}^n$, substituting $O_n = \mathcal{R}(x/\|x\|_2)$ in the last display it follows that

$$\left(\mathbf{A}_{n,k}^T \frac{x}{\|x\|_2}, \|x\|_2\right) \stackrel{(\mathrm{d})}{=} \left(\mathbf{A}_{n,k}^T \left(\mathcal{R}\left(\frac{x}{\|x\|_2}\right)\right)^{-1} \frac{x}{\|x\|_2}, \|x\|_2\right) = \left(\mathbf{A}_{n,k}^T e_1, \|x\|_2\right).$$

Since $\mathbf{A}_{n,k}^T$ is independent of $X^{(n)}$, the above relation holds when x is replaced with $X^{(n)}$. ☐

In the settings in which $\{k_n\}_{n\in\mathbb{N}}$ grows, we will first analyze the empirical measure of the k_n elements in the top row of $\sqrt{n}\mathbf{A}_{n,k_n}$. That is, let

$$\hat{\mu}^n_{\mathbf{A}} := \frac{1}{k_n}\sum_{j=1}^{k_n} \delta_{\sqrt{n}\mathbf{A}_{n,k_n}(1,j)}, \quad n \in \mathbb{N}. \tag{2.45}$$

Recall that $\mathcal{P}(\mathbb{R})$ is always equipped with the weak topology unless otherwise stated.

Lemma 2.7.3. *For $n \in \mathbb{N}$, let $X^{(n)}$ be independent of $\mathbf{A}_{n,k_n}$ and recall the definition of L^n given in (2.1). Then we have*

$$L^n(\cdot) \overset{(\mathrm{d})}{=} \hat{\mu}^n_{\mathbf{A}}(\cdot \times \sqrt{n}/\|X^{(n)}\|_2).$$

Moreover, the map $\mathcal{P}(\mathbb{R}) \times (0,\infty) \ni (\nu, c) \mapsto \nu(\cdot \times c^{-1}) \in \mathcal{P}(\mathbb{R})$ is continuous.

Proof. For each $n \in \mathbb{N}$, using the definition of L^n from (2.1) and Lemma 2.7.2, we have

$$\begin{aligned} L^n(\cdot) = \frac{1}{k_n}\sum_{j=1}^{k_n} \delta_{\|X^{(n)}\|_2\left(\mathbf{A}^T_{n,k_n}\frac{X^{(n)}}{\|X^{(n)}\|_2}\right)_j}(\cdot) &\overset{(\mathrm{d})}{=} \frac{1}{k_n}\sum_{j=1}^{k_n} \delta_{\frac{\|X^{(n)}\|_2}{\sqrt{n}}\sqrt{n}\left(\mathbf{A}^T_{n,k_n}e_1\right)_j}(\cdot) \\ &= \frac{1}{k_n}\sum_{j=1}^{k_n} \delta_{\sqrt{n}\mathbf{A}_{n,k_n}(1,j)}(\cdot \times \sqrt{n}/\|X^{(n)}\|_2), \end{aligned}$$

from which the first assertion of the lemma follows by (2.45). Continuity of the map $(\nu, c) \mapsto \nu(\cdot \times c^{-1})$ follows if and only if as $n \to \infty$ a sequence of random variables $\{B_n\}_{n\in\mathbb{N}}$ converging in distribution to B and a sequence of constants $\{c_n\}$ converging to c implies that $c_n B_n$ converges to cB in distribution as $n \to \infty$, which follows from Slutsky's theorem [53, Theorem 13.18]. $\square$

Lemma 2.7.4. *Suppose $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ and $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfy LDPs, both at speed s_n, and with GRFs $\mathbb{I}_1 : \mathcal{P}_q(\mathbb{R}) \to [0,\infty]$ and $\mathbb{I}_2 : \mathbb{R}_+ \to [0,\infty]$, respectively. Then $\{L^n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF defined by*

$$\mathbb{I}(\mu) = \inf_{\nu\in\mathcal{P}_q(\mathbb{R}), c\in\mathbb{R}_+} \left\{\mathbb{I}_1(\nu) + \mathbb{I}_2(c) : \mu = \nu(\cdot \times c^{-1})\right\}, \quad \mu \in \mathcal{P}_q(\mathbb{R}).$$

Proof. By the independence of $\mathbf{A}_{n,k_n}$ and $X^{(n)}$, $\{\hat{\mu}^n_{\mathbf{A}}, \|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF

$$\mathcal{P}_q(\mathbb{R}) \times \mathbb{R}_+ \ni (\nu, c) \mapsto \mathbb{I}_1(\nu) + \mathbb{I}_2(c) \in [0, \infty].$$

The claim follows on applying the contraction principle to the mapping $F : \mathcal{P}_q(\mathbb{R}) \times \mathbb{R}_+ \to \mathcal{P}_q(\mathbb{R})$ defined by $F(\nu, c) = \nu(\cdot \times c^{-1})$ □

Lemma 2.7.5. *Let $\{\zeta_j\}_{j\in\mathbb{N}}$ be i.i.d. $\mathcal{N}(0,1)$ random variables, $\zeta^{(n)} = (\zeta_1, \ldots, \zeta_n)$ and consider the sequence*

$$\nu_n := \frac{1}{k_n} \sum_{j=1}^{k_n} \delta_{\zeta_j}, \quad n \in \mathbb{N}. \tag{2.46}$$

Then $\{\nu_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\mathbb{R})$ with respect to the weak topology, at speed k_n, with GRF $H(\cdot|\gamma_1)$. Moreover, for a sequence $\{s_n\}_{n\in\mathbb{N}}$ such that $s_n \ll k_n$, $\{\nu_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\mathbb{R})$ with respect to the weak topology at speed s_n with GRF

$$\chi_{\gamma_1}(\nu) := \begin{cases} 0 & \text{if } \nu = \gamma_1 \\ +\infty & \text{else} \end{cases}, \quad \nu \in \mathcal{P}(\mathbb{R}). \tag{2.47}$$

Proof. The first claim is a direct conclusion of Sanov's theorem [21, Theorem 2.1.10] while the second claim follows since $H(\cdot|\gamma_1) : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ is convex with a unique minimizer at γ_1, as in the observation of Remark 1.1.3. □

We now document an elementary observation since it will be used multiple times in the sequel.

Remark 2.7.6. Let $\{\zeta_j\}_{j\in\mathbb{N}}$ be i.i.d. $\mathcal{N}(0,1)$ random variables, let $\{k_n\}_{n\in\mathbb{N}}$ be a sequence of positive integers that converges to infinity and let $C_n := \frac{1}{k_n}\sum_{j=1}^{k_n} \zeta_j^2$, $n \in \mathbb{N}$. It is easy to see that the log moment generating function of ζ_1^2 is given by $-\frac{1}{2}\log(1-2t)$ for

$t < 1/2$ and infinity, otherwise. Hence, it is an immediate consequence of Cramér's theorem [21, Theorem 2.2.1] that $\{C_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ at speed k_n with GRF $J_{\zeta^2} : \mathbb{R} \to [0,\infty]$ of the form

$$J_{\zeta^2}(t) := \sup_{s\in\mathbb{R}} \left\{ st + \frac{1}{2}\log(1-2s) \right\} = \begin{cases} \frac{t-1}{2} - \frac{1}{2}\log t, & t > 0, \\ +\infty, & t \leq 0. \end{cases} \tag{2.48}$$

2.7.2 Proof of the LDP for random projections in the constant regime

This section is devoted to the proof of Theorem 2.3.3. Throughout this section, fix $k \in \mathbb{N}$ and suppose $k_n = k$ for each $n \in \mathbb{N}$. We first give the main idea behind the proof. Due to Lemma 2.7.2 we have

$$\frac{1}{\sqrt{n}}\mathbf{A}_{n,k}^T X^{(n)} \overset{(\mathrm{d})}{=} \frac{1}{\sqrt{n}}\mathbf{A}_{n,k}^T e_1 \|X^{(n)}\|_2 = \mathbf{A}_{n,k}(1,\cdot)\frac{\|X^{(n)}\|_2}{\sqrt{n}} \tag{2.49}$$

where $\{\mathbf{A}_{n,k}(1,\cdot)\}_{n\in\mathbb{N}}$ is independent of $\{X^{(n)}\}_{n\in\mathbb{N}}$. Thus, the question essentially reduces to understanding the following: suppose two independent sequences of random vectors $\{V_n\}_{n\in\mathbb{N}}$ and $\{W_n\}_{n\in\mathbb{N}}$ (with at least one-being real-valued) satisfy LDPs at speeds $\{\beta_n\}_{n\in\mathbb{N}}$, and $\{\gamma_n\}_{n\in\mathbb{N}}$ with GRFs J_V and J_W, respectively. Then we want to understand when an LDP with a non-trivial rate function can be deduced for the product sequence $Z_n := V_n W_n$, $n \in \mathbb{N}$. When $\beta_n = \gamma_n$, then this is a simple application of the contraction principle. We will see that this is the case when Assumption 2.1.3* holds, that is, when $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n, by showing that then $\{\mathbf{A}_{n,k}(1,\cdot)\}_{n\in\mathbb{N}}$ also satisfies an LDP at speed n. On the other hand, if $\beta_n \ll \gamma_n$ or $\beta_n \gg \gamma_n$, then the idea is to find sequences $\{b_n\}_{n\in\mathbb{N}}$ and $\{s_n\}_{n\in\mathbb{N}}$ such that $\{b_n V_n\}_{n\in\mathbb{N}}$ and $\{b_n^{-1} W_n\}_{n\in\mathbb{N}}$ both satisfy non-trivial LDPs at the same speed s_n, and once again apply the contraction principle.

This falls into the case of Assumption 2.3.1, which states that $\{b_n\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$, with $b_n = \sqrt{s_n/n}$, satisfies a non-trivial LDP at speed s_n, and thus the proof in this case follows by showing that $\{b_n^{-1}\mathbf{A}_{n,k}(1,\cdot)\}_{n\in\mathbb{N}}$ also satisfies a non-trivial LDP at speed s_n.

Proof of Theorem 2.3.3. Suppose Assumption 2.3.1 holds with sequence $\{s_n\}_{n\in\mathbb{N}}$ and GRF J_X. From Lemma 2.7.2 and (2.44), denoting $\zeta^{(n)} := (\zeta_1,\ldots,\zeta_n)$ with $\{\zeta_i\}_{i\in\mathbb{N}}$ i.i.d. $\mathcal{N}(0,1)$ random variables, we have

$$\frac{1}{\sqrt{n}}\mathbf{A}_{n,k}^T X^{(n)} \overset{(d)}{=} \mathbf{A}_{n,k}(1,\cdot)\frac{\left\|X^{(n)}\right\|_2}{\sqrt{n}} \overset{(d)}{=} \frac{\zeta^{(k)}/\sqrt{s_n}}{\left\|\zeta^{(n)}\right\|_2/\sqrt{n}}\frac{\sqrt{s_n}\left\|X^{(n)}\right\|_2}{n}, \tag{2.50}$$

Now, consider the sequence of vectors

$$R_n := \left(\frac{\zeta^{(k)}}{\sqrt{s_n}}, \frac{\sqrt{n}}{\left\|\zeta^{(n)}\right\|_2}, \frac{\sqrt{s_n}\left\|X^{(n)}\right\|_2}{n}\right), \quad n \in \mathbb{N},$$

which are almost surely well-defined. Then $\zeta_i/\sqrt{s_n}$ is a $\mathcal{N}(0,1/s_n)$ random variable with $s_n \to \infty$ as $n \to \infty$. Hence, by the Gärtner-Ellis theorem, the sequence $\{\zeta^{(k)}/\sqrt{s_n}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^k$ at speed s_n with GRF $x \mapsto \|x\|^2/2$. Assumption 2.3.1 implies that $\{\sqrt{s_n}\|X^{(n)}\|_2/n\}_{n\in\mathbb{N}}$ satisfies an LDP, also at speed s_n, and with GRF J_X. Finally, by Remark 2.7.6, the contraction principle applied to $x \mapsto 1/\sqrt{x}$ and the strong law of large numbers, $\{\sqrt{n}/\|\zeta^{(n)}\|_2\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n and converges almost surely to 1, which implies that the associated GRF has a unique minimum at 1.

Together with the independence of $\{\zeta_i\}_{i=1,\ldots,k}$ and $X^{(n)}$, and Lemma 1.1.9 (with $U_n = \zeta^{(k)}/\sqrt{s_n}$, $V_n = \sqrt{s_n}\|X^{(n)}\|_2/n$, $W_n = \sqrt{n}/\|\zeta^{(n)}\|_2$ and $m = 1$), this implies that R_n satisfies an LDP with GRF given by $\mathbb{R}^k \times \mathbb{R}_+ \times \mathbb{R}_+ \ni (r_1, r_2, r_3) \mapsto \frac{\|r_1\|_2^2}{2} + J_X(r_3)$ if $r_2 = 1$ (and $+\infty$ otherwise). Combining this with (2.50) and the contraction principle for the continuous mapping $\mathbb{R}^k \times \mathbb{R}_+^2 \ni (r_1, r_2, r_3) \mapsto r_1 r_2 r_3 \in \mathbb{R}^k$, it follows that the sequence of k-dimensional vectors, $\{n^{-1/2}\mathbf{A}_{n,k}^T X^{(n)}\}_{n\in\mathbb{N}}$, satisfies an LDP at speed s_n with GRF

$$
\begin{aligned}
I_{\mathbf{A}X,k}(x) &= \inf\left\{\frac{\|r_1\|_2^2}{2} + J_X(r_3) : r_1 \in \mathbb{R}^k, r_3 > 0, x = r_1 r_3\right\} \\
&= \inf_{c>0}\left\{J_X\left(\frac{\|x\|_2}{c}\right) + \frac{c^2}{2}\right\},
\end{aligned}
$$

which coincides with the expression given in (2.12). This completes the proof of Theorem 2.3.3. ◻

2.7.3 Proof of the empirical measure LDP in the sublinear regime $1 \ll k_n \ll n$

We now present the proof of Theorem 2.4.1. We first start with an auxiliary result.

Proposition 2.7.7. *Fix $q \in [1,2)$. Suppose $\{k_n\}_{n\in\mathbb{N}}$ grows sublinearly. Then, $\{\hat{\mu}_{\mathbf{A}}^n\}_{n\in\mathbb{N}}$ defined in (2.45) satisfies an LDP in $\mathcal{P}_q(\mathbb{R})$ at speed k_n with GRF $H(\cdot|\gamma_1) : \mathcal{P}(\mathbb{R}) \to [0,\infty]$.*

Proof. Recall that $\{\zeta_j\}_{j\in\mathbb{N}}$ are i.i.d. $\mathcal{N}(0,1)$ random variables, $\zeta^{(n)} = (\zeta_1, \dots, \zeta_n)$. Due to (2.45) and Lemma 2.7.1, $\{\hat{\mu}_{\mathbf{A}}^n\}_{n\in\mathbb{N}} \overset{(d)}{=} \{\tilde{\nu}_n\}_{n\in\mathbb{N}}$, where

$$
\tilde{\nu}_n := \frac{1}{k_n}\sum_{j=1}^{k_n} \delta_{\sqrt{n}\zeta_j/\|\zeta^{(n)}\|_2}, \quad n \in \mathbb{N}.
$$

We now claim (and justify below) that $\{\nu_n\}_{n\in\mathbb{N}}$ defined in Lemma 2.7.5 is exponentially equivalent (at speed k_n) to the sequence $\{\tilde{\nu}_n\}_{n\in\mathbb{N}}$ defined in (2.46).

To prove the claim, let $z_n := \sqrt{n}/\|\zeta^{(n)}\|_2$ and let d_{BL} denote the bounded-Lipschitz metric (which metrizes weak convergence). Then, letting $\mathrm{BL}(\mathbb{R})$ denote the space of bounded Lipschitz functions $f : \mathbb{R} \to \mathbb{R}$ with Lipschitz constant 1,

$$d_{\mathrm{BL}}(\nu_n, \tilde{\nu}_n) \leq \sup_{f \in \mathrm{BL}(\mathbb{R})} \frac{1}{k_n} \sum_{j=1}^{k_n} |f(\zeta_j) - f(\zeta_j z_n)| \leq \frac{1}{k_n} \sum_{j=1}^{k_n} |\zeta_j - \zeta_j z_n| = |z_n - 1| \cdot \frac{1}{k_n} \sum_{j=1}^{k_n} |\zeta_j|.$$

Hence, for $\delta, \epsilon > 0$, we have

$$\mathbb{P}(d_{\mathrm{BL}}(\nu_n, \tilde{\nu}_n) > \delta) \leq \mathbb{P}\left(|z_n - 1| \cdot \frac{1}{k_n} \sum_{j=1}^{k_n} |\zeta_j| > \delta \right) \leq \mathbb{P}\left(\frac{1}{k_n} \sum_{j=1}^{k_n} |\zeta_j| > \frac{\delta}{\epsilon} \right) + \mathbb{P}(|z_n - 1| > \epsilon).$$

Since ζ_1 has a finite exponential moment, Cramér's theorem [21, Theorem 2.2.1] implies

$$\lim_{n \to \infty} \frac{1}{k_n} \log \mathbb{P}\left(\frac{1}{k_n} \sum_{j=1}^{k_n} |\zeta_j| > \delta/\epsilon \right) = -\mathcal{I}(\delta/\epsilon),$$

for some convex and superlinear rate function $\mathcal{I}$. Also, by Cramér's theorem for $\sum_{i=1}^{n} \zeta_i^2/n$, the continuity of $x \mapsto 1/\sqrt{x}$ on $(0, \infty)$ and the contraction principle, the sequence $\{z_n\}_{n \in \mathbb{N}}$ satisfies an LDP at speed n. Hence, due to the sublinear growth of k_n, we have

$$\lim_{n \to \infty} \frac{1}{k_n} \log \mathbb{P}(|z_n - 1| > \epsilon) = -\infty.$$

Combining the last three displays, we find that

$$\lim_{n \to \infty} \frac{1}{k_n} \log \mathbb{P}(d_{\mathrm{BL}}(\nu_n, \tilde{\nu}_n) > \delta) \leq -\mathcal{I}(\delta/\epsilon).$$

The claim of exponential equivalence follows on sending $\epsilon \to 0$, due to the superlinearity of $\mathcal{I}$.

In light of Remark 1.1.6, the last observation taken together with Lemma 2.7.5 implies that $\{\hat{\mu}^n_{\mathbf{A}}\}_{n \in \mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\mathbb{R})$ at speed k_n with GRF $H(\cdot|\gamma_1)$. In order to strengthen the LDP for $\{\hat{\mu}^n_{\mathbf{A}}\}_{n \in \mathbb{N}}$, by [21, Corollary 4.2.6], it suffices to show that $\{\hat{\mu}^n_{\mathbf{A}}\}_{n \in \mathbb{N}}$

is exponentially tight in $\mathcal{P}_q(\mathbb{R})$. Since $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}} \overset{(d)}{=} \{\tilde{\nu}_n\}_{n\in\mathbb{N}}$ and for each $j > 0$, the set $K_{2,j}$ defined in (1.4) is compact with respect to $\mathcal{P}_q(\mathbb{R})$ due to Lemma 1.1.1, if suffices to show that for $\varepsilon > 0$, there exists $M < \infty$ such that

$$\lim_{n\to\infty} \frac{1}{k_n} \log \mathbb{P}(\tilde{\nu}_n \in K^c_{2,M}) = \lim_{n\to\infty} \frac{1}{k_n} \log \mathbb{P}(M_2(\tilde{\nu}_n) > M) < -\varepsilon. \tag{2.51}$$

Now note that $\{M_2(\tilde{\nu}_n) = \frac{n}{k_n}\sum_{j=1}^{k_n} \frac{\zeta_j^2}{\|\zeta^{(n)}\|_2}\}_{n\in\mathbb{N}}$ and by Remark 2.7.6, $\{\|\zeta^{(k_n)}\|_2^2/k_n\}_{n\in\mathbb{N}}$ and $\{\|\zeta^{(n)}\|_2^2/n\}_{n\in\mathbb{N}}$ satisfy LDPs at speed k_n and n, respectively, both with the same GRF $J_{\zeta^2}(x)$ defined in (2.48). Thus, $\{n/\|\zeta^{(n)}\|_2^2\}_{n\in\mathbb{N}}$ also satisfies an LDP at speed n by the contraction principle. Together with Lemma 1.1.9 and the fact that J_{ζ^2} has a unique minimizer at 1, this implies the sequence $\{M_2(\tilde{\nu}_n)\}_{n\in\mathbb{N}}$ satisfies an LDP at speed k_n with GRF J_{ζ^2}. Now, for $\varepsilon > 0$, pick M such that $J_{\zeta^2}(M) > \varepsilon$. We then obtain

$$\lim_{n\to\infty} \frac{1}{k_n} \log \mathbb{P}(M_2(\tilde{\nu}_n) > M) \le -J_{\zeta^2}(M) < -\varepsilon,$$

which proves (2.51), and thus concludes the proof of the lemma. □

Proof of Theorem 2.4.1. By Assumption 2.1.3, $\|X^{(n)}\|_2/\sqrt{n}$ satisfies an LDP at speed s_n with GRF J_X and when $s_n \gg k_n$ by Remark 1.1.3 (and the additional assumption that J_X has a unique minimizer m), $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ also satisfies an LDP at speed k_n with (degenerate) rate function χ_m. Also, by Proposition 2.7.7 and Remark 1.1.3, $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed k_n with GRF $H(\cdot|\gamma_1)$ and, when $s_n \ll k_n$, $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ also satisfies an LDP at speed s_n with (degenerate) GRF χ_{γ_1} defined in Remark 1.1.3 (see also Lemma 2.7.5).

Combining the above observations with Lemma 2.7.4, we see that $\{L^n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $s_n \wedge k_n$ with GRF $\mathbb{I}_{L,k_n}$, where when $s_n \gg k_n$,

$$\begin{aligned}\mathbb{I}_{L,k_n}(\mu) &= \inf_{\nu\in\mathcal{P}_q(\mathbb{R}),c\in\mathbb{R}_+}\left\{H(\nu|\gamma_1) + \chi_m(c) : \mu = \nu(\cdot\times c^{-1})\right\}\\ &= H(\mu(\cdot\times m)|\gamma_1)\\ &= H(\mu|\gamma_1(\cdot\times m^{-1}))\\ &= H(\mu|\gamma_m),\end{aligned}$$

when $s_n = k_n$,

$$\begin{aligned}\mathbb{I}_{L,k_n}(\mu) &= \inf_{\nu\in\mathcal{P}_q(\mathbb{R}),c\in\mathbb{R}_+}\left\{H(\nu|\gamma_1) + J_X(c) : \mu = \nu(\cdot\times c^{-1})\right\}\\ &= \inf_{c\in\mathbb{R}_+}\left\{H(\mu|\gamma_c) + J_X(c)\right\},\end{aligned}$$

and when $s_n \ll k_n$,

$$\begin{aligned}\mathbb{I}_{L,k_n}(\mu) &= \inf_{\nu\in\mathcal{P}_q(\mathbb{R}),c\in\mathbb{R}_+}\left\{\chi_{\gamma_1}(\nu) + J_X(c) : \mu = \nu(\cdot\times c^{-1})\right\}\\ &= \begin{cases} J_X(c), & \mu = \gamma_c\\ +\infty, & \text{otherwise.}\end{cases}\end{aligned}$$

This proves Theorem 2.4.1. □

2.7.4 Proof of the empirical measure LDP in the linear regime $k_n \sim \lambda n$

Throughout this section, suppose $\{k_n\}_{n\in\mathbb{N}}$ grows linearly with rate $\lambda \in (0,1]$. As in the sublinear regime, we first analyze the sequence of empirical measures $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ of (2.45) as a precursor to the analysis of $\{L^n\}_{n\in\mathbb{N}}$ of (2.1).

Proposition 2.7.8. *Fix $q \in [1,2)$. Suppose k_n grows linearly with rate $\lambda \in (0,1]$. Then, the sequence $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}_q(\mathbb{R})$ at speed n with GRF $\mathcal{H}_\lambda$ of (2.4).*

The proof of the above result is deferred to the end of this section. Taking it as given, we now prove the main LDP in the linear regime.

Proof of Theorem 2.5.1. Suppose Assumption 2.1.3 is satisfied with corresponding sequence $\{s_n\}_{n\in\mathbb{N}}$. First, suppose $s_n = n$. Then, along with Proposition 2.7.8 and Lemma 2.7.4, this implies that $\{L^n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n with GRF $\hat{\mathbb{I}}_{L,\lambda}$, defined by

$$\begin{aligned}\hat{\mathbb{I}}_{L,\lambda}(\mu) &:= \inf_{\nu\in\mathcal{P}(\mathbb{R}),c\in\mathbb{R}_+}\left\{\mathcal{H}_\lambda(\nu) + J_X(c) : \mu(\cdot) = \nu(\,\cdot\times c^{-1})\right\},\\ &= \inf_{c\in\mathbb{R}_+}\left\{\mathcal{H}_\lambda(\mu(\cdot\times c)) + J_X(c)\right\}, \quad \mu \in \mathcal{P}(\mathbb{R}).\end{aligned}$$

In turn, since $\mathcal{H}_\lambda(\nu) = \infty$ if $M_2(\nu) > 1/\lambda$ and $M_2(\mu(\cdot\times c)) = M_2(\mu)/c^2$, this implies

$$\hat{\mathbb{I}}_{L,\lambda}(\mu) = \inf_{c>\sqrt{\lambda M_2(\mu)}}\left\{\mathcal{H}_\lambda(\mu(\cdot\times c)) + J_X(c)\right\}, \quad \mu \in \mathcal{P}(\mathbb{R}). \tag{2.52}$$

Now for $\mu \in \mathcal{P}(\mathbb{R})$ and $c > \sqrt{\lambda M_2(\mu)}$, the definitions of $\mathcal{H}_\lambda$ and h in (2.4) and (1.6), respectively, imply

$$\begin{aligned}\mathcal{H}_\lambda(\mu(\,\cdot\times c)) &= -\lambda\, h(\mu(\,\cdot\times c)) + \tfrac{\lambda}{2}\log(2\pi e) + \tfrac{1-\lambda}{2}\log\left(\tfrac{1-\lambda}{1-\lambda\, M_2(\mu(\,\cdot\times c))}\right)\\ &= -\lambda\, h(\mu) + \lambda\log(c) + \tfrac{\lambda}{2}\log(2\pi e) + \tfrac{1-\lambda}{2}\log\left(\tfrac{1-\lambda}{1-(\lambda/c^2)\, M_2(\mu)}\right)\\ &= \lambda\log(c) - \tfrac{1-\lambda}{2}\log\left(1 - \tfrac{\lambda M_2(\mu)}{c^2}\right) - \lambda h(\mu) + \tfrac{\lambda}{2}\log(2\pi e) + \tfrac{1-\lambda}{2}\log(1-\lambda).\end{aligned}$$

When substituted back into (2.52), this shows that for every $\mu \in \mathcal{P}(\mathbb{R})$, $\hat{\mathbb{I}}_{L,\lambda}(\mu) = \mathbb{I}_{L,\lambda}(\mu)$, with the latter defined as in (2.17).

Next, consider the case $s_n \ll n$. By Lemma 2.7.8, $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF $\mathcal{H}_\lambda$. Therefore, by Remark 1.1.3 and Remark 2.2.4, $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ also satisfies an LDP at speed s_n with GRF χ_{γ_1}. Then, by Assumption 2.1.3 and Lemma 2.7.4, $\{L^n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF

$$\inf_{\nu\in\mathcal{P}(\mathbb{R}),c\in\mathbb{R}_+}\left\{\chi_{\gamma_1}(\nu) + J_X(c) : \mu = \nu(\cdot \times c^{-1})\right\}$$
$$= \begin{cases} J_X(c), & \mu = \gamma_c, \\ +\infty, & \text{otherwise,} \end{cases}$$

which coincides with the expression for $I_{L,\lambda}(\mu)$ given in Theorem 2.5.1 for the case $s_n \ll n$. □

Remark 2.7.9. Note that the above proof in particular shows that, under Assumption 2.1.3 the rate function $\mathbb{I}_{L,\lambda}(\mu)$ from Theorem 2.5.1 in the linear regime in the case $s_n = n$, satisfies, for $\lambda > 0$,

$$\mathbb{I}_{L,\lambda}(\mu) = \inf_{c\in\mathbb{R}_+}\left\{\mathcal{H}_\lambda(\mu(\cdot \times c)) + J_X(c)\right\},$$

with $\mathcal{H}_\lambda$ defined as in (2.4).

The rest of this section is devoted to the proof of Proposition 2.7.8, which is broken down into the intermediate steps given by Lemmas 2.7.10 and 2.7.11 below.

Lemma 2.7.10. *Suppose $k_n/n \to \lambda \in (0,1]$ as $n \to \infty$, and $\{\zeta_j\}_{j\in\mathbb{N}}$ is an i.i.d. sequence with common law γ_1 (the standard normal distribution). If $\lambda \in (0,1)$, then the sequence*

$$\left(\frac{1}{k_n}\sum_{j=1}^{k_n}\delta_{\zeta_j}, \quad \frac{1}{n-k_n}\sum_{j=k_n+1}^{n}\delta_{\zeta_j}, \quad \frac{1}{n}\sum_{j=1}^{n}\zeta_j^2\right), \quad n \in \mathbb{N}, \tag{2.53}$$

satisfies an LDP in $[\mathcal{P}(\mathbb{R})]^2 \times \mathbb{R}_+$ *at speed* n *with GRF* $I_{1,\lambda}$ *defined by*

$$I_{1,\lambda}(\mu,\nu,s) := \lambda\, H(\mu|\gamma_1) + (1-\lambda)\, H(\nu|\gamma_1) + \tfrac{1}{2}\left[s - \lambda\, M_2(\mu) - (1-\lambda)\, M_2(\nu)\right],$$

if $\lambda\, M_2(\mu) + (1-\lambda)\, M_2(\nu) \le s$, *and* $I_{1,\lambda}(\mu,\nu,s) := +\infty$ *otherwise. On the other hand, if* $\lambda = 1$ *then the sequence*

$$\left(\frac{1}{k_n}\sum_{j=1}^{k_n}\delta_{\zeta_j}\,, \quad \frac{1}{n}\sum_{j=1}^{n}\zeta_j^2\right), \quad n \in \mathbb{N},$$

satisfies an LDP in $\mathcal{P}(\mathbb{R}) \times \mathbb{R}_+$ *at speed* n *with GRF* $I_{1,1}$ *defined by*

$$I_{1,1}(\mu,s) := H(\mu|\gamma_1) + \tfrac{1}{2}\left[s - M_2(\mu)\right],$$

if $M_2(\mu) \le s$, *and* $I_{1,1}(\mu,s) := +\infty$ *otherwise.*

Proof. Our approach is to apply the approximate contraction principle of Proposition 2.10.1 and Corollary 2.10.2, with the following parameters:

- $\Sigma := \mathbb{R}$;
- $\mathcal{X} := \mathbb{R} =: \mathcal{X}^*$;
- $\mathbf{c}(x) = x^2$, for $x \in \mathbb{R}$;
- for $n \in \mathbb{N}$, let $\mathcal{L}_n^{(1)} := \frac{1}{k_n}\sum_{j=1}^{k_n}\delta_{\zeta_j} \in \mathcal{P}(\mathbb{R})$, and $\mathcal{L}_n^{(2)} := \frac{1}{n-k_n}\sum_{j=k_n+1}^{n}\delta_{\zeta_j} \in \mathcal{P}(\mathbb{R})$;
- for $n \in \mathbb{N}$, let $C_n^{(i)} := \int \mathbf{c}\mathcal{L}_n^{(i)} = \int_{\mathbb{R}} \mathbf{c}(x)\mathcal{L}_n^{(i)}(dx)$, for $i = 1,2$.

First, let $\lambda \in (0,1)$, consider $(\mathcal{L}_n, C_n) = (\mathcal{L}_n^{(1)}, C_n^{(1)})$. Then the domain $\mathcal{D}$ of (2.75) takes the form $\mathcal{D} = \{\alpha \in \mathbb{R} : \log\mathbb{E}[e^{\lambda^{-1}\alpha\zeta_1^2}] < \infty\} = (-\infty, \frac{\lambda}{2})$, and so $0 \in \mathcal{D}^\circ = \mathcal{D}$. Thus, $F(x) = \sup_{\alpha<\lambda/2}\alpha x$, which is equal to $\lambda x/2$ if $x \ge 0$, and equal to ∞ otherwise, and

$\int \mathbf{c} d\mu = M_2(\mu)$. Thus, by Corollary 2.10.2, the sequence $\{\mathcal{L}_n^{(1)}, C_n^{(1)}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n and with GRF $J^{(1)} = J_\lambda^{(1)}$, defined by

$$J^{(1)}(\mu,t) := \begin{cases} \lambda H(\mu|\gamma_1) + \frac{\lambda}{2}[t - M_2(\mu)] & \text{if } M_2(\mu) \le t; \\ +\infty & \text{else.} \end{cases}$$

Similarly, another application of Corollary 2.10.2 shows that the sequence $\{\mathcal{L}_n^{(2)}, C_n^{(2)}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n with GRF $J^{(2)} = J_\lambda^{(2)}$, defined by

$$J^{(2)}(\nu,u) := \begin{cases} (1-\lambda) H(\nu|\gamma_1) + \frac{1-\lambda}{2}[u - M_2(\nu)] & \text{if } M_2(\nu) \le u; \\ +\infty & \text{else.} \end{cases}$$

Due to the independence of ζ_j, $j \in \mathbb{N}$, and the contraction principle applied to the mapping $\mathcal{P}(\mathbb{R}) \times \mathbb{R} \times \mathcal{P}(\mathbb{R}) \times \mathbb{R} \ni (\mu, t, \nu, u) \mapsto (\mu, \nu, \lambda t + (1-\lambda)u)$, the sequence

$$\left\{ \mathcal{L}_n^{(1)},\, \mathcal{L}_n^{(2)},\, \lambda \int \mathbf{c}\, d\mathcal{L}_n^{(1)} + (1-\lambda) \int \mathbf{c}\, d\mathcal{L}_n^{(2)} \right\}_{n\in\mathbb{N}} \tag{2.54}$$

satisfies an LDP at speed n with GRF

$$(\mu,\nu,s) \mapsto \inf_{t,u\in\mathbb{R}} \left\{ J^{(1)}(\mu,t) + J^{(2)}(\nu,u) : \lambda t + (1-\lambda)u = s \right\} = I_{1,\lambda}(\mu,\nu,s).$$

To complete the proof, note that because $k_n/n \to \lambda$ deterministically, the sequences

$$\left\{ \frac{\lambda}{k_n} \sum_{j=1}^{k_n} \zeta_j^2 + \frac{1-\lambda}{n-k_n} \sum_{j=k_n+1}^{n} \zeta_j^2 \right\}_{n\in\mathbb{N}} \quad \text{and} \quad \left\{ \frac{1}{n} \sum_{j=1}^{n} \zeta_j^2 \right\}_{n\in\mathbb{N}}$$

are exponentially equivalent (recall Definition 1.1.5). Hence, the sequence in (2.54) is also exponentially equivalent to (2.53), and the latter thus satisfies an LDP at speed n with GRF $I_{1,\lambda}$.

Now, let $\lambda = 1$. We see from the derivation above that the sequence $\{\mathcal{L}_n^{(1)}, C_n^{(1)}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n with GRF $J^{(1)}$, and that the two sequences $\{\frac{1}{k_n}\sum_{j=1}^{k_n}\zeta_j^2\}_{n\in\mathbb{N}}$ and $\{\frac{1}{n}\sum_{j=1}^{n}\zeta_j^2\}_{n\in\mathbb{N}}$ are exponentially equivalent. Hence, from Remark 1.1.6, we obtain the desired LDP. □

Lemma 2.7.11. *Suppose $k_n/n \to \lambda \in (0,1]$ and let $\{\zeta_j\}_{j\in\mathbb{N}}$ be as in Lemma 2.7.10. If $\lambda \in (0,1)$, the sequence of pairs of measures*

$$\left(\frac{1}{k_n}\sum_{j=1}^{k_n}\delta_{\sqrt{n}\zeta_j/\|\zeta^{(n)}\|_2}, \quad \frac{1}{n-k_n}\sum_{j=k_n+1}^{n}\delta_{\sqrt{n}\zeta_j/\|\zeta^{(n)}\|_2}\right), \quad n\in\mathbb{N}, \tag{2.55}$$

satisfies an LDP in $[\mathcal{P}(\mathbb{R})]^2$ at speed n with GRF $I_{2,\lambda}$ defined by

$$\begin{aligned} I_{2,\lambda}(\mu,\nu) &:= I_{1,\lambda}(\mu,\nu,1) \\ &= \lambda\, H(\mu|\gamma_1) + (1-\lambda)\, H(\nu|\gamma_1) + \tfrac{1}{2}\left(1 - \lambda\, M_2(\mu) - (1-\lambda)\, M_2(\nu)\right), \end{aligned}$$

if $\lambda\, M_2(\mu) + (1-\lambda)\, M_2(\nu) \le 1$, and $I_{2,\lambda}(\mu,\nu) := \infty$ otherwise. On the other hand, if $\lambda = 1$, then the sequence $\{\frac{1}{k_n}\sum_{j=1}^{k_n}\delta_{\sqrt{n}\zeta_j/\|\zeta^{(n)}\|_2}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\mathbb{R})$ at speed n with GRF $I_{2,1}$ defined by

$$I_{2,1}(\mu) = H(\mu|\gamma_1) + \tfrac{1}{2}\left(1 - M_2(\mu)\right),$$

if $M_2(\mu) \le 1$, and $I_{2,1}(\mu,\nu) := +\infty$ otherwise.

Proof. Below we present the proof only for the case $\lambda \in (0,1)$ since the case $\lambda = 1$ can be argued in an exactly analogous fashion. Due to Slutsky's theorem, the map

$$[\mathcal{P}(R)]^2 \times \mathbb{R}_+ \ni (\bar{\mu}, \bar{\nu}, s) \mapsto \left(\bar{\mu}(\cdot \times s^{1/2}),\, \bar{\nu}(\cdot \times s^{1/2})\right),$$

is continuous. Then, applying the contraction principle with this map to the LDP of Lemma 2.7.10, we find that the sequence in (2.55) satisfies an LDP with GRF:

$$\begin{aligned}
(\mu,\nu) \mapsto & \inf_{\bar\mu,\bar\nu\in\mathcal{P}(\mathbb{R}),s\in\mathbb{R}_+} \Big\{ I_{1,\lambda}(\bar\mu,\bar\nu,s) : \mu = \bar\mu(\,\cdot\times s^{1/2}), \nu = \bar\nu(\,\cdot\times s^{1/2}) \Big\} \\
&= \inf_{s\in\mathbb{R}_+} \Big\{ \lambda\, H(\mu(\,\cdot\times s^{-1/2})|\gamma_1) + (1-\lambda)\, H(\nu(\,\cdot\times s^{-1/2})|\gamma_1) \\
&\qquad + \tfrac{1}{2}\left[s - \lambda s\, M_2(\mu) - (1-\lambda) s\, M_2(\nu) \right] \,:\, s \geq \lambda s\, M_2(\mu) + (1-\lambda) s\, M_2(\nu) \Big\} \\
&= \inf_{s\in\mathbb{R}_+} \Big\{ \lambda\, H(\mu|\gamma_1(\,\cdot\times s^{1/2})) + (1-\lambda)\, H(\nu|\gamma_1(\,\cdot\times s^{1/2})) \\
&\qquad + \tfrac{s}{2}\left[1 - \lambda\, M_2(\mu) - (1-\lambda)\, M_2(\nu) \right] \,:\, 1 \geq \lambda\, M_2(\mu) + (1-\lambda)\, M_2(\nu) \Big\}
\end{aligned}$$

Assuming $1 \geq \lambda\, M_2(\mu) + (1-\lambda)\, M_2(\nu)$, noting that

$$\log\left(\frac{d\gamma_1}{d\gamma_1(\cdot\times s^{1/2})}(x) \right) = -\frac{1}{2}\log s - (1-s)\frac{x^2}{2},$$

and using the chain rule for relative entropy, the right-hand side of the previous display is equal to

$$\begin{aligned}
& \lambda\, H(\mu|\gamma_1) + (1-\lambda)\, H(\nu|\gamma_1) \\
&\qquad + \inf_{s\in\mathbb{R}_+} \Big\{ -\lambda \left(\tfrac{1}{2}\log s + \tfrac{1-s}{2} M_2(\mu) \right) - (1-\lambda) \left(\tfrac{1}{2}\log s + \tfrac{1-s}{2} M_2(\nu) \right) \\
&\qquad + \tfrac{s}{2}\left[1 - \lambda\, M_2(\mu) - (1-\lambda)\, M_2(\nu) \right] \Big\} \\
&= I_{2,\lambda}(\mu,\nu) - \tfrac{1}{2} + \tfrac{1}{2} \inf_{s\in\mathbb{R}_+} \{ s - \log s \} \\
&= I_{2,\lambda}(\mu,\nu).
\end{aligned}$$

On the other hand, if $1 < \lambda\, M_2(\mu) + (1-\lambda)\, M_2(\nu)$ is violated, then (by the convention that the infimum over an empty set is infinite) is infinite, which is again equal to $I_{2,\lambda}(\mu,\nu)$. This completes the proof. $\square$

We now turn to the proof of Proposition 2.7.8. First, for $\lambda \in (0,1]$, define

$$\mathbf{J}_{2,\lambda}(z) := \begin{cases} \frac{\lambda}{2}\log\left(\frac{\lambda}{z^2}\right) + \frac{1-\lambda}{2}\log\left(\frac{1-\lambda}{1-z^2}\right), & 0 < z < 1, \\ \infty, & \text{otherwise}, \end{cases} \tag{2.56}$$

where we use the convention $0\log 0 = 0\log(0/0) = 0$.

Proof of Proposition 2.7.8. We first prove that $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\mathbb{R})$ with respect to the weak topology, with GRF $\mathcal{H}_\lambda$ of (2.4). Due to the representation for $\mathbf{A}_{n,k_n}(1,\cdot)$ established in Lemma 2.7.1, it suffices to show that the sequence

$$\hat{\nu}_n := \frac{1}{k_n}\sum_{j=1}^{k_n} \delta_{\sqrt{n}\zeta_j/\|\zeta^{(n)}\|_2}, \quad n \in \mathbb{N}, \tag{2.57}$$

satisfies an LDP in $\mathcal{P}(\mathbb{R})$ at speed n with GRF $\mathcal{H}_\lambda$.

We deduce in the following the case for $\lambda \in (0,1)$. The case for $\lambda = 1$ can be shown using a similar calculation and is hence omitted. Combining the LDP established in Lemma 2.7.11 with the contraction principle applied to the projection mapping: $(\mathcal{P}(\mathbb{R}))^2 \ni (\pi_1,\pi_2) \mapsto \pi_1 \in \mathcal{P}(\mathbb{R})$, it follows that $\{\hat{\nu}_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n with GRF given by $\mu \mapsto \inf_{\nu\in\mathcal{P}(\mathbb{R})} I_{2,\lambda}(\mu,\nu)$. Now, note that for $\mu,\nu \in \mathcal{P}(\mathbb{R})$ such that $\lambda M_2(\mu) + (1-\lambda)M_2(\nu) \le 1$, we have

$$\begin{aligned} I_{2,\lambda}(\mu,\nu) = -\lambda h(\mu) &+ \tfrac{\lambda}{2}\log(2\pi) + \tfrac{\lambda}{2}M_2(\mu) \\ &- (1-\lambda)h(\nu) + \tfrac{1-\lambda}{2}\log(2\pi) + \tfrac{1-\lambda}{2}M_2(\nu) \\ &+ \tfrac{1}{2}(1 - \lambda\, M_2(\mu) - (1-\lambda)\, M_2(\nu)) \\ = -\lambda h(\mu) &- (1-\lambda)h(\nu) + \tfrac{1}{2}\log(2\pi e), \end{aligned}$$

where h is the entropy functional defined in (1.6). Thus,

$$\inf_{\nu\in\mathcal{P}(\mathbb{R})} I_{2,\lambda}(\mu,\nu) = -\lambda h(\mu) + \tfrac{1}{2}\log(2\pi e) + \inf_{\nu\in\mathcal{P}(\mathbb{R})} \{-(1-\lambda)h(\nu) : (1-\lambda)\, M_2(\nu) \le 1 - \lambda\, M_2(\mu)\}$$
$$= -\lambda h(\mu) + \tfrac{1}{2}\log(2\pi e) + (1-\lambda) \inf_{\nu\in\mathcal{P}(\mathbb{R})} \left\{-h(\nu) : M_2(\nu) \le \frac{1-\lambda\, M_2(\mu)}{1-\lambda}\right\}.$$

Since $M_2(\nu) \ge 0$, the right-hand side above is equal to infinity if $1 < \lambda M_2(\mu)$. On the other hand, if $\lambda M_2(\mu) \le 1$, then recalling that the maximum entropy probability measure under a second moment upper bound of z is the Gaussian measure with mean zero and variance z (see, e.g., Section 12 of [20]), we have

$$\begin{aligned}\inf_{\nu\in\mathcal{P}(\mathbb{R})} I_{2,\lambda}(\mu,\nu) &= -\lambda h(\mu) + \tfrac{1}{2}\log(2\pi e) - \tfrac{1-\lambda}{2}\log\left(2\pi e\tfrac{1-\lambda\, M_2(\mu)}{1-\lambda}\right) \\ &= -\lambda h(\mu) + \tfrac{\lambda}{2}\log(2\pi e) + \tfrac{1-\lambda}{2}\log\left(\tfrac{1-\lambda}{1-\lambda\, M_2(\mu)}\right) \\ &= \mathcal{H}_\lambda(\mu).\end{aligned}$$

Thus, we have shown that the rate function is $\mathcal{H}_\lambda(\mu)$, as desired.

To strengthen the LDP on $\mathcal{P}(\mathbb{R})$ to an LDP on $\mathcal{P}_q(\mathbb{R})$ (i.e., with respect to the q-Wasserstein topology), via an appeal to [21, Corollary 4.2.6] it suffices to show that $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ is exponentially tight in $\mathcal{P}_q(\mathbb{R})$. Since for each $j > 0$, the set $K_{2,j}$ defined in (1.4) is compact in the q-Wasserstein topology by Lemma 1.1.1. By the definition of $K_{2,j}$, the definition of M_2 in (1.3), and the identity $\hat{\mu}^n_A \overset{(d)}{=} \hat{\nu}_n$ established in (2.45) and Lemma 2.7.1, and the definition of $\hat{\nu}_n$ in (2.57),

$$\{\hat{\mu}^n_{\mathbf{A}} \in K^c_{2,M}\} = \{\hat{\nu}_n \in K^c_{2,M}\} = \{M_2(\hat{\nu}_n) > M\} = \left\{\sqrt{\frac{n}{k_n}}\frac{\|\zeta^{(k_n)}\|_2}{\|\zeta^{(n)}\|_2} > \sqrt{M}\right\}. \tag{2.58}$$

When $k_n/n \to \lambda \in (0,1]$, by [3, Lemma 4.2], $\{\|\zeta^{(k_n)}\|_2/\|\zeta^{(n)}\|_2\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n with GRF $\mathbf{J}_{2,\lambda}$ defined in (2.56), and $\{\sqrt{M_2(\hat{\nu}_n)}\}_{n\in\mathbb{N}}$ is exponentially equivalent to $\{\|\zeta^{(k_n)}\|_2/(\sqrt{\lambda}\|\zeta^{(n)}\|_2)\}_{n\in\mathbb{N}}$ at speed n by Remark 1.1.7. Combining this with (2.58) and Remark 1.1.6, when $M = 4/\lambda$, we obtain

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{\mu}^n_{\mathbf{A}} \in K^c_{2,M}) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{\|\zeta^{(k_n)}\|_2}{\sqrt{\lambda}\|\zeta^{(n)}\|_2} > \sqrt{M} \right) \le -\mathbf{J}_{2,\lambda}(2) = -\infty.$$

This proves the desired exponential tightness of $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ in $\mathcal{P}_q(\mathbb{R})$. □

2.8 Proofs of q-norm LDPs in the sublinear regime, $1 \ll k_n \ll n$

In this section we prove Theorems 2.4.4 and 2.4.5. Recall from (2.7) that $Y^n_{2,k_n} = \|\mathbf{A}^\top_{n,k_n} X^{(n)}\|_2$.

Proof of Theorem 2.4.4. Fix $\{k_n\}_{n\in\mathbb{N}}$ that grows sublinearly. From Lemma 2.7.2 and (2.44), we have the following distributional identity,

$$n^{-1/2} Y^n_{2,k_n} \overset{(d)}{=} \frac{\left\|\zeta^{(k_n)}\right\|_2}{\left\|\zeta^{(n)}\right\|_2} \frac{\left\|X^{(n)}\right\|_2}{\sqrt{n}} = \frac{\left\|\zeta^{(k_n)}\right\|_2 / \sqrt{k_n}}{\left\|\zeta^{(n)}\right\|_2 / \sqrt{n}} \frac{\sqrt{k_n}\left\|X^{(n)}\right\|_2}{n}. \tag{2.59}$$

Suppose Assumption 2.1.3* holds, in which case $\{\left\|X^{(n)}\right\|_2 / \sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n with GRF J_X. On the other hand, by [3, Lemma 4.2], $\{\|\zeta^{(k_n)}\|_2/\|\zeta^{(n)}\|_2\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n with GRF

$$J_\zeta(x) := \begin{cases} -\frac{1}{2}\log(1-x^2), & x \in [0,1), \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.60}$$

Due to the independence of $\zeta^{(n)}$ and $X^{(n)}$, the result follows by Lemma 1.1.9.

Now, suppose Assumption 2.4.2 holds with sequence $\{s_n\}_{n\in\mathbb{N}}$, $r \in [0,\infty]$ and GRF $J_X^{(r)}$. We will make repeated use of the following simple observation: by Remark 2.7.6 and the contraction principle applied to the map $x \mapsto \sqrt{x}$, $\{\|\zeta^{(k_n)}\|_2/\sqrt{k_n}\}_{n\in\mathbb{N}}$ and $\{\|\zeta^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfy LDPs at speed k_n and n, respectively, with the same GRF $J_{\zeta^2}(x^2)$, and hence, $\{\sqrt{n}/\|\zeta^{(n)}\|_2\}_{n\in\mathbb{N}}$ also satisfies an LDP at speed n, and all three sequences converge almost surely to 1 by the strong law of large numbers. We now consider three cases.

Case 2.8.1. Suppose $r = 0$. Then $s_n \ll k_n \ll n$. By the case assumption, $\{\sqrt{k_n}\|X^{(n)}\|_2/n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF $J_X^{(0)}$. Hence, the result in this case follows by (2.59), the observation above and a dual application of Lemma 1.1.9 with (V_n, W_n) therein first being $(\sqrt{k_n}\|X^{(n)}\|_2/n, \|\zeta^{(k_n)}\|_2/\sqrt{k_n})$ and then $(\|\zeta^{(k_n)}\|_2\|X^{(n)}\|_2/n, \sqrt{n}/\|\zeta^{(n)}\|_2)$.

Case 2.8.2. Suppose $r \in (0,\infty)$. By Remark 2.4.3, $\{\sqrt{k_n}\|X^{(n)}\|_2/n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed k_n and with GRF $rJ_X^{(r)}(\sqrt{r}x)$. Then (2.59), the independence of $X^{(n)}$ and $\zeta^{(n)}$, the observation above and Lemma 1.1.9 together show that $\{n^{-1/2}Y^n_{2,k_n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed k_n with GRF

$$\mathbb{J}_2^{\mathrm{sub}}(x) := \inf_{c>0}\left\{\frac{c^2-1}{2} - \log c + rJ_X^{(r)}\left(\frac{\sqrt{r}x}{c}\right)\right\}.$$

Case 2.8.3. Suppose $r = \infty$. Again from the reformulation (2.59), we have

$$n^{-1/2}Y^n_{2,k_n} \overset{(d)}{=} \frac{\left\|\zeta^{(k_n)}\right\|_2/\sqrt{s_n}}{\left\|\zeta^{(n)}\right\|_2/\sqrt{n}} \frac{\sqrt{s_n}\left\|X^{(n)}\right\|_2}{n}. \tag{2.61}$$

Since

$$\frac{\left\|\zeta^{(k_n)}\right\|_2}{\sqrt{s_n}} = \left(\frac{1}{s_n}\sum_{i=1}^{k_n}\zeta_i^2\right)^{1/2},$$

and (because $r = \infty$) $k_n/s_n \to 0$ as $n \to \infty$, by Lemma 2.6.1 (with $p = 2$) and the contraction principle applied to $t \mapsto \sqrt{t}$, $\{\|\zeta^{(k_n)}\|_2/\sqrt{s_n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF $\mathcal{J}_\zeta(t) = t^2/2$ if $t \geq 0$, and $\mathcal{J}_\zeta(t) = \infty$ otherwise. By the case assumption, $\{\sqrt{s_n}\|X^{(n)}\|_2/n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF $J_X^{(\infty)}$. The independence of $\{\zeta_i\}_{i\in\mathbb{N}}$ and $X^{(n)}$, together with the contraction principle applied to the product mapping $(x, y) \mapsto xy$, then implies that $\{V_n := \|\zeta^{(k_n)}\|_2\|X^{(n)}\|_2/n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF $\inf_{c>0}\{c^2/2 + J_X^\infty(x/c)\}$. The lemma follows on applying Lemma 1.1.9 with $\{V_n\}_{n\in\mathbb{N}}$ defined above, $W_n := (\|\zeta^{(n)}\|_2/\sqrt{n})^{-1}$, $n \in \mathbb{N}$, and $m = 1$.

□

Proof of Theorem 2.4.5. Fix $q \in [1, 2]$ and suppose k_n grows sublinearly. From Lemma 2.7.2 and (2.44), we have the following reformulation

$$k_n^{-1/q} Y_{q,k_n}^n \overset{(d)}{=} \frac{\left\|\zeta^{(k_n)}\right\|_q / k_n^{1/q}}{\left\|\zeta^{(n)}\right\|_2 / \sqrt{n}} \frac{\left\|X^{(n)}\right\|_2}{\sqrt{n}}, \tag{2.62}$$

where $\zeta^{(n)} := (\zeta_1, \ldots, \zeta_n)$ with $\{\zeta_i\}_{i\in\mathbb{N}}$ being i.i.d. $\mathcal{N}(0, 1)$ random variables. Consider the following sequence of random vectors

$$R_n := \left(\frac{\left\|\zeta^{(k_n)}\right\|_q}{k_n^{1/q}}, \frac{\left\|\zeta^{(n)}\right\|_2}{\sqrt{n}}, \frac{\left\|X^{(n)}\right\|_2}{\sqrt{n}} \right), \quad n \in \mathbb{N}.$$

By Assumption 2.1.3, $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF J_X. By Cramér's theorem and the contraction principle, $\{\|\zeta^{(k_n)}\|_q/k_n^{1/q}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed k_n with GRF $J_{\zeta,q}(x) := \Lambda_q^*(x^q)$ if $x \geq 0$, with Λ_q defined in (2.15), and $J_{\zeta,q}(x) = \infty$ if $x < 0$. Note that Λ_q^* is strictly convex with unique minimizer $\mathcal{M}_q^{1/q}$. Similarly, $\{\|\zeta^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n with GRF $J_{\zeta,2}(x) := \Lambda_2^*(x^2)$ if $x \geq 0$ (and is equal to infinity otherwise) with unique minimizer 1.

To prove the theorem, we will use Lemma 1.1.9 to show that in each of the three regimes, $\{R_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $s_n \wedge k_n$ with a GRF J_R that we identify in each case. In view of (2.62) and the contraction principle, this would imply that $\{k_n^{-1/q} Y^n_{q,k_n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $s_n \wedge k_n$ with GRF

$$\widehat{\mathbb{J}}_q^{\mathrm{sub}}(u) := \inf\left\{ J_R(x,y,z) : u = \frac{xz}{y}, x,y,z > 0 \right\}. \tag{2.63}$$

Case 2.8.4. Suppose $k_n \ll s_n$. Then we also have $k_n \ll n$, and by Lemma 1.1.9 and the additional assumption that J_X has a unique minimizer m, $\{R_n\}_{n\in\mathbb{N}}$ is exponentially equivalent at speed k_n to $(\|\zeta^{(k_n)}\|_q / k_n^{1/q}, 1, m)$, which satisfies an LDP at speed k_n with GRF $J_R(x,y,z) = J_{\zeta,q}(x)$ when $y = 1$ and $z = m$, and $J_R(x,y,z) = +\infty$ otherwise. Then (2.63) shows that $\widehat{\mathbb{J}}_q^{\mathrm{sub}}(u) = J_{\zeta,q}(u/m) = \Lambda_q^*(u^q/m^q)$ for $u \geq 0$, and is equal to positive infinity otherwise.

Case 2.8.5. Suppose $s_n = k_n$. Since $k_n \ll n$, again invoking Lemma 1.1.9 and the additional assumption that J_X has a unique minimizer m, we have the exponential equivalence at speed k_n between $\{R_n\}_{n\in\mathbb{N}}$ and $\{(\|\zeta^{(k_n)}\|_q / k_n^{1/q}, 1, \|X^{(n)}\|_2/\sqrt{n})\}_{n\in\mathbb{N}}$. Hence, $\{R_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed k_n with GRF $J_R(x,y,z) = J_{\zeta,q}(x) + J_X(z)$ when $y = 1$, and $J_R(x,y,z) = +\infty$ otherwise. Moreover, (2.63) shows that $\widehat{\mathbb{J}}_q^{\mathrm{sub}}(u) = \inf_{x>0}\{\Lambda_q^*(x^q) + J_X(u/x)\}$ if $u > 0$, (and is equal to $+\infty$ otherwise).

Case 2.8.6. Suppose $s_n \ll k_n$. Then $k_n \ll n$ implies $s_n \ll n$, and so Lemma 1.1.9 and the observation that the GRF $J_{\zeta,q}$ has $M_q^{1/q}$ as its unique minimizer imply that $\{R_n\}_{n\in\mathbb{N}}$ is exponentially equivalent at speed s_n to the sequence $\{(\mathcal{M}_q^{1/q}, 1, \|X^{(n)}\|_2/\sqrt{n})\}_{n\in\mathbb{N}}$, and therefore satisfies an LDP at speed s_n with GRF $J_R(x,y,z) = J_X(z)$ when $x = \mathcal{M}_q^{1/q}$ and $y = 1$, and $J_R(x,y,z) = +\infty$ otherwise. When combined with (2.63) this implies that $\widehat{\mathbb{J}}_q^{\mathrm{sub}}(u) = J_X(u/\mathcal{M}_q^{1/q})$ if $u \geq 0$ (and is equal to $+\infty$ otherwise).

□

2.9 Proofs of q-norm LDPs in the linear regime

The goal of this section is to prove Theorem 2.5.2. Throughout, fix $\lambda \in (0,1]$, and assume $k_n \sim \lambda n$. Also, for $q \in [1,2]$ and $n,k \in \mathbb{N}, k \leq n$, recall the definition $n^{-1/q}Y^n_{q,k_n} = n^{-1/q}\|\mathbf{A}^T_{n,k_n}X^{(n)}\|_q$ given in (2.7). Section 2.9.1 contains a simple proof that is valid when $q \in [1,2)$. Section 2.9.2 is devoted to the more involved case of $q = 2$ in the linear regime, which also then provides an alternative proof and alternative form of the rate function in the case $q \in [1,2)$.

2.9.1 The case $q \in [1,2)$

Proof of Theorem 2.5.2 when $q \in [1,2)$. Fix $q \in [1,2)$, and observe that with M_q, L^n, Y^n_{q,k_n} and $\mathbf{A}_{n,k_n}$ defined as in (1.3), (2.1), (2.7) and Section 2.1.1, respectively, it follows that

$$\left(\frac{k_n}{n}M_q(L^n)\right)^{1/q} = n^{-1/q}\|\mathbf{A}^T_{n,k_n}X^{(n)}\|_q = n^{-1/q}Y^n_{q,k_n}.$$

Then, the LDP for $\{L^n\}_{n\in\mathbb{N}}$ in Theorem 2.5.1 and the contraction principle applied to the continuous map $\mathcal{P}_q(\mathbb{R}) \ni \nu \mapsto (\lambda M_q)^{1/q}(\nu) \in \mathbb{R}$ imply that $\{\lambda M_q(L^n)\}_{n\in\mathbb{N}}$ satisfies an LDP with GRF

$$\mathbb{J}^{\mathrm{lin}}_{q,\lambda}(x) \quad := \quad \inf_{\mu\in\mathcal{P}(\mathbb{R})}\{\mathbb{I}_{L,\lambda}(\mu) : (\lambda M_q(\mu))^{1/q} = x\}.$$

Since $\frac{k_n}{n}M_q(L^n)$ is exponentially equivalent to $\lambda M_q(L^n)$ at speed n by Remark 1.1.7, this implies (by Remark 1.1.6) that $\{n^{-1/q}Y^n_{q,k_n}\}_{n\in\mathbb{N}}$ also satisfies an LDP at speed n with GRF $\mathbb{J}^{\mathrm{lin}}_{q,\lambda}$.

If $x < 0$, we infimize over an empty set, and so $\mathbb{J}^{\text{lin}}_{q,\lambda}(x) = -\infty$. Now, if $s_n = n$, using the expression for $\mathbb{I}_{L,\lambda}$ given in Remark 2.7.9, we see that for $x \geq 0$,

$$\begin{aligned}
\mathbb{J}^{\text{lin}}_{q,\lambda}(x) &= \inf_{c\in\mathbb{R}_+,\, \mu\in\mathcal{P}(\mathbb{R})} \left\{ \mathcal{H}_\lambda(\mu(\cdot \times c)) + J_X(c) : [\lambda M_q(\mu)]^{1/q} = x \right\}, \\
&= \inf_{c\in\mathbb{R}_+,\, \nu\in\mathcal{P}(\mathbb{R})} \left\{ \mathcal{H}_\lambda(\nu) + J_X(c) : [\lambda M_q(\nu(\cdot \times c^{-1}))]^{1/q} = x \right\}, \\
&= \inf_{c\in\mathbb{R}_+,\, \nu\in\mathcal{P}(\mathbb{R})} \left\{ \mathcal{H}_\lambda(\nu) + J_X(c) : [\lambda M_q(\nu)]^{1/q} = \frac{x}{c} \right\},
\end{aligned}$$

which coincides with (2.18).

On the other hand, if $s_n \ll n$, using the identity $\mathbb{I}_{L,\lambda}(\mu) = J_X(c)$ if $\mu = \gamma_c$ (and infinity, otherwise) established in Theorem 2.5.1, we have for $x \geq 0$,

$$\begin{aligned}
\mathbb{J}^{\text{lin}}_{q,\lambda}(x) &= \inf_{c\in\mathbb{R}_+} \{ J_X(c) : (\lambda M_q(\gamma_c))^{1/q} = x \} \\
&= \inf_{c\in\mathbb{R}_+} \{ J_X(c) : (\lambda c^q \mathcal{M}_q)^{1/q} = x \} \\
&= J_X\left(\frac{x}{(\lambda \mathcal{M}_q)^{1/q}} \right),
\end{aligned}$$

where $\mathcal{M}_q$ is as defined in (2.8). This proves (2.18). □

2.9.2 The case $q = 2$ and alternative proof for $q \in [1, 2)$

We now provide an alternative proof of Theorem 2.5.2 for $q \in [1, 2)$, which also extends to the case $q = 2$. This yields an alternative representation for the rate function $\mathbb{J}^{\text{lin}}_{q,\lambda}$ of (2.18). To introduce this representation, fix $q \in [1, 2]$ and define the following functions. Let

$$\Lambda_{\mathrm{A},q}(t_1, t_2) := \log \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(t_1 |x|^q + (t_2 - \tfrac{1}{2}) x^2 \right) dx, \quad (t_1, t_2) \in \mathbb{R}^2. \tag{2.64}$$

Note that for $q \in [1,2)$, we have $\Lambda_{\mathrm{A},q}(t_1,t_2) < \infty$ when $t_1 \in \mathbb{R}$ and $t_2 < \frac{1}{2}$. On the other hand, for $q = 2$, we have $\Lambda_{\mathrm{A},q}(t_1,t_2) < \infty$ when $t_1 + t_2 < \frac{1}{2}$. We also define

$$\Lambda_{\mathrm{B}}(t_3) := \log \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left((t_3 - \tfrac{1}{2})x^2\right) dx, \quad t_3 \in \mathbb{R}. \tag{2.65}$$

Note that $\Lambda_{\mathrm{B}}(t_3) < \infty$ for $t_3 < \frac{1}{2}$. Let $\Lambda^*_{\mathrm{A},q}$ and Λ^*_{B} denote the Legendre-Fenchel transforms of $\Lambda_{\mathrm{A},q}$ and Λ_{B}, respectively. For $q \in [1,2)$ and $\lambda \in (0,1)$, define

$$\mathbf{J}_{q,\lambda}(z) := \inf_{(x_1,x_2,x_3)\in\mathbb{R}^3} \left\{ \lambda \Lambda^*_{\mathrm{A},q}\left(\frac{(x_1,x_2)}{\lambda}\right) + (1-\lambda)\Lambda^*_{\mathrm{B}}\left(\frac{x_3}{1-\lambda}\right) \,:\, z = \frac{x_1^{1/q}}{(x_2+x_3)^{1/2}} \right\}, \tag{2.66}$$

and define for $\lambda = 1$,

$$\mathbf{J}_{q,1}(z) := \inf_{(x_1,x_2)\in\mathbb{R}^3} \left\{ \Lambda^*_{\mathrm{A},q}\left((x_1,x_2)\right) \,:\, z = \frac{x_1^{1/q}}{(x_2)^{1/2}} \right\}. \tag{2.67}$$

Also, recall the definition of $\mathbf{J}_{2,\lambda}$, $\lambda \in (0,1]$, from (2.56).

We then have the following result.

Proposition 2.9.1. *Fix $q \in [1,2]$. Suppose $\{k_n\}_{n\in\mathbb{N}}$ grows linearly with rate $\lambda \in (0,1]$, and that Assumption 2.1.3 holds with associated GRF J_X. Then, the sequence of scaled ℓ_q^n norms of random projections, $\{n^{-1/q}Y^n_{q,k_n} = n^{-1/q}\|\mathbf{A}^T_{n,k_n}X^{(n)}\|_q\}_{n\in\mathbb{N}}$, satisfies an LDP at speed n with GRF*

$$\mathbb{J}^{\mathrm{lin}}_{q,\lambda}(x) := \inf_{y,z\in\mathbb{R}} \left\{ \mathbf{J}_{q,\lambda}(z) + J_X(y) \,:\, x = yz \right\}, \quad x \in \mathbb{R}_+. \tag{2.68}$$

The proof of Proposition 2.9.1 relies on an auxiliary result stated in Lemma 2.9.2, which concerns an LDP related to the top row of the matrix $\mathbf{A}_{n,k_n}$. As in Section 2.7, let $\zeta_1, \zeta_2, \ldots$ denote a sequence of i.i.d. standard Gaussian random variables, independent of $X^{(n)}$, and let $\zeta^{(n)} := (\zeta_1, \dots, \zeta_n) \in \mathbb{R}^n$. Due to Lemmas 2.7.2 and 2.7.1, we have

$$\mathbf{A}_{n,k_n}^T X^{(n)} = \mathbf{A}_{n,k_n}^T \tfrac{X^{(n)}}{\|X^{(n)}\|_2} \|X^{(n)}\|_2 \stackrel{(\mathrm{d})}{=} \mathbf{A}_{n,k_n}^T e_1 \|X^{(n)}\|_2 \stackrel{(\mathrm{d})}{=} \frac{(\zeta_1, \dots, \zeta_{k_n})}{\|\zeta^{(n)}\|_2} \|X^{(n)}\|_2 \in \mathbb{R}^{k_n}.$$

Therefore, for $n \in \mathbb{N}$ and all $q \in [1, \infty]$, we have

$$n^{-1/q} Y_{q,k_n}^n = n^{-1/q} \|\mathbf{A}_{n,k_n}^T X^{(n)}\|_q \stackrel{(\mathrm{d})}{=} n^{1/2-1/q} \frac{\|\zeta^{(k_n)}\|_q}{\|\zeta^{(n)}\|_2} \frac{\|X^{(n)}\|_2}{\sqrt{n}}. \tag{2.69}$$

Given (2.69) and Assumption 2.1.3, a natural step to proving Proposition 2.9.1 is to establish an LDP for the sequence $\{n^{1/2-1/q}\|\zeta^{(k_n)}\|_q / \|\zeta^{(n)}\|_2\}_{n\in\mathbb{N}}$.

Lemma 2.9.2. *Suppose $\{k_n\}_{n\in\mathbb{N}}$ grows linearly with rate $\lambda \in [0, 1]$. For $q \in [1, 2]$, the sequence*

$$n^{1/2-1/q} \frac{\|\zeta^{(k_n)}\|_q}{\|\zeta^{(n)}\|_2} = n^{1/2-1/q} \frac{\|(\zeta_1, \dots, \zeta_{k_n})\|_q}{\|(\zeta_1, \dots, \zeta_n)\|_2}, \quad n \in \mathbb{N}, \tag{2.70}$$

satisfies an LDP in $\mathbb{R}$ at speed n with GRF $\mathbf{J}_{q,\lambda}$ of (2.66) and (2.67) for $q \in [1, 2)$ and (2.56) for $q = 2$.

Proof. Fix $q \in [1, 2)$. For $n \in \mathbb{N}$. We start with the observation that the quantity in (2.70) can be represented in terms of a continuous mapping of a vector of scaled i.i.d. sums: for each $n \in \mathbb{N}$,

$$n^{1/2-1/q} \frac{\|\zeta^{(k_n)}\|_q}{\|\zeta^{(n)}\|_2} = T(Z_n), \quad Z_n := \left(\frac{1}{n} \sum_{i=1}^{k_n} |\zeta_i|^q, \frac{1}{n} \sum_{i=1}^{k_n} \zeta_i^2, \frac{1}{n} \sum_{j=k_n+1}^{n} \zeta_j^2 \right), \tag{2.71}$$

where T is the continuous map

$$T : \mathbb{R}^3 \ni (x_1, x_2, x_3) \mapsto \frac{x_1^{1/q}}{(x_2 + x_3)^{1/2}} \in \mathbb{R}.$$

We now establish an LDP for the sequence $\{Z_n\}_{n\in\mathbb{N}}$. To this end, for fixed $\lambda \in (0,1)$, we first establish LDPs of the related sequences

$$\mathrm{A}_{n,q} := \frac{\lambda}{k_n}\sum_{i=1}^{k_n}(|\zeta_i|^q, \zeta_i^2), \qquad \mathrm{B}_n := \frac{1-\lambda}{n-k_n}\sum_{j=k_n+1}^{n}\zeta_j^2, \quad n\in\mathbb{N}.$$

Note that for $q\in[1,2]$, the origin $(0,0)$ lies in the interior of the domain of $\Lambda_{\mathrm{A},q}$, which is the logarithmic moment generating function of $(|\zeta_1|^q, \zeta_1^2)$. Since the $\{\zeta_i\}_{i\in\mathbb{N}}$ are i.i.d., by Cramér's theorem [21, Theorem 2.2.1] the sequence $\{\mathrm{A}_{n,q}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^2$ at speed k_n with GRF $\Lambda^*_{\mathrm{A},q}(\cdot/\lambda)$, and hence (by Remark 1.1.4 and the fact that $k_n/n\to\lambda$) also satisfies an LDP in $\mathbb{R}^2$ at speed n with GRF $\lambda\Lambda^*_{\mathrm{A},q}(\cdot/\lambda)$. Similarly, since 0 belongs to the interior of the domain of Λ_{B}, the logarithmic moment generating function of ζ_1^2, the sequence $\{\mathrm{B}_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ at speed n with GRF $(1-\lambda)\Lambda^*_{\mathrm{B}}(\cdot/(1-\lambda))$. Since $k_n/n\to\lambda$ and $(n-k_n)/n\to 1-\lambda$ deterministically, by Remarks 1.1.6 and 1.1.7, $\{\frac{1}{n}\sum_{i=1}^{k_n}(|\zeta_i|^q,\zeta_i^2)\}_{n\in\mathbb{N}}$ and $\{\frac{1}{n}\sum_{j=k_n+1}^{n}\zeta_j^2\}_{n\in\mathbb{N}}$ satisfy LDPs at speed n with the same GRFs as $\{\mathrm{A}_{n,q}\}_{n\in\mathbb{N}}$ and $\{\mathrm{B}_n\}_{n\in\mathbb{N}}$, respectively. Combined with the independence of $\mathrm{A}_{n,q}$ and B_n, and Lemma 1.1.9, it follows that the sequence $\{Z_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^3$ at speed n with GRF

$$(x_1,x_2,x_3)\mapsto \lambda\,\Lambda^*_{\mathrm{A},q}\left(\frac{(x_1,x_2)}{\lambda}\right) + (1-\lambda)\,\Lambda^*_{\mathrm{B}}\left(\frac{x_3}{1-\lambda}\right), \qquad x\in\mathbb{R}^3_+,$$

and with the GRF equal to $-\infty$ otherwise. Finally, use (2.71), the LDP for $\{Z_n\}_{n\in\mathbb{N}}$ and the contraction principle for the map T defined above to show that the sequence in (2.70) satisfies an LDP with the GRF $\mathbf{J}_{q,\lambda}$ of (2.66).

Next, suppose $\lambda = 1$. We see that the sequence $\{\mathrm{A}_{n,q}\}_{n\in\mathbb{N}}$ is exponentially equivalent to $\{(\frac{1}{n}\sum_{i=1}^{k_n}|\zeta_i|^q, \frac{1}{n}\sum_{i=1}^{n}\zeta^2)\}_{n\in\mathbb{N}}$ by Remark 1.1.7. An application of the contraction principle to the map $(x_1,x_2)\mapsto x_1^{1/q}/x_2^{1/2}$ yields the LDP for the sequence in (2.70) with speed

n and GRF (2.67).

The case $q = 2$ was also considered in [3, Lemma 4.2]), but can be obtained via (a simpler form of) the argument given above. To be self-contained, we provide some details. In this case, the quantity of interest is $\tilde{T}(\tilde{Z}_n)$, where $\tilde{Z}_n$, is the $\mathbb{R}^2$-valued random vector that coincides with the second and third components of Z_n, and $\tilde{T}$ maps $(x_2, x_3) \mapsto x_2/(x_2+x_3)^{1/2}$, and the arguments above show that $\{\tilde{Z}_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n with GRF $\lambda\Lambda_{\mathrm{B}}^*(\cdot/\lambda)+(1-\lambda)\Lambda_{\mathrm{B}}^*(\cdot/(1-\lambda))$ and thus, by the contraction principle, $\{\|\zeta^{(k_n)}\|/\|\zeta^{(n)}\|\}_{n\in\mathbb{N}}$ satisfies an LDP with GRF

$$\mathbf{J}_{2,\lambda}(z) := \inf_{(x_2,x_3)\in\mathbb{R}^2} \left\{ \lambda\Lambda_{\mathrm{B}}^*\left(\frac{x_2}{\lambda}\right) + (1-\lambda)\Lambda_{\mathrm{B}}^*\left(\frac{x_3}{1-\lambda}\right) : z^2 = \frac{x_2}{x_2+x_3} \right\}.$$

Noting that Λ_B is the log moment generating function of the chi-squared distribution, it follows from Remark 2.7.6 that $\Lambda_B^*(x) = J_{\zeta^2}(x) = \frac{1}{2}(x - 1 - \log x)$, for $x > 0$ and infinity, otherwise. Substituting this into the expression for $\mathbf{J}_{2,\lambda}(z)$ in the last display and solving the optimization problem yields (2.56). □

Proof of Proposition 2.9.1. Recall from Lemma 2.9.2 that the sequence $\{n^{1/2-1/q}\|\zeta^{(k_n)}\|_q/\|\zeta^{(n)}\|_2\}_{n\in\mathbb{N}}$ satisfies an LDP with GRF $\mathbf{J}_{q,\lambda}$. In addition, Assumption 2.1.3 states that the sequence $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP with GRF J_X. Given the equality in distribution of (2.69), the contraction principle applied to the continuous function $(y,z) \mapsto yz$ yields that the sequence of scaled ℓ_q^n norms of random projections, $\{Y^n_{q,k_n}\}_{n\in\mathbb{N}}$, satisfies an LDP at speed n with GRF $\bar{\mathbb{J}}^{\mathrm{lin}}_{q,\lambda}$ of (2.68). □

Remark 2.9.3. Note that it follows from Proposition 2.9.1 and the proof of Theorem 2.5.2 for $q \in [1,2)$ in Section 2.9.1 that $\bar{\mathbb{J}}^{\mathrm{lin}}_{q,\lambda} = \mathbb{J}^{\mathrm{lin}}_{q,\lambda}$ for all $q \in [1,2)$.

To complete the proof of Theorem 2.5.2 we show that this relation also holds for

$q = 2$.

Proof of Theorem 2.5.2 in the case $q = 2$. First, consider the case $s_n = n$. Fix $\lambda \in (0,1)$. The result for $\lambda = 1$ follows on using the convention that $0 \log 0 = 0 \log 0/0 = 0$ everywhere in the derivation below. By Proposition 2.9.1, it suffices to show that $\bar{\mathbb{J}}^{\text{lin}}_{2,\lambda} = \mathbb{J}^{\text{lin}}_{2,\lambda}$. In view of (2.18), (2.68), and (2.56), it clearly suffices to show that for $z^2 \in (0,1)$,

$$\mathbf{J}_{2,\lambda}(z) = \frac{\lambda}{2} \log\left(\frac{\lambda}{z^2}\right) + \frac{1-\lambda}{2} \log\left(\frac{1-\lambda}{1-z^2}\right) = \inf_{\nu \in \mathcal{P}(\mathbb{R}): z^2 = \lambda M_2(\nu)} \mathcal{H}_\lambda(\nu). \tag{2.72}$$

since when $z^2 \notin (0,1)$ both $\mathbf{J}_{2,\lambda}(z)$ and the right-hand side above are infinite. For $z^2 \in (0,1)$, substituting the expression for $\mathcal{H}_\lambda$ from (2.4), we obtain

$$\begin{aligned}\inf_{\nu \in \mathcal{P}(\mathbb{R}): z^2 = \lambda M_2(\nu)} \mathcal{H}_\lambda(\nu) &= \inf_{\nu \in \mathcal{P}(\mathbb{R}): z^2 = \lambda M_2(\nu)} \left\{-\lambda\, h(\nu) + \tfrac{\lambda}{2}\log(2\pi e) + \tfrac{1-\lambda}{2}\log\left(\tfrac{1-\lambda}{1-\lambda\, M_2(\nu)}\right)\right\} \\ &= -\sup_{\nu \in \mathcal{P}(\mathbb{R}): z^2 = \lambda M_2(\nu)} \lambda\, h(\nu) + \tfrac{\lambda}{2}\log(2\pi e) + \tfrac{1-\lambda}{2}\log\left(\tfrac{1-\lambda}{1-z^2}\right).\end{aligned}$$

Recall that the maximum entropy probability measure constrained to have second moment z^2/λ is the Gaussian measure with mean zero and variance z^2/λ (see, e.g., Ex. 12.2.1 of [20]). Since the entropy of such a Gaussian equals $\frac{1}{2}\log(2\pi e z^2/\lambda)$, we have

$$\inf_{\nu \in \mathcal{P}(\mathbb{R}): z^2 = \lambda M_2(\nu)} \mathcal{H}_\lambda(\nu) = -\tfrac{\lambda}{2}\log(2\pi e z^2/\lambda) + \tfrac{\lambda}{2}\log(2\pi e) + \tfrac{1-\lambda}{2}\log\left(\tfrac{1-\lambda}{1-z^2}\right) = \mathbf{J}_{2,\lambda}(z),$$

which proves (2.72). This completes the proof of the equality $\bar{\mathbb{J}}^{\text{lin}}_{2,\lambda} = \mathbb{J}^{\text{lin}}_{2,\lambda}$ in this case.

Now, suppose $s_n \ll n$. By (2.69), it follows that $n^{-1/2} Y^n_{2,k_n} \overset{(d)}{=} V_n W_n W'_n$, where $V_n = \|X^{(n)}\|_2/\sqrt{n}$, $W_n := \sqrt{k_n/n}\|\zeta^{(k_n)}\|_2/\sqrt{k_n}$ and $W'_n = \sqrt{n}/\|\zeta^{(n)}\|_2$. Noting that $k_n/n \to \lambda$ as $n \to \infty$, we see that $\{W_n\}_{n\in\mathbb{N}}$ and $\{W'_n\}_{n\in\mathbb{N}}$ converge almost surely to $(\lambda \mathcal{M}_2)^{1/2}$ and 1, respectively (by the strong law of large numbers), and satisfy LDPs at speed n (by

Remark 2.7.6 and the contraction principle). Since by assumption, $\{V_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $s_n \ll n$ with GRF J_X, by Lemma 1.1.9 the sequence $\{V_n' := V_n W_n'\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF J_X. Since the GRF of $\{W_n'\}_{n\in\mathbb{N}}$ has a unique minimum at $m := (\lambda \mathcal{M}_2)^{1/2}$, by Lemma 1.1.9 $\{V_n' W_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF $J_X(\cdot/m)$, which coincides with the expression in (2.18). □

Remark 2.9.4. Recall that in the sublinear/linear regime, when $q = 2$ the contraction principle cannot be applied because the LDP of Theorem 2.5.1 holds with respect to the q-Wasserstein topology only for $q \in [1, 2)$. Moreover, the LDP for $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ of (2.45) as stated in Proposition 2.7.8 also holds with respect to the q-Wasserstein topology only for $q < 2$. However, in the case $q = 2$, (2.72) still establishes a variational problem that explicitly relates the rate function $\mathbf{J}_{2,\lambda}$ of (2.56) to the rate function $\mathcal{H}_\lambda$ of (2.4), in the same manner as if the contraction principle were applicable. We claim that this inconvenient gap is related to a more fundamental obstacle. To illustrate this concretely, consider the case where $X_1, X_2, \ldots$ are i.i.d. exponential random variables with mean 1. It is known that:

1. Due to Sanov's theorem, the sequence of empirical measures $\{\frac{1}{n}\sum_{i=1}^n \delta_{X_i}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\mathbb{R})$ with GRF $H(\cdot\| \operatorname{Exp}(1))$, where we write Exp(1) to denote the exponential distribution with mean 1.

2. Due to Cramér's theorem [21, Theorem 2.2.1], the sequence of empirical means $\{\frac{1}{n}\sum_{i=1}^n X_i\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ with GRF

$$L(\beta) := \beta - \log \beta - 1.$$

3. An explicit calculation establishes the expression

$$L(\beta) = \inf_{\mu \in \mathcal{P}(\mathbb{R})} \left\{ H(\mu \| \operatorname{Exp}(1)) : \int_{\mathbb{R}} x \, d\mu = \beta \right\}.$$

Note that if the map $\mu \mapsto \int_{\mathbb{R}} x \, d\mu$ were continuous, then the LDP of point 2. above would follow from point 1., point 3., and an application of the contraction principle. However, the map $\mu \mapsto \int_{\mathbb{R}} x \, d\mu$ is *not* continuous with respect to the weak topology on probability measures. Moreover, the result of [81, Theorem 1.1] applied to the exponential distribution indicates that the LDP of point 1. *does not* hold with respect to the 1-Wasserstein topology. This suggests that the apparently cryptic transition at $q = 2$ in the nature of the proof of Theorem 2.5.2 and the result of Theorem 2.5.1 is in a sense a manifestation of a more common sticking point in large deviations theory. In other words, even in the simple setting of i.i.d. random variables, the continuity required by the contraction principle fails to hold, but the consequences (i.e., a large deviation principle and a variational formula for the rate function) *do* still hold in many instances.

2.10 An approximate contraction principle

In this section, we recall the approximate contraction principle established in Section 6.2 of [14], and establish a corollary of it that we use in the proof of the LDP in the linear regime in Section 2.7.4. Recall that given $a, b \in \mathbb{R}$, we will use $a \vee b$ and $a \wedge b$ to denote $\max(a, b)$ and $\min(a, b)$, respectively.

Let Σ be a Polish space. Let $\mathcal{X}$ be a separable Banach space with topological dual space $\mathcal{X}^*$, and let $\langle \cdot, \cdot \rangle : \mathcal{X}^* \times \mathcal{X} \to \mathbb{R}$ denote the associated dual pairing. Fix a continuous map $\mathbf{c} : \Sigma \to \mathcal{X}$, and let $\{\mathcal{L}_n\}_{n \in \mathbb{N}}$ be a sequence of $\mathcal{P}(\Sigma)$-valued random elements. For

$r \in (0, \infty]$ and a continuous function $W : \Sigma \to \mathbb{R}$ such that $\mathbb{P}$-a.s., $\int_\Sigma (W(x) \vee 0) \mathcal{L}_n(dx) < \infty$ for all $n \in \mathbb{N}$, let

$$\Lambda_r(W) := \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[e^{n\left(\int_\Sigma W(x)\mathcal{L}_n(dx) \wedge r\right)}\right], \tag{2.73}$$

and also let

$$\bar{\Lambda}(W) := \sup_{r>0} \Lambda_r(W). \tag{2.74}$$

We introduce the "domain" of $\bar{\Lambda}$, defined in the following manner: let

$$\mathcal{D} := \{\alpha \in \mathcal{X}^* : \bar{\Lambda}(\langle \alpha, \mathbf{c}(\cdot)\rangle) < \infty\}, \tag{2.75}$$

$$\mathcal{D}_\circ := \{\alpha \in \mathcal{X}^* : \exists p > 1, p\alpha \in \mathcal{D}\}.$$

Then, for $x \in \mathcal{X}$, let

$$F(x) := \sup_{\alpha \in \mathcal{D}_\circ} \langle \alpha, x\rangle. \tag{2.76}$$

Lastly, for $n \in \mathbb{N}$, we define the $\mathcal{X}$-valued random variable $C_n := \int_\Sigma \mathbf{c}(x)\, \mathcal{L}_n(dx)$.

Proposition 2.10.1 (Proposition 6.4 of [14]). *Suppose that:*

1. *$\{\mathcal{L}_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\Sigma)$ at speed n with GRF I_0;*

2. *$\{\mathcal{L}_n, C_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\Sigma) \times \mathcal{X}$ at speed n with some convex GRF $\mathbb{I}$;*

3. *for any sequence $\{W_n\}_{n\in\mathbb{N}}$, in the set*

$$\{V + \langle \alpha, \mathbf{c}(\cdot)\rangle : V : \Sigma \to \mathbb{R} \text{ continuous and bounded}, \alpha \in \mathcal{D}_\circ\} \tag{2.77}$$

such that $W_n \downarrow W_\infty$ to a limit $W_\infty : \Sigma \to \mathbb{R}$ that is continuous and bounded above, we have

$$\limsup_{n\to\infty} \bar{\Lambda}(W_n) \le \bar{\Lambda}(W_\infty). \tag{2.78}$$

Then, we have the following representation for the GRF $\mathbb{I}$, for all $\mu \in \mathcal{P}(\Sigma)$ and $s \in \mathcal{X}$:

$$\mathbb{I}(\mu, s) := \begin{cases} I_0(\mu) + F\left(s - \int_\Sigma \mathbf{c}\, d\mu\right) & \text{if } I_0(\mu) < \infty, \\ +\infty & \text{else,} \end{cases} \tag{2.79}$$

with F as defined in (2.76).

The following corollary considers a special case where the conditions of Proposition 2.10.1 can be easily verified.

Corollary 2.10.2. *Let Σ be a Polish space and $\mathcal{X}$ be a separable Banach space. Suppose that $\{k_n\}_{n\in\mathbb{N}}$ grows sublinearly with rate $\lambda \in (0, 1]$, each $\mathcal{L}_n$ is the empirical measure of k_n i.i.d. Σ-valued random variables $\eta_1, \ldots, \eta_{k_n}$ with common distribution μ (that does not depend on n), and for continuous $W : \Sigma \to \mathbb{R}$, define*

$$\widehat{\Lambda}(W) := \log \mathbb{E}[e^{\lambda^{-1} W(\eta_1)}] \tag{2.80}$$

Also, let $\mathbf{c} : \Sigma \to \mathcal{X}$ be a continuous map such that 0 lies in the interior $\mathcal{D}^\circ$ of the set

$$\mathcal{D} := \left\{\alpha \in \mathcal{X}^* : \widehat{\Lambda}(\langle \alpha, \mathbf{c}(\cdot)\rangle) < \infty\right\}, \tag{2.81}$$

and let $C_n := \int_\Sigma \mathbf{c}(x)\mathcal{L}_n(dx)$. Then $\{\mathcal{L}_n, C_n\}$ satisfies an LDP with GRF given by (2.79), where $I_0(\nu) := \lambda H(\nu|\mu)$ and $F(x) = \sup_{\alpha \in \mathcal{D}^\circ} \langle \alpha, x\rangle$.

Proof. We start by verifying the conditions of Proposition 2.10.1.

The fact that $\{\mathcal{L}_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\Sigma)$ at speed k_n with GRF $H(\cdot|\mu)$ follows from Sanov's theorem on the Polish space $\mathcal{P}(\Sigma)$ (see, e.g., Theorem 6.6.9 of [45]). Since $k_n/n \to \lambda$, this immediately implies that $\{\mathcal{L}_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n with GRF

$I_0(\cdot) := \lambda H(\cdot|\mu)$, so condition 1. of Proposition 2.10.1 is satisfied.

Next, note that $(\mathcal{L}_n, C_n) = \frac{1}{k_n}\sum_{j=1}^{k_n}(\delta_{\eta_j}, \mathbf{c}(\eta_j)) \in \mathcal{P}(\Sigma) \times \mathcal{X}$. Therefore, it follows from Cramér's theorem on any locally convex topological space (see, e.g., Theorem 6.1.3 and Corollary 6.16 of [21]), with an appeal to the assumption that 0 lies in $\mathcal{D}^\circ$, the interior of the set $\mathcal{D}$ of (2.81), that $\{(\mathcal{L}_n, C_n)\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\Sigma) \times \mathcal{X}$ at speed k_n with a convex GRF. Since $k_n/n \to \lambda \in (0, 1]$, condition 2. of Proposition 2.10.1 is satisfied.

As for condition 3., first consider any sequence $\{W_n\}_{n\in\mathbb{N}}$ such that for each $n \in \mathbb{N}$, $W_n = V_n + \langle\alpha_n, \mathbf{c}(\cdot)\rangle$ for V_n bounded and continuous, and $\alpha_n \in \mathcal{D}$. Due to the assumption that $\alpha_1 \in \mathcal{D}_\circ$ and the boundedness of V_1, we have $\widehat{\Lambda}(W_1) < \infty$, so if $W_n \downarrow W_\infty$, then by the dominated convergence theorem,

$$\lim_{n\to\infty} \widehat{\Lambda}(W_n) = \widehat{\Lambda}(W_\infty). \tag{2.82}$$

To complete the proof, it clearly suffices to show that $\mathcal{D} = \mathcal{D}$, for the domains $\mathcal{D}$ and $\mathcal{D}$ defined in (2.81) and (2.75), respectively, and that relation (2.82) holds when $\widehat{\Lambda}$ is replaced with $\bar{\Lambda}$. In turn, to show the latter, it suffices to prove that $\bar{\Lambda} = \lambda\widehat{\Lambda}$. Indeed, note that for any continuous $W : \Sigma \to \mathbb{R}$ such that $\mathbb{P}$-a.s., $\int_\Sigma (W(x) \vee 0)\mathcal{L}_n(dx) < \infty$ for all $n \in \mathbb{N}$, by the definitions of Λ_∞ and $\bar{\Lambda}$ from (2.73) and (2.74), respectively, the i.i.d. assumption on $\{\eta_i\}_{i=1}^{k_n}$ and the fact that $k_n/n \to \lambda$, we have

$$\begin{aligned}
\widehat{\Lambda}(W) &= \lambda^{-1}\Lambda_\infty(W) \\
&\geq \lambda^{-1}\bar{\Lambda}(W) \\
&\geq \lambda^{-1}\lim_{R\to\infty}\bar{\Lambda}(W \wedge R) \\
&= \lambda^{-1}\lim_{R\to\infty}\Lambda_\infty(W \wedge R) \\
&= \lim_{R\to\infty}\widehat{\Lambda}(W \wedge R)
\end{aligned}$$

$$= \widehat{\Lambda}(W),$$

where the first inequality uses the elementary observation that $\Lambda_r \le \Lambda_\infty$ for every r implies $\bar{\Lambda} \le \Lambda_\infty$, and the second equality uses this observation along with the fact that the converse inequality also holds when both functions are evaluated at $W \wedge R$, since then $\int_\Sigma (W(x) \wedge R)\mathcal{L}_n(dx) \le R$ implies $\Lambda_\infty(W \wedge R) = \Lambda_R(W \wedge R) \le \bar{\Lambda}(W \wedge R)$. Thus, we have shown $\widehat{\Lambda} = \lambda^{-1}\Lambda_\infty$, which completes the verification of condition 3.

The corollary then follows from Proposition 2.10.1 and the identity $\mathcal{D}^\circ = \mathcal{D}_0$.

□

2.11 Properties of the function $\mathcal{J}$ of (2.33)

We recall here the definition of the subdifferential of a convex function. Let $f : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ be a convex function. Then the subdifferential at a point $x \in \mathbb{R}^n$ is defined to be

$$\partial f(x) := \{s \in \mathbb{R}^n : f(y) \ge f(x) + \langle s, y - x\rangle \quad \text{fo all} \quad y \in \mathbb{R}^n\}.$$

Note that when f is differentiable at x, the subdifferential $\partial f(x)$ has only one element, which is the gradient $\nabla f(x)$. For further discussion of subdifferentials, the reader is referred to Chapter X in [36].

Proof of Lemma 2.6.8. Before proving the individual properties, we make a general observation. Recall that D_V is the domain of V. For $u, v \in \mathbb{R}_+$, define $H : \mathbb{R}_- \times \mathbb{R} \to \mathbb{R}$ to be

$$H(s,t) := su + tv - \log \int_{D_V} e^{sV(x)+tx^2} dx. \tag{2.83}$$

By definition, H is differentiable. We first show that H is concave. For this, by (2.83) it suffices to show that $(s,t) \mapsto \log \int_{D_V} e^{sV(x)+tx^2} dx$ is convex. Indeed, by Hölder's inequality, for $\lambda \in (0,1)$ and (s_1,t_1), $(s_2,t_2) \in \mathbb{R}_- \times \mathbb{R}$, we see that

$$\begin{aligned}
&\log \int_{D_V} e^{(s_1\lambda + s_2(1-\lambda))V(x)+(t_1\lambda+t_2(1-\lambda))x^2} dx \\
&\quad \le \log\left(\left(\int_{D_V} e^{s_1\lambda V(x)+t_1\lambda x^2} dx \right)^{1/\lambda} \left(\int_{D_V} e^{s_2(1-\lambda)V(x)+t_2(1-\lambda)x^2} dx \right)^{1/(1-\lambda)} \right) \\
&\quad = \frac{1}{\lambda} \log \int_{D_V} e^{s_1\lambda V(x)+t_1\lambda x^2} dx + \frac{1}{1-\lambda} \log \int_{D_V} e^{s_2(1-\lambda)V(x)+t_2(1-\lambda)x^2} dx.
\end{aligned}$$

In particular, H is strictly concave if and only if $V \not\equiv 0$ in the domain of V.

We now turn to the proof of property 1. Let $(u,v) \in \mathbb{R}_+^2$ be in the interior of the domain of $\mathcal{J}$. Then by [36, Theorem 1.4.2], since $\mathcal{J}$ is convex, the subdifferential at (u,v) is nonempty, $\partial\mathcal{J}(u,v) \neq \emptyset$. Therefore there exists $s \in \mathbb{R}_-$, $t \in \mathbb{R}$ such that $(s,t) \in \partial\mathcal{J}(u,v)$. Since $\mathcal{J}$ and $(s,t) \mapsto \log \int_{D_V} e^{sV(x)+tx^2} dx$ are both convex, both functions satisfy the condition in [36, Theorem 1.4.1]. Hence, we see that there exist $s \in \mathbb{R}_-$ and $t \in \mathbb{R}$ such that $\mathcal{J}(u,v) = H(s,t)$. If H is strictly concave, then automatically (s,t) is the unique maximizer in the supremum of $\mathcal{J}(u,v)$ in (2.33). If H is not strictly concave, then $V \equiv 0$ in its domain, and so by (2.33), $\mathcal{J}(v) = \sup_{s<0, t\in\mathbb{R}_+} \left\{ su + tv - \log(\int_{D_V} e^{tx^2} dx) \right\}$, from which it clearly follows that $(0,t)$ is the unique maximizer (since $v \geq 0$). Next, again by the last display and [36, Theorem 1.4.1], we see that $(u,v) \in \partial \log \int_{D_V} e^{sV(x)+tx^2} dx$. By the definition of a subdifferential, we conclude that

$$u \geq \frac{d}{ds} \log\left(\int_{D_V} e^{sV(x)+tx^2} dx \right) = \int_{D_V} V(x) \nu_{s,t}(dx) = M_V(\nu_{s,t}),$$

$$v = \frac{d}{dt}\log\left(\int_{D_V} e^{sV(x)+tx^2}dx\right) = \int_{D_V} x^2\nu_{s,t}(dx) = M_2(\nu_{s,t}), \tag{2.84}$$

where (2.84) holds due to the differentiability of $t \mapsto \log\int_{D_V} e^{sV(x)+tx^2}dx$. This proves property 1.

To see why property 2 holds, note that by the duality of the Legendre transform [84, Equation (12)], the minimizer of $\mathcal{J}(1,\cdot)$ is obtained at m such that

$$\frac{d}{dv}\mathcal{J}(1,v)\bigg|_{v=m} = t(m) = 0. \tag{2.85}$$

Substituting this relation back into (2.84), we obtain $1 = M_V(\nu_{s(m),0})$ and $m = M_2(\nu_{s(m),0})$. Setting $b^* = -s(m) > 0$ and observing that $\nu_{s,0} = \mu_{V,-s}$ for any s, we conclude that $1 = M_V(\nu_{s(m),0}) = M_V(\mu_{V,b^*})$ and $m = M_2(\nu_{s(m),0}) = M_2(\mu_{V,b^*})$. Now, (2.35) follows since the supremum on the right-hand side of (2.35) is attained at $1 = M_V(\nu_{s(m),0})$ and (2.84) is uniquely solvable. This proves property 2.

For the remaining properties, first note that by the duality of the Legendre transform, the supremum in the definition (2.33) of $\mathcal{J}(u,v)$, when finite, is attained at $(\partial_u\mathcal{J}(u,v), \partial_v\mathcal{J}(u,v)) \in \mathbb{R}_- \times \mathbb{R}$. This proves property 4. By the convexity of $v \mapsto \mathcal{J}(1,v)$ and the fact that it uniquely attains its minimum at the value m defined above, we see that $\partial_v\mathcal{J}(1,v) > 0$ if $v > m$ and $\partial_v\mathcal{J}(1,v) < 0$ if $0 < v < m$. This proves property 3, and completes the proof of the lemma. □

Chapter Three

Annealed sharp large deviation estimates

3.1 Introduction

Asymptotic convex geometry is concerned with properties of convex bodies in high dimension. Due to a useful representation of uniform random variables on ℓ_p^n, a series of work on the volume of the intersections of two ℓ_p^n balls was studied in [13,73,75,76]. Recently, [42] extends the study of volumetric properties to Orlicz balls. They study the volume ratio and relate the uniform measure on an Orlicz ball to Gibbs measures. On the other hand, the q-norm of ℓ_p^n balls are studied in [43] and [44]. A refinement of the previous results is obtained in [47]. Finally, the random projections of vector sampled from a convex body is also of interests [3,4,34]. All these properties of convex bodies are important in asymptotic convex geometry. The aim of this paper is to use the technique of sharp large deviation estimates to shed insights into all problems mentioned above.

Limit theorems in probability, including the concentration inequality, the central limit theorem and the large deviation principles, have well been used to study high-dimensional convex bodies. Concentration inequality for norms of high-dimensional measures [8, 28, 29, 51, 52, 67, 68] is closely connected with the Kannan-Lovász-Simonovits conjecture. The CLT's and LDP's for norms of ℓ_p^n balls is studied in [43, 44, 49]. Antilla, Ball and Perissinnnaki showed in [8] that the "thin-shell condition" implies that the projection along most directions are approximately Gaussian. Klartag [52] later showed the central limit theorem for convex sets that uniform measures on compact convex sets satisfy the thin-shell condition. The result of [8] was further extended by Meckes to more general measures and a quantitative bound is obtained [61]. Apart from the central limit theorem, the large deviation principle was introduce in asymptotic convex geometry in [34] to study the tails of random projections of ℓ_p^n balls. We see that probability theory helps with finding new insights in convex bodies. Recently, the sharp large deviation estimate is also used to study the random projection and norms of ℓ_p^n balls [42,47,58] but these are still less well-studied.

In this chapter, we introduce another aspect of probability theory to asymptotic convex geometry. The refinement for large deviation principles was first shown by Bahadur and Ranga-Rao [9]. For a sequence of independent and identically distributed (i.i.d.) random variables $\{X_i\}_{i\in\mathbb{N}}$ under suitable condition on each X_i, we have the following estimate

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq a\right) = \frac{e^{-n\mathbb{I}(a)}}{\sigma_a\tau_a\sqrt{2\pi n}}(1+o(1)). \tag{3.1}$$

The first results of sharp large deviation estimates in asymptotic convex geometry was given in [58] where they obtained a sharp estimate of the tail of quenched random projections of ℓ_p^n balls and spheres. Another instance of sharp large deviation estimates is in [42]. Kabluchko and Prochno derived various volumetric of Orlicz balls using the technique of sharp large deviation estimates. Very recently, Kaufmann [47] shows the sharp tail estimate of q-norms of uniform vectors distributed on ℓ_p^n balls and spheres. We extend and generalize in this article many of the results mentioned above.

Our first contribution is a sharp estimate of the volume of the intersections of two Orlicz balls, Theorem 3.2.12. The sharp estimate is then used to resolve an open problem for volume ratios in [42] under the critical case. Secondly, we find both the upper tail and lower tail of Orlicz norms of uniform measures on Orlicz balls, Theorem **??** and **??**. The LDP for Orlicz balls are given in [49] while here we obtain the sharp large deviation estimates. We also extend the results in [47], where Kaufmann obtained the upper tail of q-norm for ℓ_p^n balls. Finally, we propose an "asymptotic thin-shell estimate" which is a sufficient condition to obtain a sharp large deviation estimates for random projections. A sufficient condition for the LDP of random projections was given previously in [49] and the "asymptotic thin-shell estimate" is a refinement of the LDP results.

3.2 Main results

3.2.1 Sharp large deviation estimates for random projections

Our first main result is sharp large deviation estimates for multidimensional random projections of high-dimensional vectors. For $n \in \mathbb{N}$, let $X^{(n)}$ be a random vector in $\mathbb{R}^n$. For $k \in \mathbb{N}$, let I_k denote the $k \times k$ identity matrix and for $n > k$ denote the Stiefel manifold of k-frames in $\mathbb{R}^n$ to be

$$\mathbb{V}_{n,k} := \{A \in \mathbb{R}^{n \times k} : A^T A = I_k\}.$$

For $k_n < n$, let $\mathbf{A}_{n,k_n}$ be an $n \times k_n$ random matrix drawn from the Haar measure on the Stiefel manifold $\mathbb{V}_{n,k_n}$. We consider the orthogonal projection of $X^{(n)}$ onto the k_n-dimensional subspace spanned by the column of $\mathbf{A}_{n,k_n}$ for different regimes of the sequence $\{k_n\}_{n \in \mathbb{N}}$, denoted $\mathbf{A}^T_{n,k_n} X^{(n)}$.

Previously, Antilla, Ball and Perissinnaki [8] showed that when $X^{(n)}$ satisfies a "thin-shell condition" , the random projections are close to Gaussian, thus leading to a central limit theorem for random projections. Subsequently, Kim, Liao and Ramanan [49] proposed an "asymptotic thin shell condition" and showed that this implies that the sequence of random projections satisfy an LDP. In this article, we aim to find a sufficient condition under which we obtain a sharp LDP for random projections.

We state our main assumption:

Assumption 3.2.1 (Thin shell density estimate)**.** The sequence $\{X^{(n)}\}_{n \in \mathbb{N}}$ satisfies the following property. For each $n \in \mathbb{N}$, $\|X^{(n)}\|_2^2/n$ has a density f^n. Moreover, the sequence $\{f_n\}_{n \in \mathbb{N}}$ satisfies a (J, h, β)-representation by which we mean there exist a lower semicontinuous function $J : \mathbb{R} \to [0, \infty]$, a differentiable function $h : \mathbb{R} \to [0, \infty)$, $\alpha \in \mathbb{R}$ and a domain D_J such that

$$f^n(x) = n^\beta h(x) e^{-nJ(\sqrt{x})}(1 + o(1)), \quad \forall x \in D_J, \tag{3.2}$$

uniformly on any compact neighborhood of x in D_J.

Remark 3.2.2. Assumption 3.2.1 implies the asymptotic thin shell condition introduced in [49], i.e., the sequence $\{\left\|X^{(n)}\right\| / \sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP with rate function J.

The additional condition in Assumption 3.2.1 requires that the density of $\|X^{(n)}\|_2^2/n$ takes the form given in (3.2). This allows us to obtain the sharp large deviation estimates.

In Sections 3.2.1.1–3.2.1.3 below, we report results for three different scaling regimes of the projection dimension k_n: constant ($k_n \equiv k$), α-sublinear $k_n \ll n^\alpha$ and (λ, α)-linear ($k_n = \lambda n + r_n$, $\lambda \in (0,1)$ and $r_n \ll n^\alpha$). In Section 3.2.1.4, we present examples that satisfy the sharp density estimate, which include the uniform measure on normalized ℓ_p^n balls, $p \geq 2$, and Orlicz balls.

3.2.1.1 Sharp estimates for projections in the constant regime

In the constant regime, where $k_n \equiv k$, the sequence $\{n^{-1/2}\mathbf{A}_{n,k}^T X^{(n)}\}_{n\in\mathbb{N}}$ was shown in [49, Theorem 2.8] to satisfy an LDP in $\mathbb{R}^k$ with GRF $I_{\mathbf{A}X,k} : \mathbb{R}^k \mapsto [0,\infty]$ defined by

$$I_{\mathbf{A}X,k}(x) := \inf_{c\in(0,1)} \left\{ J\left(\frac{\|x\|_2}{c}\right) - \frac{1}{2}\log\left(1 - c^2\right) \right\}. \tag{3.3}$$

Define $\bar{J}^{\text{con}} : \mathbb{R}^k \times \mathbb{R}_+ \to [0, \infty]$ to be

$$\bar{J}^{\text{con}}(x, y) := \begin{cases} -\frac{1}{2}\log\left(1 - \|x\|_2^2\right) + J\left(\sqrt{y}\right), & \|x\|_2 < 1, \\ \infty, & \text{otherwise.} \end{cases} \tag{3.4}$$

and for a measurable set $\Upsilon \subset \mathbb{R}^k$, define

$$\mathcal{D}_\Upsilon := \{(x, y) \in \mathbb{R}^k \times \mathbb{R}_+ : x\sqrt{y} \in \Upsilon, \|x\|_2 \le 1\}. \tag{3.5}$$

Theorem 3.2.3 (Constant $k_n \equiv k$)**.** *Fix $k \in \mathbb{N}$. Suppose $\{X^n\}_{n\in\mathbb{N}}$ satisfies Assumption 3.2.1 with parameters (J, h, β). Let $\Upsilon \subset \mathbb{R}^k$ be a measurable set such that $\inf_{x\in\Upsilon} I_{\mathbf{A}X,k}(\Upsilon) < \infty$. Suppose there exists $(x_\Upsilon, y_\Upsilon) \in \partial\mathcal{D}_\Upsilon \subset \mathbb{R}^k \times \mathbb{R}$ that uniquely attains the infimum of $\bar{J}^{\text{con}}$ over $\mathcal{D}_\Upsilon$. Moreover, suppose $\mathcal{D}_\Upsilon$ is smooth at (x_Υ, y_Υ) and $y_\Upsilon \in D_J$. Then there exists a constant $\gamma^{\text{con}}_{\Upsilon,k} \in \mathbb{R}_+$ such that*

$$\mathbb{P}\left(n^{-1/2}\mathbf{A}^T_{n,k}X^{(n)} \in \Upsilon\right) = n^{\beta-1}C^{\text{con}}_{\Upsilon,k}e^{-nI_{\mathbf{A}X,k}(\Upsilon)}(1 + o(1)), \tag{3.6}$$

where

$$C^{\text{con}}_{\Upsilon,k} := \frac{h(y_\Upsilon)}{\gamma^{\text{con}}_{\Upsilon,k}}\left(1 - \|x_\Upsilon\|_2^2\right)^{-k/2-1}.$$

We first have the following observation on the rate function $I_{\mathbf{A}X,k}$, defined in (3.3).

Lemma 3.2.4. *1. $x \mapsto I_{\mathbf{A}X,k}(x)$ is spherical symmetric.*

2. $I_{\mathbf{A}X,k}(x) = \hat{I}_{\mathbf{A}X,k}(\|x\|_2)$, where $\hat{I}_{\mathbf{A}X,k} : \mathbb{R}_+ \to [0, \infty]$ is an increasing function defined to be

$$\hat{I}_{\mathbf{A}X,k}(r) := \inf_{0<c<1}\left\{-\frac{1}{2}\log(1 - c^2) + J\left(\frac{r}{c}\right)\right\}$$

The proof of Theorem 3.2.3 is given in Section 3.3.4.1 and Lemma 3.2.4 is deferred to Section 3.7.

Remark 3.2.5. We contrast (3.6) with the sharp tail estimates for the quenched 1-dimensional projection ℓ_p^n ball obtained in [58]. Let $X^{(n,p)}$ be uniformly distributed on the ℓ_p^n ball

$$\mathbb{B}_p^n := \{x \in \mathbb{R}^n : \sum_{i=1}^n |x_i|^p \leq n\},$$

for $p \geq 2$. Given $\theta = \{\theta^{(n)}\}_{n \in \mathbb{N}}$ with $\theta^{(n)} \in S^{n-1}$, let $\mathbb{P}_\theta$ denote the probability $\mathbb{P}$ conditioned on $\mathbf{A}_{n,1} = \theta^{(n)}$. Let σ be any probability measure on $\prod_{n\in\mathbb{N}} S^{n-1}$ such that its image σ^{n-1} under the coordinate projection coincides with the unique rotation invariant measure on S^{n-1}. Then it was shown in [58] that for every $a > 0$, there exist constant $\kappa_{a,p} \in \mathbb{R}_+$ and mappings $C_a^{n,p}, R_a^{n,p} : \mathbb{R}^n \to \mathbb{R}$ such that

$$\mathbb{P}_\theta\left(n^{-1/2}\mathbf{A}_{n,1}^T X^{(n,p)} > a\right) = \frac{C_a^{n,p}(\theta^{(n)})}{\kappa_{a,p}\sqrt{2\pi n}} e^{-n\mathbb{I}_p(a) + \sqrt{n} R_a^{n,p}(\theta^{(n)})}(1 + o(1)).$$

Note that in the case of ℓ_p^n balls, $p \geq 2$, we show in Section 3.2.1.4 that Assumption 3.2.1 is satisfied with $\beta = 1/2$. We can see from Theorem 3.2.3, the main difference is in the exponent of the refined estimates. We do not have an extra term here because we are averaging over all directions but not quenching on a particular direction sequence.

Another immediate observation from the estimate above is that in the constant regime, only the constant $C_{\Upsilon,k}^{\mathrm{con}}$ depends on the projection dimension k. We will see that this is not the case for the sublinear and linear regimes.

Next, we consider different choice of sets Υ in Theorem 3.2.3. In particular, we consider the norm of the projection $n^{-1/2}\mathbf{A}_{n,k}^T X^{(n)}$. For $q \in [1, \infty)$ and $a \in \mathbb{R}_+$, define

$$\Upsilon_{q,a} := \{x \in \mathbb{R}^k : \|x\|_q > a\}. \tag{3.7}$$

One thing to note here is that the proof used in Theorem 3.2.3 does not work in the following theorem since the assumption for the set $\Upsilon_{q,a}$ is not satisfied and a slight different approach should be taken.

For $i = 1, \ldots, k$, let $e_i \in \mathbb{R}^k$ be the standard basis vector. From Lemma 3.2.4, we see that for different q the set of minimizers $\mathcal{W}_{q,k}(a)$ of $I_{\mathbf{A}X,k}$ over $\Upsilon_{q,a}$ have fundamentally different structure:

1. For $q \in [1,2)$, $\mathcal{W}_{q,k}(a) = \{\frac{a}{k^{1/q}} \sum_{i=1}^k \iota_i e_i : \iota_i \in \{-1,1\}\}$, consists of 2^k distinct points.
2. For $q = 2$, $\mathcal{W}_{2,k}(a) = \{x \in \mathbb{R}^k : \|x\|_2 = a\}$ is a $(k-1)$-dimensional surface.
3. For $q \in (2,\infty)$, $\mathcal{W}_{q,k}(a) = \{\iota e_i : \iota = \{-1,1\}, i = 1,\ldots,k\}$ consists of $2k$ distinct points.

Theorem 3.2.6. *Suppose $\{X^n\}_{n\in\mathbb{N}}$ satisfies Assumption 3.2.1 with parameters (J,h,β). Assume that $\inf_{u\in\Upsilon_{q,k}} I_{\mathbf{A}X,k}(u) < \infty$.*

1. *For $q \neq 2$, let $w_{q,a}$ be any point in $\mathcal{W}_{q,k}(a)$. Suppose $(x_{q,a}, y_{q,a})$ attain the minimum of $I_{\mathbf{A}X,k}(w_{q,a})$ in (3.3) and $y_{q,a} \in D_J$. Let $C^{\text{con}}_{a,q,k}$ be defined as in Theorem 3.2.3 with (x_a, y_a) replaced by $(x_{q,a}, y_{q,a})$.*

   (a) If $q \in [1,2)$, then

$$\mathbb{P}\left(n^{-1/2}\left\|\mathbf{A}_{n,k}^T X^{(n)}\right\|_q > a\right) = 2^k n^{\beta-1} C^{\text{con}}_{a,q,k} e^{-nI_{\mathbf{A}X,k}(\Upsilon_{q,a})}(1+o(1)). \tag{3.8}$$

   (b) If $q \in (2,\infty)$, then

$$\mathbb{P}\left(n^{-1/2}\left\|\mathbf{A}_{n,k}^T X^{(n)}\right\|_q > a\right) = 2kn^{\beta-1} C^{\text{con}}_{a,q,k} e^{-nI_{\mathbf{A}X,k}(\Upsilon_{q,a})}(1+o(1)). \tag{3.9}$$

2. *For $q = 2$, let $(r_{2,a}, y_{2,a})$ attain the infimum of the 1-dimensional*

$$I_{\mathbf{A}X,k}(a) := \inf_{r\in\mathbb{R},|r|\leq 1, y\in\mathbb{R}_+, r\sqrt{y}=a,} \left\{-\frac{1}{2}\log(1-|r|^2) + J(\sqrt{y})\right\}.$$

Then, there exists a constant $\gamma_{2,a}^{\text{con}}$ such that

$$\mathbb{P}\left(n^{-1/2}\left\|\mathbf{A}_{n,k}^T X^{(n)}\right\|_2 > a\right) = n^{\beta+\frac{k-3}{2}} C_{a,2,k}^{\text{con}} e^{-nI_{\mathbf{A}X,k}(\Upsilon_{2,a})}(1+o(1)), \tag{3.10}$$

where

$$C_{a,2,k}^{\text{con}} := \frac{k\pi}{\Gamma(k/2+1)2^{k/2-1/2}} \frac{h(y_{2,a})}{\gamma_{2,a}^{\text{con}}} \left(1-\left|r_{2,a}\right|_2^2\right)^{-k/2-1}.$$

In the special case when $k = 1$, the three expressions in (3.8), (3.9) and (3.10) coincide because there is no difference between $\Upsilon_{q,a}$ for different q when $k = 1$. The proof of Theorem 3.2.6 is given in Section 3.3.4.2. The proof is a modification of the proof of Theorem 3.2.3. In the first two cases, $q \neq 2$, we first isolate each minimizer, apply Theorem 3.2.3 and then sum up each tail estimate using symmetry. On the other hand, when $q = 2$, we have a whole surface of minimizers. Here, we instead utilize the spherical symmetry to reduce the problem to a 1-dimensional problem.

We will show in Section 3.2.1.4 that when $X^{(n)}$ is uniformly distributed on an Orlicz ball, $y_{2,a}$ always lies in the admissible domain D_J. In the theorem above, we see that the growth of the prefactor depends on k and the dependence on k is due to the non-uniqueness of the minimizer of $(x, y) \mapsto -\frac{1}{2}\log(1-\|x\|_2^2) + J(\sqrt{y})$ over $\mathcal{D}_\Upsilon$.

3.2.1.2 Sharp estimates for norms of projections in the sublinear regime

We next turn to sublinear regime. Define

$$Y^n_{2,k_n} := n^{-1/2}\|\mathbf{A}^T_{n,k_n}X^{(n)}\|_2 \tag{3.11}$$

When $k_n \ll n$, by [49, Theorem 2.11], the sequence $\{Y^n_{2,k_n}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ with GRF $\mathbb{J}_2^{\text{sub}} : \mathbb{R} \mapsto [0,\infty]$, defined by

$$\mathbb{J}_2^{\text{sub}}(x) := \inf_{c\in(0,1)} \left\{ -\frac{1}{2}\log(1-c^2) + J\left(\frac{x}{c}\right) \right\}. \tag{3.12}$$

Define $\bar{J}^{\text{sub}} : \mathbb{R}^2 \to [0,\infty]$ to be

$$\bar{J}^{\text{sub}}(x,y) := \begin{cases} -\frac{1}{2}\log\left(1-|x|^2\right) + J(\sqrt{y}), & |x|<1, \\ \infty, & \text{otherwise.} \end{cases} \tag{3.13}$$

and for $a > 0$ define

$$\mathcal{D}_a := \{(x,y) \in \mathbb{R}\times\mathbb{R}_+ : x\sqrt{y} > a, |x| \le 1\}. \tag{3.14}$$

Theorem 3.2.7 (α-sublinear regime)**.** *Fix a sequence $\{k_n\}_{n\in\mathbb{N}} \subset \mathbb{N}$ and $\alpha < 1$ such that $k_n/n^\alpha \to 0$ as $n\to\infty$. Suppose $\{X^n\}_{n\in\mathbb{N}}$ satisfies Assumption 3.2.1 with parameters (J,h,β). For $a>0$ such that $\inf_{x>a}\mathbb{J}_2^{\text{sub}}(x) < \infty$, suppose there exists $(x_a,y_a)\in\partial\mathcal{D}_a$ that uniquely attains the infimum of $\bar{J}^{\text{sub}}$ over $\mathcal{D}_a$. Moreover, suppose (x_a,y_a) is a smooth boundary point of $\mathcal{D}_a$ and $y_a \in D_J$. Then there exists a constant $\gamma_a^{\text{sub}}\in\mathbb{R}_+$ such that*

$$\mathbb{P}\left(Y^n_{2,k_n} > a\right) = C_a^{\text{sub}}(x_a,y_a)\, n^{\beta-\frac{3}{2}} k_n^{\frac{1}{2}} e^{-n\inf_{x>a}\mathbb{J}_2^{\text{sub}}(x) + k_n \log n M_n^{\text{sub}}(x_a)}(1+o(1)), \tag{3.15}$$

where

$$C_a^{\mathrm{sub}}(x,y) := \frac{\sqrt{2}h(y)}{\gamma_a^{\mathrm{sub}} x^{3/2}\left(1-x^2\right)},$$
$$M_n^{\mathrm{sub}}(x) := \frac{1}{2}\left(1 - \frac{\log k_n}{\log n} + \frac{1}{\log n}\log\left(\frac{x^2}{1-x^2}\right) + \frac{1-\frac{n}{k_n}}{\log n}\log\left(1-\frac{k_n}{n}\right)\right). \tag{3.16}$$

The proof is given in Section 3.3.4.3

3.2.1.3 Sharp estimates for norms of projections in the linear regime

Finally, suppose that k_n grows linearly with n at rate λ, $k_n/n \to \lambda$ as $n \to \infty$. It is shown in [49] that $\{Y^n_{2,k_n}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ with GRF $\mathbb{J}^{\mathrm{sub}}_{2,\lambda} : \mathbb{R} \mapsto [0,\infty]$, defined by

$$\mathbb{J}^{\mathrm{lin}}_{2,\lambda}(x) := \inf_{c\in(0,1)}\left\{\frac{\lambda}{2}\log\frac{\lambda}{c^2} + \frac{1-\lambda}{2}\log\frac{1-\lambda}{1-c^2} + J\left(\frac{x}{c}\right)\right\}. \tag{3.17}$$

Define $\bar{J}^{\mathrm{lin}} : \mathbb{R}^2 \to [0,\infty]$ to be

$$\bar{J}^{\mathrm{lin}}(x,y) := \begin{cases} \frac{\lambda}{2}\log\frac{\lambda}{x^2} + \frac{1-\lambda}{2}\log\frac{1-\lambda}{1-x^2} + J(\sqrt{y}), & |x|<1, \\ \infty, & \text{otherwise.} \end{cases} \tag{3.18}$$

Theorem 3.2.8 ((λ,α)-linear regime)**.** *Fix $\lambda \in (0,1)$ and $\alpha < 1$. Let $n, k_n, r_n \in \mathbb{N}$ be such that $k_n = \lambda n + r_n$ and $r_n/n^\alpha \to 0$ as $n \to \infty$. Suppose $\{X^n\}_{n\in\mathbb{N}}$ satisfies Assumption 3.2.1 with parameters (J, h, β). For $a > 0$ such that $\inf_{x>a}\mathbb{J}^{\mathrm{lin}}_{2,\lambda}(x) < \infty$, suppose there exists $(x_a, y_a) \in \partial\mathcal{D}_a$ that uniquely attains the infimum of $\bar{J}^{\mathrm{lin}}$ over $\mathcal{D}_a$. Moreover, suppose (x_a, y_a) is a smooth boundary point of $\mathcal{D}_a$ and $y_a \in D_J$. Then there exists a constant $\gamma_a^{\mathrm{lin}} \in \mathbb{R}_+$ such that*

$$\mathbb{P}\left(Y^n_{2,k_n} > a\right) = C^{\mathrm{lin}}_{a,\lambda}(x_a,y_a)n^{\beta-1}e^{-n\inf_{x>a}\mathbb{J}^{\mathrm{lin}}_{2,\lambda}(x) + r_n M^{\mathrm{lin}}_{a,\lambda}(x_a)}(1+o(1)). \tag{3.19}$$

where

$$C^{\mathrm{lin}}_{a,\lambda}(x,y) := \frac{\sqrt{2}(\lambda-\lambda^2)^{\frac{1}{2}} h(y)}{\gamma_a^{\mathrm{lin}} x^{\frac{3}{2}}\,(1-x^2)},$$

$$\begin{aligned} M^{\mathrm{lin}}_{\lambda}(x) := \frac{1}{2}\Bigg(&-\log\left(\lambda-\lambda^2\right)+\log\left(\frac{x^2}{1-x^2}\right)-\left(1+\frac{\lambda n}{r_n}\right)\log\left(1+\frac{r_n}{\lambda n}\right)\\ &-\left(1+\frac{(1-\lambda)n}{r_n}\right)\log\left(1-\frac{r_n}{(1-\lambda)n}\right)\Bigg). \end{aligned} \tag{3.20}$$

The proof of Theorem 3.2.8 are given in Section 3.3.4.4. In the sublinear and linear regimes, we see that in the exponent of the tail estimates, there is not only a term coming from the LDP, but also a sublinear (in n) correction term. Once again, similar to the constant regime in Theorem 3.2.6, when $X^{(n)}$ is uniformly distributed on an Orlicz ball, y_a lies in the admissible domain D_J for both the sublinear and linear regime.

3.2.1.4 Orlicz balls and Assumption 3.2.1

Let V be an Orlicz function as defined below.

Definition 3.2.9. A function $V : \mathbb{R} \to \mathbb{R}$ with $V(0) = 0$ is said to be an Orlicz function if it is convex, symmetric, that is, $V(x) = V(-x)$ for $x \in \mathbb{R}$, and satisfies $\lim_{x\to\infty} V(x) = \infty$.

For $R \in \mathbb{R}_+$, define the Orlicz ball

$$B^n_V(R) := \left\{ x \in \mathbb{R}^n : \sum_{i=1}^{n} V(x_i) \le nR \right\}. \tag{3.21}$$

For Orlicz functions V_1, V_2 with $V_1(x)/V_2(x) \to \infty$ as $x \to \infty$, define

$$\Lambda_{V_1,V_2}(t_1,t_2) := \log \int_{\mathbb{R}} e^{t_1 V_1(x)+t_2 V_2(x)} dx, \quad (t_1,t_2) \in \mathbb{R}_- \times \mathbb{R}, \tag{3.22}$$

and the Legendre transform of Λ,

$$\mathcal{J}_{V_1,V_2}(t_1,t_2) := \sup_{s_1<0, s_2\in\mathbb{R}} \{s_1 t_2 + s_2 t_2 - \Lambda(s_1,s_2)dx\}. \tag{3.23}$$

When $V_2(x) = x^2$ and V_1 is superquadratic, in the sense that $V_1(x)/x^2 \to \infty$ as $x \to \infty$, $\mathcal{J}_{V_1,x^2}$ was used in [49] to establish an LDP for the 2-norm of a random vector uniformly distributed on the Orlicz ball $B^n_{V_1}(1)$. Below, we provide a substantial generalization of this result that provides more refined asymptotics for more general V_2.

For $s \in \mathbb{R}_-$ and $t \in \mathbb{R}$ define

$$\mu_{s,t}(dx) := \frac{1}{Z_{s,t}} e^{sV_1(x)+tV_2(x)} dx \quad \text{where} \quad Z_{s,t} := \int_{\mathbb{R}} e^{sV_1(x)+tV_2(x)} dx, \tag{3.24}$$

which is finite because of the assumption that V_1 is convex and dominates V_2 at infinity. Moreover, for a function $V : \mathbb{R} \to \mathbb{R}$, we define the moment map to be

$$M_V(\mu_{s,t}) := \int_{\mathbb{R}} V(x) \mu_{s,t}(dx). \tag{3.25}$$

Since V_1 and V_2 are convex functions, the derivatives of V_1 and V_2 exist almost everywhere and we set for a.e. $(x_1,x_2) \in \mathbb{R}^2$

$$\mathcal{G}(x_1,x_2) := \left| V_1'(x_1)V_2'(x_2) - V_1'(x_2)V_2'(x_1) \right|. \tag{3.26}$$

We make the following assumptions.

Assumption 3.2.10. V_1 and V_2 are differentiable functions that satisfy the following conditions:

1. $V_1'(x)/V_2'(x)$ is strictly increasing in $\mathbb{R}_+$, which implies that $V_1(x)/V_2(x) \to \infty$ as $x \to \infty$.
2. There exist $r_1, r_2 > 0$ and such that $1/\mathcal{G}$, defined in (3.26), lies in $L^{r_1}(\mathcal{N}) \cap L^{r_2}(\mathcal{N}^c)$ for some neighborhood $\mathcal{N} \subset \mathbb{R}^2$ of the origin.
3. Further assume there exists $T \in [-\infty, 0]$ such that $M_{V_1}(\mu_{0,t})$, defined in (3.25), is finite for all $t < T$ and is infinite for $t \geq T$.

Recall $\mathcal{J}_{V_1,V_2}$ defined in (3.23). In the special case when $V_2(x) = x^2$, define for $z \in \mathbb{R}_+$,

$$J_{V,x^2}(z) := \mathcal{J}_{V,x^2}(1, z^2) - \Psi_V^*(1), \tag{3.27}$$

and

$$\Lambda_{V,x^2}(s_1, s_2) := \log \int_{\mathbb{R}} e^{s_1 V(x) + s_2 x^2} dx.$$

Let D_{V,x^2} be the admissible domain of $\mathcal{J}_{V_1,x^2}$ so that for $(x_1, x_2) \in D_{\mathcal{J}}$, $(t_1, t_2) = \nabla \Lambda_{V,x^2}(x_1, x_2)$ attains the supremum of $\mathcal{J}_{V_1,V_2}(x_1, x_2)$ in (3.23).

Theorem 3.2.11. *Let V be an Orlicz function. Suppose the pair of Orlicz functions (V, x^2) satisfies Assumption 3.2.10. For $n \in \mathbb{N}$, let $X^{n,V}$ be uniformly distributed on the Orlicz ball $B_V^n(1)$. Then there exist $\sigma > 0$ and $\tau < 0$ such that $\{X^{n,V}\}_{n \in \mathbb{N}}$ satisfies Assumption 3.2.1 with parameters, $J := J_V$, $\beta := 1/2$, and*

$$h(x) := \frac{\sigma\,|\tau|}{\sqrt{2\pi}} \frac{\det\left(\mathcal{H}\mathcal{J}_{V,x^2}(1,x)\right)^{1/2}}{\left|\partial_1 \mathcal{J}_{V,x^2}(1,x)\right|}, \quad D_J := \{x \in \mathbb{R}_+ : (1.x) \in D_{V,x^2}\},$$

and moreover, for $a > 0$ such that the infimum of $\bar{J}^{\mathrm{sub}}$, defined in (3.13), over $\mathcal{D}_a$ is finite, if (x_a, y_a) is the minimizer, then $y_a \in D_J$.

The proof of the theorem is given in Section 3.5.1. The constants τ and σ coincide with the constants τ_1 and σ_1 defined in Proposition 3.4.7. Theorem 3.2.11, when combined with the main theorems in Section 3.2.1, implies sharp large deviation estimates for annealed random projections for uniform random variables on Orlicz balls. This generalizes previous results in [49].

3.2.2 Volumetric properties of Orlicz balls

Given two Orlicz functions V_1 and V_2, we are interested in the asymptotic behavior of the volume of intersections $\left|B^n_{V_1}(R_1) \cap B^n_{V_2}(R_2)\right|$. Now we state our result on the asymptotic volume of intersections.

Theorem 3.2.12. *Let V_1 and V_2 be two Orlicz functions that satisfy Assumption 3.2.10 with $T \in [-\infty, 0]$. Fix $R_1 > 0$.*

1. *There exists a unique $\tau_1(R_1) < 0$ such that $R_1 = M_{V_1}(\mu_{\tau_1(R_1),0})$ and if $T > -\infty$, then there exists a unique $\tau_2(R_1) < 0$ such that $R_1 = M_{V_1}(\mu_{0,\tau_2(R_1)})$.*

2. *Define $m_1 = m_1(R_1) := M_{V_2}(\mu_{0,\tau_2(R_1)})$ if $T > -\infty$ and 0 otherwise. Also define $m_2 = m_2(R_1) := M_{V_2}(\mu_{\tau_1(R_1),0})$. Then, setting $\mathcal{J} = \mathcal{J}_{V_1,V_2}$, we have*

   $$\left|B^n_{V_1}(R_1) \cap B^n_{V_2}(R_2)\right|$$

$$= \begin{cases} \frac{1}{2\pi n} \frac{\det(\operatorname{Hess}\mathcal{J}(R_1,R_2))^{1/2}}{|\partial_1\mathcal{J}(R_1,R_2)||\partial_2\mathcal{J}(R_1,R_2)|} e^{-n\mathcal{J}(R_1,R_2)}(1+o(1)), & \textit{if} \quad m_1 < R_2 < m_2, \\ \frac{1}{2\sqrt{2\pi n}} \frac{\det(\operatorname{Hess}\mathcal{J}(R_1,m_2))^{1/2}}{|\partial_1\mathcal{J}(R_1,m_2)| \sqrt{\partial_{22}\mathcal{J}(R_1,m_2)}} e^{-n\mathcal{J}(R_1,m_2)}(1+o(1)), & \textit{if} \quad R_2 = m_2, \\ \frac{1}{\sqrt{2\pi n}} \frac{\det(\operatorname{Hess}\mathcal{J}(R_1,m_2))^{1/2}}{|\partial_1\mathcal{J}(R_1,m_2)| \sqrt{\partial_{22}\mathcal{J}(R_1,m_2)}} e^{-n\mathcal{J}(R_1,m_2)}(1+o(1)), & \textit{if} \quad R_2 > m_2. \end{cases}$$

Remark 3.2.13. As R_2 increases, we see that the asymptotic volume exhibits a phase transition at $R_2 = m_2(R_1)$, when the prefactor (or pre-multiplicative factor in front of the exponentially decaying term) changes its order from n^{-1} to $n^{-1/2}$ and no longer depends on R_2. The latter occurs because roughly speaking, in the so-called high overlap regime $R_2 > m_2(R_2)$, $B^n_{V_1}(R_1)$ is essentially contained in $B^n_{V_2}(R_2)$ and so the volume of the intersection asymptotically equals that of $B^n_{V_1}(R_1)$. On the other hand, in the lower overlap regime $R_2 < m_2(R_1)$, the intersection is of a smaller order and depends on R_2, in addition to the shape of V_2.

The proof of the theorem is given in Section 3.4. In the theorem above, we see a transition for different values of R_2. In particular, when $R_2 > m_2$, the estimate does no depend on R_2 because, intuitively, $B^n_{V_1}(R_1)$ is mostly contained in $B^n_{V_2}(R_2)$, and hence in the asymptotics the volume of the intersection approximately equals the volume of $B^n_{V_1}(R_1)$. For the region where R_1 below m_1, we do not have an estimate of the intersection.

To obtain the sharp volume of a single Orlicz ball, $B^n_{V_1}(R_1)$, we may naively take $R_2 = +\infty$ in the theorem and obtain

$$\left|B^n_{V_1}(R_1)\right| = \frac{1}{\sqrt{2\pi n}} \frac{\det(\operatorname{Hess}\mathcal{J}(R_1, m_2))^{1/2}}{|\partial_1\mathcal{J}(R_1, m_2)| \sqrt{\partial_{22}\mathcal{J}(R_1, m_2)}} e^{-n\mathcal{J}(R_1,m_2)}(1+o(1)).$$

We show in Proposition 3.4.7 that the volume is indeed given as in the last display, although this requires a different proof.

A related question that has been studied in the literature is the asymptotics of the sequence of ratio

$$\frac{\left|B^n_{V_1}(R_1) \cap B^n_{V_2}(R_2)\right|}{\left|B^n_{V_1}(R_1)\right|}. \tag{3.28}$$

In [73], Schechtman and Schmuckenschläger first considered the problem of finding the limit of (3.28) for the subclass of ℓ_p^n balls. They showed that the ratio of the volumes of an ℓ_p^n ball and an ℓ_q^n ball converges to either 0 or 1 in the low overlap and high overlap regimes. The critical case was resolved after a decade by [76], who showed the ratio converges to 1/2 as $n \to \infty$. One reason to consider ℓ_p^n ball is that we have a convenience representation of uniform random variables on an ℓ_p^n balls. However, in the general setting of Orlicz balls, no such representation is available. In [49], they first looked at an LDP for the 2-norm of uniform random variables on Orlicz balls. It is shown that the proof in [49] relies on estimates, at an exponential scale, of the volumes of the intersection of ℓ_2^n balls and Orlicz balls. In [42], Kabluchko and Prochno extended the ideas further and related an asymptotic representation of Orlicz balls to a class of Gibbs measures, and used that to show that in the subcritical and supercritical regimes, respectively, once again the ratio converges to 0 or 1 exponentially fast. Theorem 3.2.12 leads us to the following immediate corollary, which resolves an open question raised in [42] under Assumption 3.2.10. We use a different proof based on the sharp asymptotic estimates for intersections of Orlicz balls obtained in Theorem 3.2.12.

Corollary 3.2.14. *Let V_1 and V_2 be two Orlicz functions that satisfy Assumption 3.2.10. Let $m_2(\cdot)$ be defined as in Theorem 3.2.12. Then as $n \to \infty$,*

$$\frac{\left|B^n_{V_1}(R_1) \cap B^n_{V_2}(R_2)\right|}{\left|B^n_{V_1}(R_1)\right|} \to \frac{1}{2} \quad \text{if} \quad R_2 = m_2(R_1).$$

The proof of the corollary is given in Section 3.4.4.

3.3 Tail estimates for random projections

We start with multidimensional reformulations. Let $\{\xi_i\}_{i\in\mathbb{N}}$ be a sequence of i.i.d. standard Gaussian random variables and denote $\xi^{(n)} := (\xi_1, \ldots, \xi_n)$. By [49, Lemma 4.2], for $k_n \leq n$, we have

$$n^{-1/2}\mathbf{A}_{n,k_n}^T X^{(n)} \overset{(d)}{=} \frac{\xi^{(k_n)}}{\left\|\xi^{(n)}\right\|_2} \frac{\left\|X^{(n)}\right\|_2}{\sqrt{n}}, \tag{3.29}$$

and

$$Y_{2,k_n}^n = n^{-1/2}\|\mathbf{A}_{n,k}^T X^{(n)}\|_2 \overset{(d)}{=} \frac{\left\|\xi^{(k_n)}\right\|_2}{\left\|\xi^{(n)}\right\|_2} \frac{\left\|X^{(n)}\right\|_2}{\sqrt{n}}. \tag{3.30}$$

Define

$$\bar{S}^n := \left(\frac{\xi^{(k_n)}}{\|\xi^{(n)}\|_2}, \frac{\|X^{(n)}\|_2^2}{n}\right), \quad \text{and} \quad \tilde{S}^n := \left(\frac{\|\xi^{(k_n)}\|_2}{\|\xi^{(n)}\|_2}, \frac{\|X^{(n)}\|_2^2}{n}\right). \tag{3.31}$$

Since $\{\xi\}_{i\in\mathbb{N}}$ and $X^{(n)}$ are independent and both with densities (the latter by Assumption 3.2.1), both $\bar{S}^n$ and $\tilde{S}^n$ have densities, which we denote by $f_{\bar{S}^n}$ and $f_{\tilde{S}^n}$, respectively.

Recall the definitions of $\mathcal{D}_\Upsilon$ and $\mathcal{D}_a$ in (3.5) and (3.14), respectively. By (3.29), for a measurable set $\Upsilon \subset \mathbb{R}^{k_n}$, we can rewrite the tail probability in (3.6) as an integral

$$\mathbb{P}\left(n^{-1/2}\mathbf{A}_{n,k_n}^T X^{(n)} \in \Upsilon\right) = \mathbb{P}\left(\frac{\xi^{(k_n)}}{\left\|\xi^{(n)}\right\|_2} \frac{\left\|X^{(n)}\right\|_2}{\sqrt{n}} \in \Upsilon\right) = \int_{\mathcal{D}_\Upsilon} f_{\bar{S}^n}(x,y)dxdy.$$

Similarly, for $a > 0$, the tail probabilities of interest in (3.15) and (3.19) can be written as

$$\mathbb{P}\left(Y_{2,k_n}^n > a\right) = \mathbb{P}\left(\frac{\left\|\xi^{(k_n)}\right\|_2}{\left\|\xi^{(n)}\right\|_2} \frac{\|X^{(n)}\|_2}{\sqrt{n}} > a\right) = \int_{\mathcal{D}_a} f_{\tilde{S}^n}(x,y)dxdy. \tag{3.32}$$

Thus, to obtain tail estimates, we first understand the density of $\xi^{(k_n)}/\|\xi^{(n)}\|_2$ in Section 3.3.1 and then use that to obtain expansions for the densities of $\bar{S}^n$ and $\tilde{S}^n$ in Section 3.3.2. In Section 3.3.3, we recall a generalized Laplace approximation. These results are used to prove Theorems 3.2.3, 3.2.7 and 3.2.8 in Section 3.3.4.

3.3.1 Density estimates related to $\xi^{(n)}$

Note that the first k_n coordinates of $\bar{S}^n$ is related to Dirichlet random variables.

Lemma 3.3.1. *Fix $k, n \in \mathbb{N}$ and $k < n$. The density of the $\mathbb{R}^k$-valued random vector*

$$G := \frac{(\xi_1, \cdots, \xi_k)}{\left(\sum_{i=1}^n \xi_i^2\right)^{\frac{1}{2}}}$$

is given by

$$f_G(x) := \begin{cases} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^k \Gamma\left(\frac{n-k}{2}\right)} \left(1 - \sum_{i=1}^k x_i^2\right)^{\frac{n-k}{2}-1}, & x \in (-1,1)^k \quad \textit{such that} \quad \|x\|_2 \leq 1, \\ 0, & \textit{otherwise}. \end{cases} \tag{3.33}$$

Proof. Define the $\mathbb{R}^{k+1}$-valued and $\mathbb{R}^k$-valued random vectors

$$\tilde{G} := \frac{\left(\xi_1^2, \cdots, \xi_k^2, \sum_{i=k+1}^n \xi_i^2\right)}{\sum_{i=1}^n \xi_i^2}, \quad \text{and} \quad \hat{G} := \frac{(\xi_1^2, \ldots, \xi_k^2)}{\sum_{i=1}^n \xi_i^2}.$$

Since for $i = 1, \ldots, n$, ξ_i^2 is chi-square, or equivalently Gamma$(1/2, 2)$, distributed, $\tilde{G}$ has a Dirichlet distribution with parameter $(1/2, \cdots, 1/2, (n-k)/2)$, which has density

$$\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^k \Gamma\left(\frac{n-k}{2}\right)} x_1^{-\frac{1}{2}} \cdots x_k^{-\frac{1}{2}} x_{k+1}^{\frac{n-k}{2}-1}, \quad x \in (0,1)^{k+1} \quad \text{such that} \quad \sum_{i=1}^{k+1} x_i = 1.$$

Hence, the density of $\hat{G}$ is given by

$$\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^k \Gamma\left(\frac{n-k}{2}\right)} x_1^{-\frac{1}{2}} \cdots x_k^{-\frac{1}{2}} \left(1 - \sum_{i=1}^k x_i\right)^{\frac{n-k}{2}-1}, \quad x \in (0,1)^k \quad \text{such that} \quad \sum_{i=1}^k x_i \leq 1.$$

Using a change of variables $y = \sqrt{x}$ and exploiting the symmetry of the Gaussian distribution, one then obtains (3.33). The density of $\sqrt{\hat{G}}$ is then given by

$$f_{\sqrt{\hat{G}}}(x) = \frac{2^k \Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^k \Gamma\left(\frac{n-k}{2}\right)} \left(1 - \sum_{i=1}^k x_i^2\right)^{\frac{n-k}{2}-1}, \quad x \in (0,1)^k \quad \text{such that} \quad \sum_{i=1}^k x_i \leq 1.$$

Finally, by the symmetry of the Gaussian distribution, we obtain the density of G,

$$f_G(x) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^k \Gamma\left(\frac{n-k}{2}\right)} \left(1 - \sum_{i=1}^k x_i^2\right)^{\frac{n-k}{2}-1}, \quad x \in (-1,1)^k \quad \text{such that} \quad \sum_{i=1}^k x_i^2 \leq 1.$$

□

Using similar arguments, we obtain the following density.

Lemma 3.3.2. *Fix $k_n, n \in \mathbb{N}$, with each $k_n < n$. Let $\bar{G} := \|\xi^{(k_n)}\|_2 / \|\xi^{(n)}\|_2$. Then the density of $\bar{G}$ is given by*

$$f_{\bar{G}}(x) = \begin{cases} \frac{2\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{k_n}{2}\right)\Gamma\left(\frac{n-k_n}{2}\right)} x^{k_n - \frac{3}{2}} (1-x^2)^{-\frac{k_n}{2}-1} (1-x^2)^{\frac{n}{2}}, & x \in (0,1), \\ 0, & \textit{otherwise}. \end{cases} \tag{3.34}$$

Proof. Since $\bar{G}^2 = Y/(Y+Z)$ where $Y = \sum_{i=1}^{k_n} \xi_i^2$ and $Z = \sum_{i=k_n+1}^{n} \xi_i^2$ are independent chi-square distribution with k_n and $n-k_n$ degree of freedom, respectively, $\bar{G}^2 \sim$Beta$(k_n/2, (n-$

$k_n)/2)$ and has density

$$f_{\bar{G}^2}(x) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{k_n}{2}\right)\Gamma\left(\frac{n-k_n}{2}\right)} x^{\frac{k_n}{2}-1}(1-x)^{\frac{n-k_n}{2}-1}, \quad x \in (0,1).$$

The lemma follows by a simple change of variables along with the symmetry of $\xi^{(k_n)}$. $\square$

3.3.2 Density estimates for $\bar{S}^n$ and $\tilde{S}^n$

We first obtain an asymptotic expansion for the density $f_{\bar{S}^n}$ of the vector $\bar{S}^n$ in the following lemma, which combines Lemma 3.3.1 and Assumption 3.2.1. The following lemma will be used in the proof of the tail estimate in the constant regime. Define

$$D_\xi := \{x \in \mathbb{R}^k : \|x\|_2 \le 1\}. \tag{3.35}$$

Lemma 3.3.3. *Fix $k \in \mathbb{N}$. Recall the function $\bar{J}^{\rm con}$ defined in (3.4). Suppose $\{X^n\}_{n\in\mathbb{N}}$ satisfies Assumption 3.2.1 with parameters (J, h, β). Let $\bar{S}^n$ be defined in (3.31) and $f_{\bar{S}^n}$ be its density. Then*

$$f_{\bar{S}^n}(x,y) = n^\beta \left(\frac{n}{2\pi}\right)^{k/2} h(y) \left(1 - \|x\|_2^2\right)^{-\frac{k}{2}-1} e^{-n\bar{J}^{\rm con}(x,y)}(1+o(1)), \tag{3.36}$$

uniformly in any compact neighborhood of $(x,y) \in D_\xi \times D_J$. In particular, for $(x,y) \in D_\xi \times D_J$, $f_{\bar{S}^n}(x,y) = \hat{f}_{\bar{S}^n}(\|x\|_2, y)$, where

$$\hat{f}_{\bar{S}^n}(r,y) = n^\beta \left(\frac{n}{2\pi}\right)^{k/2} h(y) \left(1 - r^2\right)^{-k/2-1} e^{-n\bar{J}^{\rm con}(r,y)}(1+o(1)), \tag{3.37}$$

uniformly in any compact neighborhood of $(r,y) \in [0,1] \times D_J$.

Proof. Fix $k, n \in \mathbb{N}$, $n > k$. Define

$$B_{n,k} := \frac{\Gamma\left(\frac{1}{2}\right)^k \Gamma\left(\frac{n-k}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}.$$

By Assumption 3.2.1, the independence of $\{\xi_i\}_{i\in\mathbb{N}}$ and $X^{(n)}$, Lemma 3.3.1,, and the expression (3.4) for $\bar{J}^{\text{con}}$, $\bar{S}^n := (\xi^{(k)}/\|\xi^{(n)}\|_2, \|X^{(n)}\|_2^2/n)$ has the asymptotic density:

$$\begin{aligned} f_{\bar{S}^n}(x,y) &= \frac{1}{B_{n,k}}\left(1 - \sum_{i=1}^k x_i^2\right)^{-k/2-1} \exp\left(\frac{n}{2}\log\left(1 - \sum_{i=1}^k x_i^2\right)\right) n^\beta h(y) e^{-nJ(\sqrt{y})}(1+o(1)) \\ &= \frac{n^\beta}{B_{n,k}} h(y)\left(1 - \sum_{i=1}^k x_i^2\right)^{-k/2-1} e^{-n\bar{J}^{\text{con}}(x,y)}(1+o(1)), \end{aligned} \tag{3.38}$$

uniformly in any compact neighborhood of $(x,y) \in D_\xi \times D_J$. By Stirling's approximation,

$$\begin{aligned} B_{n,k} = \frac{\Gamma\left(\frac{1}{2}\right)^k \Gamma\left(\frac{n-k}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} &= \Gamma\left(\frac{1}{2}\right)^k \frac{\sqrt{2\pi}\left(\frac{n-k}{2} - 1\right)^{\frac{n-k}{2}-\frac{1}{2}} e^{-\frac{n-k}{2}+1}}{\sqrt{2\pi}\left(\frac{n}{2} - 1\right)^{\frac{n}{2}-\frac{1}{2}} e^{-\frac{n}{2}+1}}(1+o(1)) \\ &= \left(\frac{2\pi}{n}\right)^{k/2}(1+o(1)). \end{aligned} \tag{3.39}$$

Combining (3.38) and (3.39), we obtain (3.36). The expression in (3.37) follows as an immediate consequence since $f_{\bar{S}^n}(x,y)$ is spherically symmetric in (3.36) and for $r \in (0,1)$ the neighborhood of $\|x\|_2 = r$ is bounded and contained in D_ξ. ◻

Next, we obtain asymptotic expressions for the density $f_{\tilde{S}^n}$ of the vector $\tilde{S}^n$ defined in (3.31) in the sublinear and linear regimes. This will be used in the proofs of Theorems 3.2.7 and 3.2.8 in Section 3.3.4.

Lemma 3.3.4. *Recall the definitions of $\bar{J}^{\text{sub}}$, M^{sub}, $\bar{J}^{\text{lin}}$ and M_λ^{lin} in (3.13), (3.18), (3.16) and (3.20), respectively. Fix $k_n, n \in \mathbb{N}$ and $k_n < n$. Suppose $\{X^n\}_{n\in\mathbb{N}}$ satisfies Assumption 3.2.1* *with parameters* (J, h, β)*. Let* $f_{\tilde{S}^n}$ *be the density of* $\tilde{S}^n$*. Then*

1. *(sublinear regime) for* $k_n \ll n$*, we have*

$$f_{\tilde{S}^n}(x, y) = \frac{n^\beta k_n^{\frac{1}{2}}}{\sqrt{\pi} x^{\frac{3}{2}}(1 - x^2)} h(y) e^{-n\bar{J}^{\text{sub}}(x,y)+(k_n \log n)M_n^{sub}(x)}(1 + o(1)), \tag{3.40}$$

uniformly in any compact neighborhood of $(x, y) \in (0, 1) \times D_J$*;*

2. *(linear regime) for* $\lambda \in (0, 1)$*,* $k_n = \lambda n + r_n$ *and* $r_n = o(n)$*,*

$$f_{\tilde{S}^n}(x, y) = \frac{n^{\beta+\frac{1}{2}}}{\sqrt{\pi}} \frac{(\lambda - \lambda^2)^{\frac{1}{2}} h(y)}{x^{\frac{3}{2}}(1 - x^2)} e^{-n\bar{J}^{\text{lin}}(x,y)+r_n M_\lambda^{\text{lin}}(x)}(1 + o(1)), \tag{3.41}$$

uniformly in any compact neighborhood of $(x, y) \in (0, 1) \times D_J$*.*

Proof. Fix $n, k_n \in \mathbb{N}$, $k_n < n$. Define

$$B_{n,k_n} := \frac{2\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{k_n}{2}\right)\Gamma\left(\frac{n-k_n}{2}\right)}.$$

By Assumption 3.2.1, the independence of $\{\xi_i\}_{i\in\mathbb{N}}$ and $X^{(n)}$, Lemma 3.3.2, and the expression (3.13) for $\bar{J}^{\text{sub}}$, we have

$$f_{\tilde{S}^n}(x, y) = \frac{n^\beta B_{n,k_n}}{x^{\frac{3}{2}}(1 - x^2)} x^{k_n}(1 - x^2)^{-\frac{k_n}{2}} h(y) \exp\left(-n\bar{J}^{\text{sub}}(x, y)\right)(1 + o(1)), \tag{3.42}$$

uniformly in any compact neighborhood of $(x, y) \in (0, 1) \times D_J$. Applying Stirling's formula we see that

$$B_{n,k_n} = \frac{2\sqrt{2\pi}\left(\frac{n}{2} - 1\right)^{\frac{n}{2}-\frac{1}{2}} e^{-\frac{n}{2}+1}}{\sqrt{2\pi}\left(\frac{k_n}{2} - 1\right)^{\frac{k_n}{2}-\frac{1}{2}} e^{-\frac{k_n}{2}+1} \sqrt{2\pi}\left(\frac{n-k_n}{2} - 1\right)^{\frac{n-k_n}{2}-\frac{1}{2}} e^{-\frac{n-k_n}{2}+1}}(1 + o(1)). \tag{3.43}$$

We now consider the two regimes separately.

1. If $k_n \ll n$, then

$$
\begin{aligned}
B_{n,k_n} &= \frac{1}{\sqrt{\pi}} \frac{n^{n/2}}{k_n^{k_n/2}(n-k_n)^{(n-k_n)/2}} k_n^{1/2}(1+o(1)) \\
&= \frac{1}{\sqrt{\pi}} \left(\frac{n}{k_n}\right)^{k_n/2} \frac{1}{(1-\frac{k_n}{n})^{(n-k_n)/2}} k_n^{1/2}(1+o(1)) \\
&= \sqrt{\frac{k_n}{\pi}} \exp\left(\frac{k_n \log n}{2}\left(1 - \frac{\log k_n}{\log n} + \frac{1-\frac{n}{k_n}}{\log n}\log\left(1-\frac{k_n}{n}\right)\right)\right).
\end{aligned}
$$

   Hence, using (3.42) and the definitions of $\bar{J}^{\text{sub}}$ in (3.13) and M_n^{sub} in (3.16), we obtain (3.40).

2. If $k_n = \lambda n + r_n$ and $r_n = o(n)$, then substituting it in (3.43), we have

$$
\begin{aligned}
B_{n,k_n} &= \frac{1}{\sqrt{\pi}} k_n^{1/2}(1-\lambda)^{1/2} \frac{n^{n/2}}{k_n^{k_n/2}(n-k_n)^{(n-k_n)/2}}(1+o(1)) \\
&= \sqrt{\frac{n}{\pi}} \lambda^{1/2}(1-\lambda)^{1/2} \exp\left(-n\left(\frac{\lambda}{2}\log\lambda + \frac{1-\lambda}{2}\log(1-\lambda)\right)\right) \\
&\quad \times \exp\left(\frac{r_n}{2}\left(\log\left(1-\frac{1}{\lambda}\right) - \left(1+\frac{\lambda n}{r_n}\right)\log\left(1+\frac{r_n}{\lambda n}\right)\right.\right. \\
&\qquad\qquad \left.\left. + \left(1-\frac{(1-\lambda)n}{r_n}\right)\log\left(1-\frac{r_n}{(1-\lambda)n}\right)\right)\right)(1+o(1)),
\end{aligned}
$$

   Hence, using (3.42) and the definitions of $\bar{J}^{\text{lin}}$ in (3.18) and M_λ^{lin} in (3.20), we obtain (3.41).

$\square$

3.3.3 A generalized Laplace approximation

Finally, we state an Laplace type asymptotic integral from [58] that is used multiple times in the proof.

Definition 3.3.5. Given $m, d \in \mathbb{N}$, and a bounded domain $D \subset \mathbb{R}^{m+d}$, we say that the sequence $h^n : \mathbb{R}^{m+d} \to \mathbb{R}$, $n \in \mathbb{N}$, admits a (f, x^*, α, g^n)-representation if for each $n \in \mathbb{N}$,

$$h^n(x) = g^n(x)e^{-nf(x)}, \quad x \in D,$$

where

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

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

Remark 3.3.6. Suppose there exist smooth functions t and r such that g^n has the form $g^n(x) = t(x)\exp(k_n r(x))$. Moreover, $t(x) \neq 0$ for x in a neighborhood of x^* and $k_n/n^\alpha \to 0$ as $n \to \infty$ for $\alpha < 1$. Then g^n satisfies the second property in the definition because there exists $C \in (0, \infty)$ such that for x in a neighborhood of x^*, $\left|\log t(x) - \log t(x^*)\right| \leq C\|x - x^*\|$ and $|r(x) - r(x^*)| \leq C\|x - x^*\|$.

Proposition 3.3.7 (Proposition 5.3 [58])**.** *Let $\mathcal{D} \subset \mathbb{R}^d$ be a bounded domain whose boundary is a differentiable $(d-1)$-dimensional hypersurface. Let $h^n : \mathbb{R}^{m+d} \to \mathbb{R}$, $n \in \mathbb{N}$, be a sequence* *of functions admitting a (f, x^*, α, g^n)-representation in $\mathcal{D}$ in the sense of Definition 4.6.5. Then*

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

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

$$C_1 := \{y : f(y) = f(x^*)\} \qquad \textit{and} \qquad C_2 := \partial\mathcal{D}.$$

3.3.4 Proofs of tail estimates for random projections

3.3.4.1 Constant regime

Proof of Theorem 3.2.3. Invoking the representation in (3.29) and the definition of $\bar{S}^n$ in (3.31), we can write

$$\mathbb{P}\left(n^{-1/2}\mathbf{A}_{n,k}^T X^{(n)} \in \Upsilon\right) = \mathbb{P}\left(\frac{\xi^{(k)}}{\left\|\xi^{(n)}\right\|_2}\frac{\|X^{(n)}\|_2}{\sqrt{n}} \in \Upsilon\right) = \mathbb{P}\left(\bar{S}^n \in \mathcal{D}_\Upsilon\right),$$

where $\mathcal{D}_\Upsilon$ is defined in (3.5). By the assumption that $(x_\Upsilon, y_\Upsilon) \in \partial\mathcal{D}_\Upsilon$ and $y_\Upsilon \in D_J$, we have $(x_\Upsilon, y_\Upsilon) \in D_\xi \times D_J$, where D_ξ is defined in (3.35). For any bounded measurable neighborhood U of (x_Υ, y_Υ), write

$$\mathbb{P}\left(n^{-1/2}\mathbf{A}_{n,k}^T X^{(n)} \in \Upsilon\right) = \mathbb{P}\left(\bar{S}^n \in \mathcal{D}_\Upsilon \cap U\right) + \mathbb{P}\left(\bar{S}^n \in \mathcal{D}_\Upsilon \cap U^c\right). \tag{3.44}$$

For the first term in the last display, Lemma 3.3.3 implies

$$\mathbb{P}\left(\bar{S}^n \in \mathcal{D}_\Upsilon \cap U\right) = \int_{\mathcal{D}_\Upsilon \cap U} f_{\bar{S}^n}(x, y)dxdy$$

$$= n^{\beta}\left(\frac{n}{2\pi}\right)^{k/2}\int_{\mathcal{D}_{\Upsilon}\cap U} h(y)\left(1-\|x\|_2^2\right)^{-\frac{k}{2}-1} e^{-n\bar{J}^{\mathrm{con}}(x,y)}dxdy(1+o(1)),$$

where we used the boundedness of U and the fact that $(x_{\Upsilon}, y_{\Upsilon}) \in D_{\xi} \times D_J$. Since by assumption, $(x_{\Upsilon}, y_{\Upsilon})$ uniquely attains the infimum of $\bar{J}^{\mathrm{con}}$ over $\mathcal{D}_{\Upsilon}$, from the definition of $I_{\mathbf{A}X,k}$ in (3.3), we have the following relation

$$\begin{aligned}\inf_{(x,y)\in\mathcal{D}_{\Upsilon}} \bar{J}^{\mathrm{con}}(x,y) &= \inf_{xy\in\Upsilon, y\in\mathbb{R}_+, \|x\|_2\in(0,1)} \bar{J}^{\mathrm{con}}(x,y)\\ &= \inf_{x\in\Upsilon, c\in(0,1)}\left\{J\left(\frac{\|x\|_2}{c}\right)-\frac{1}{2}\log\left(1-c^2\right)\right\}\\ &= \inf_{x\in\Upsilon} I_{\mathbf{A}X,k}(x). \end{aligned} \tag{3.45}$$

Since $(x_{\Upsilon}, y_{\Upsilon})$ lies on the smooth part of the boundary $\mathcal{D}_{\Upsilon}$, we can invoke (3.45) and Proposition 4.6.7 with d, g^n, $\mathcal{D}$ and f therein replaced with $k+1$, 1, $\mathcal{D}_{\Upsilon}\cap U$ and $\bar{f}$, respectively, to obtain

$$\mathbb{P}\left(\bar{S}^n \in \mathcal{D}_{\Upsilon}\cap U\right) = n^{\beta}\left(\frac{n}{2\pi}\right)^{k/2}\frac{(2\pi)^{k/2}}{n^{(k+2)/2}}\frac{h(y_{\Upsilon})}{\gamma^{\mathrm{con}}_{\Upsilon,k}}\left(1-\|x_{\Upsilon}\|_2^2\right)^{-\frac{k}{2}-1} e^{-nI_{\mathbf{A}X,k}(\Upsilon)}(1+o(1)), \tag{3.46}$$

for some constant $\gamma^{\mathrm{con}}_{\Upsilon,k} \in \mathbb{R}_+$, which can be explicitly calculated using Proposition 4.6.7.

For the second term in (4.105), since $(x_{\Upsilon}, y_{\Upsilon}) \in U$ is the unique minimizer of $\bar{J}^{\mathrm{con}}$ over $\mathcal{D}_{\Upsilon}$, from (3.45), there exists $\delta > 0$ such that

$$\inf_{(x,y)\in\mathcal{D}_{\Upsilon}\cap U^c} \bar{J}^{\mathrm{con}}(x,y) > \inf_{(x,y)\in\mathcal{D}_{\Upsilon}} \bar{J}^{\mathrm{con}}(x,y) + \delta = I_{\mathbf{A}X,k}(\Upsilon)+\delta,$$

where we used the relation (3.45) in the last equality. It is shown in [49] that if $k_n \equiv k \in \mathbb{N}$, then the sequence $\{\bar{S}^n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n with rate function $\bar{J}^{\mathrm{con}}$ defined in (3.4). By the LDP for $\{\bar{S}^n\}_{n\in\mathbb{N}}$, for the closed set $\mathcal{D}_{\Upsilon}\cap U^c$ we have the following upper bound,

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\bar{S}^n\in\mathcal{D}_\Upsilon\cap U^c\right)\leq -\inf_{(x,y)\in\mathcal{D}_\Upsilon\cap U^c}\bar{J}^{\mathrm{con}}(x,y)<-I_{\mathbf{A}X,k}(\Upsilon)-\delta.$$

Hence, the second term in (4.105) is negligible with respect to (4.111). Finally, combining (4.105) and (4.111), we obtain (3.6). □

3.3.4.2 Proof of norm tail estimates in the constant regime

Proof of Theorem 3.2.6. We discuss each case in the following.

Case (1)(a). First, fix $k\in\mathbb{N}$, $q\in[1,2)$ and $w_{q,a}\in\mathcal{W}_{q,k}(a)$. Consider a neighborhood $U\subset\mathbb{R}^k$ of $w_{q,a}$. From (3.3) and (3.7), we see that

$$\begin{aligned}
I_{\mathbf{A}X,k}(\Upsilon_{q,a}) &= I_{\mathbf{A}X,k}(w_{q,a})\\
&= \inf_{x\in\mathbb{R}^k,\|x\|_2\leq 1,y\in\mathbb{R}_+,x\sqrt{y}=w_{q,a}}\left\{-\frac{1}{2}\log(1-\|x\|_2^2)+J(\sqrt{y})\right\}\\
&= -\frac{1}{2}\log(1-\left\|x_{q,a}\right\|_2^2)+J(\sqrt{y_{q,a}}).
\end{aligned}$$

By assumption, $y_{q,a}\in D_J$, we may apply Theorem 3.2.3 to obtain

$$\mathbb{P}\left(n^{-1/2}\mathbf{A}_{n,k}^TX^{(n)}\in U\right)=2^k n^{\beta-1}C^{\mathrm{con}}_{a,q,k}e^{-nI_{\mathbf{A}X,k}(\Upsilon_{q,a})}(1+o(1)).$$

Now, note that the expression on the right-hand side does not depend on the particular point $w_{q,a}\in\mathcal{W}_{q,k}(a)$. Therefore, since $\left|\mathcal{W}_{q,k}(a)\right|=2^k$ for $q<2$,we combine the contribution at each minimizers and , to obtain

$$\mathbb{P}\left(n^{-1/2}\mathbf{A}_{n,k}^TX^{(n)}\in\Upsilon_{q,a}\right)=2^k\mathbb{P}\left(n^{-1/2}\mathbf{A}_{n,k}^TX^{(n)}\in\Upsilon_{q,a}\cap U\right).$$

Case (1)(b). For $q \in (2, \infty)$, we have $2k$ minimizers. Hence again by the symmetry and Theorem 3.2.3, we have

$$\mathbb{P}\left(n^{-1/2}\mathbf{A}_{n,k}^T X^{(n)} \in \Upsilon_{q,a}\right) = 2kn^{\beta-1}C^{\mathrm{con}}_{a,q,k}e^{-nI_{\mathbf{A}X,k}(\Upsilon_{q,a})}(1+o(1)).$$

Case (2). For $q = 2$, since the minimizers are not isolated, we may not apply Theorem 3.2.3 directly. Instead, let $\bar{S}^n$ be defined in (3.31) and as usual $f_{\bar{S}^n}$ denote its density. By Lemma 3.3.3,

$$\begin{aligned}
\mathbb{P}\left(n^{-1/2}\mathbf{A}_{n,k}^T X^{(n)} \in \Upsilon_{2,a}\right) &= \mathbb{P}\left(\frac{\left\|\xi^{(k)}\right\|_2}{\left\|\xi^{(n)}\right\|_2}\frac{\|X^{(n)}\|_2}{\sqrt{n}} > a\right)\\
&= \int_0^\infty \int_{\|x\|_2\sqrt{y}>a} f_{\bar{S}^n}(x,y)dydx\\
&= \int_0^\infty \int_{r\sqrt{y}>a} \frac{k\pi^{k/2}}{\Gamma(k/2+1)} f_{\bar{S}^n}(r,y)drdx.
\end{aligned}$$

Recall the definition of $\mathcal{D}_a$ in (3.14) and $(r_{2,a}, y_{2,a})$ in the theorem. Pick a neighborhood U of $(r_{2,a}, y_{2,a})$ in $\mathcal{D}_a$. Then

$$\mathbb{P}\left(n^{-1/2}\mathbf{A}_{n,k}^T X^{(n)} \in \Upsilon_{2,a}\right) = \frac{k\pi^{k/2}}{\Gamma(k/2+1)}\left(\int_{\mathcal{D}_a\cap U} f_{\bar{S}^n}(r,y)drdx + \int_{\mathcal{D}_a\cap U^c} f_{\bar{S}^n}(r,y)drdx\right). \quad (3.47)$$

For the first term above, we have the following asymptotic expansion by [15, Equation (8.2.24)] and the fact that $y_{q,a} \in D_J$,

$$\int_{\mathcal{D}_a\cap U} f_{\bar{S}^n}(r,y)drdx = \left(\frac{n}{2\pi}\right)^{k/2} n^{1/2}h(y_{2,a})\left(1-\left|r_{2,a}\right|_2^2\right)^{-k/2-1}\frac{(2\pi)^{1/2}n^{\beta}}{n^2\gamma^{\mathrm{con}}_{2,a}}e^{-nI_{\mathbf{A}X,k}(\Upsilon_{2,a})}(1+o(1)).$$

Therefore, the second term on the right-hand side of (3.47) is negligible with respect to the first term by the LDP for $\{n^{-1/2}\mathbf{A}_{n,k}^T X^{(n)}\}_{n\in\mathbb{N}}$ [49, Theorem 2.7]. Hence, we obtain, using the definition of $C^{\text{con}}_{a,2,k}$ below (3.10),

$$\begin{aligned}
&\mathbb{P}\left(n^{-1/2}\mathbf{A}_{n,k}^{T}X^{(n)} \in \Upsilon_{2,a}\right)\\
&\quad= \frac{k\pi}{\Gamma(k/2+1)}\frac{n^{\beta+\frac{k-3}{2}}}{2^{k/2-1/2}}\frac{h(y_{2,a})}{\gamma^{\text{con}}_{2,a}}\left(1-\left|r_{2,a}\right|_2^2\right)^{-k/2-1} e^{-nI_{\mathbf{A}X,k}(\Upsilon_{2,a})}(1+o(1))\\
&\quad= n^{\beta+\frac{k-3}{2}}C^{\text{con}}_{a,2,k}e^{-nI_{\mathbf{A}X,k}(\Upsilon_{2,a})}(1+o(1)).
\end{aligned}$$

□

3.3.4.3 Proof of norm tail estimates in the sublinear regime

Proof of Theorem 3.2.7. Recall the definition of $\mathcal{D}_a$ in (3.14). By (3.32) and (3.31), for any bounded measurable neighborhood U of (x_a, y_a), we split the probability into two parts

$$\mathbb{P}\left(Y^{n}_{2,k_n} > a\right) = \mathbb{P}\left(\tilde{S}^n \in \mathcal{D}_a \cap U\right) + \mathbb{P}\left(\tilde{S}^n \in \mathcal{D}_a \cap U^c\right). \tag{3.48}$$

For the first term in the last display, since $(x_a, y_a) \in (0,1)\times D_J$, we have the following estimate by (3.40) and the boundedness of U:

$$\begin{aligned}
\mathbb{P}\left(\tilde{S}^n \in \mathcal{D}_a \cap U\right) &= \int_{\mathcal{D}_a\cap U} f_{\tilde{S}^n}(x,y)dxdy\\
&= \frac{n^{\beta}k_n^{1/2}}{\sqrt{\pi}}\int_{\mathcal{D}_a\cap U}\frac{h(y)}{x^{3/2}(1-x^2)}e^{-n\bar{J}^{\text{sub}}(x,y)+(k_n\log n)M^{\text{sub}}_n(x)}dxdy(1+o(1)).
\end{aligned}$$

Since by assumption, (x_a, y_a) is the unique minimizer of $\bar{J}^{\text{sub}}$ over $\mathcal{D}_a$. By the definition of $\mathbb{J}^{\text{sub}}_2$, we have the following relation

$$\bar{J}^{\text{sub}}(x_a, y_a) = \inf_{(x,y)\in\mathcal{D}_a}\bar{J}^{\text{sub}}(x,y) = \inf_{x>a}\mathbb{J}^{\text{sub}}_2(x). \tag{3.49}$$

Since in the sublinear regime, by assumption, $k_n/n^\alpha \to 0$ as $n \to \infty$, we see from the definition of M_n^{sub} in (3.16) that,

$$\lim_{n\to\infty} \frac{(k_n \log n) M_n^{\mathrm{sub}}(x_a, y_a)}{n^\alpha} = 0.$$

By assumption, (x_a, y_a) lies in the smomoth part of $\partial\mathcal{D}_a$. We then have the following asymptotic integral by Proposition 4.6.7 and Remark 4.3.1 with d, g^n, $\mathcal{D}$ and f therein replaced with 2, $x^{-3/2}(1-x^2)^{-1}h(y)\exp(k_n \log n M_n^{\mathrm{sub}}(x))$, $\mathcal{D}_a \cap U$ and $\bar{J}^{\mathrm{sub}}$, respectively,

$$\mathbb{P}\left(\tilde{S}^n \in \mathcal{D}_a \cap U\right) = \frac{n^\beta k_n^{1/2}}{\sqrt{\pi}} \frac{(2\pi)^{1/2}}{\gamma_a^{\mathrm{sub}} n^{3/2}} \frac{h(y_a)}{x_a^{3/2}(1-x_a^2)} e^{-n \inf_{x>a} \mathbb{J}_2^{\mathrm{sub}}(x) + k_n \log n M_n^{\mathrm{sub}}(x_a)}(1+o(1)) \tag{3.50}$$

for some constant $\gamma_a^{\mathrm{sub}} \in \mathbb{R}_+$ where we also use the relation (3.49).

For the second term in (3.48), since $(x_a, y_a) \in U$ is the unique minimizer of $\bar{J}^{\mathrm{sub}}$ over $\mathcal{D}_a$, there exists $\delta > 0$ such that

$$\inf_{(x,y)\in\mathcal{D}_a\cap U^c} \bar{J}^{\mathrm{sub}}(x,y) > \inf_{(x,y)\in\mathcal{D}_a} \bar{J}^{\mathrm{sub}}(x,y) + \delta = \mathbb{J}_2^{\mathrm{sub}}(a) + \delta.$$

where we used (3.49) in the last equality. It is shown in [49] that $\{\tilde{S}^n\}_{n\in\mathbb{N}}$ satisfies an LDP with rate function $\bar{J}^{\mathrm{sub}}$ in the sublinear regime. By the LDP upper bound for the sequence $\{\tilde{S}^n\}_{n\in\mathbb{N}}$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\tilde{S}^n \in \mathcal{D}_a \cap U^c\right) \le -\inf_{(x,y)\in\mathcal{D}_a\cap U^c} \bar{J}^{\mathrm{sub}}(x,y) < -\mathbb{J}_2^{\mathrm{sub}}(a) - \delta.$$

Hence, the second term in (3.48) is negligible with respect to (4.111). Finally, combining (3.48) and (3.50), we obtain (3.15). $\square$

3.3.4.4 Proof of norm tail estimates in the linear regime

Proof of Theorem 3.2.8. Recall the definition of $\mathcal{D}_a$ in (3.14). By (3.32) and (3.31), for any bounded measurable neighborhood U of (x_a, y_a), we split the probability into two parts

$$\mathbb{P}\left(\frac{\left\|\xi^{(k_n)}\right\|_2}{\left\|\xi^{(n)}\right\|_2}\frac{\|X^{(n)}\|_2}{\sqrt{n}} \in \Upsilon\right) = \mathbb{P}\left(\tilde{S}^n \in \mathcal{D}_a \cap U\right) + \mathbb{P}\left(\tilde{S}^n \in \mathcal{D}_a \cap U^c\right). \tag{3.51}$$

For the first term in the last display, we have the following estimate by (3.41) and the boundedness of U,

$$\begin{aligned}\mathbb{P}\left(\tilde{S}^n \in \mathcal{D}_a \cap U\right) &= \int_{\mathcal{D}_a \cap U} f_{\tilde{S}^n}(x,y)dxdy \\ &= \frac{n^{\beta+\frac{1}{2}}(\lambda-\lambda^2)^{1/2}}{\sqrt{\pi}} \int_{\mathcal{D}_a\cap U} \frac{h(y)}{x^{3/2}(1-x^2)} e^{-n\bar{J}^{\mathrm{lin}}(x,y)+r_n M_\lambda^{\mathrm{lin}}(x)} dxdy(1+o(1))\end{aligned}$$

By the assumption, the infimum of $\bar{J}^{\mathrm{lin}}$ over $\mathcal{D}_a$ is attained uniquely at a smooth boundary point $(x_a, y_a) \in \partial\mathcal{D}_a$. By the definition of $\mathbb{J}^{\mathrm{lin}}_{2,\lambda}$, we have the following relation

$$\bar{J}^{\mathrm{lin}}(x_a, y_a) = \inf_{(x,y)\in\mathcal{D}_a} \bar{J}^{\mathrm{lin}}(x,y) = \inf_{x>a} \mathbb{J}^{\mathrm{lin}}_{2,\lambda}(x). \tag{3.52}$$

Since r_n/n^α as $n \to \infty$, we then have the following asymptotic integral by Proposition 4.6.7 and Remark 4.3.1,

$$\mathbb{P}\left(\tilde{S}^n \in \mathcal{D}_a \cap U\right) = \frac{n^{\beta+\frac{1}{2}}}{\sqrt{\pi}}(\lambda-\lambda^2)^{1/2} \frac{h(y_a)}{x_a^{3/2}(1-x_a^2)} \frac{(2\pi)^{1/2}}{n^{3/2}} e^{-n\inf_{x>a}\mathbb{J}^{\mathrm{lin}}_{2,\lambda}(x)+r_n M_\lambda^{\mathrm{lin}}(x_a)}(1+o(1)) \tag{3.53}$$

for some constant $\gamma_a^{\mathrm{lin}} \in \mathbb{R}_+$.

For the second term in (3.51), since $(x_a, y_a) \in U$ is the unique minimizer of $\bar{J}^{\text{lin}}$ over $\mathcal{D}_a$, there exists $\delta > 0$ such that

$$\inf_{(x,y)\in\mathcal{D}_a\cap U^c} \bar{J}^{\text{lin}}(x,y) > \inf_{(x,y)\in\mathcal{D}_a} \bar{J}^{\text{lin}}(x,y) + \delta = \mathbb{J}^{\text{lin}}_{2,\lambda}(a) + \delta.$$

where we used (3.52) in the last equality. It is shown in [49] that $\{\tilde{S}^n\}_{n\in\mathbb{N}}$ satisfies an LDP with rate function $\bar{J}^{\text{sub}}$ in the sublinear regime. By the LDP upper bound for the sequence $\{\tilde{S}^n\}_{n\in\mathbb{N}}$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\tilde{S}^n \in \mathcal{D}_a \cap U^c\right) \le - \inf_{(x,y)\in\mathcal{D}_a\cap U^c} \bar{J}^{\text{lin}}(x,y) < -\mathbb{J}^{\text{lin}}_{2,\lambda}(a) - \delta.$$

Hence, the second term in (3.51) is negligible with respect to (3.53). Finally, combining (3.51) and (3.53), we obtain (3.19). □

3.3.4.5 Proof of Lemma 3.3.4

3.4 Volumetric estimates of Orlicz balls

Fix Orlicz functions V_1 and V_2, $s < 0$ and $t \in \mathbb{R}$, and consider the vector $(\sum_{i=1}^n V_1(X_i), \sum_{i=1}^n V_2(X_i))$, where $\{X_i\}_{i=1}^n$ is a sequence of i.i.d. random variables with distribution $\mu_{s,t}$, defined in (3.24). We proceed to find an asymptotic density estimate of the vector $(\sum_{i=1}^n V_1(X_i), \sum_{i=1}^n V_2(X_i))$ and then obtain sharp estimates of the intersection on integrating the density over the appropriate domain. However, there are two main obstacles. First, the properties of the rate function defined in (3.23), obtained in Lemma 3.4.1 are more involved than the one-dimensional counterpart. Second, technical issues arise in the use of Fourier analysis to obtain a sharp density estimate since for each $i \in \mathbb{N}$, the vector $(V_1(X_i), V_2(X_i))$ does not have a density. After dealing with these two problems in Lemma 3.4.1 and 3.4.4, we give the proof of Theorem 3.2.12 in the end of this section.

3.4.1 Preliminary results

3.4.1.1 Properties of function $\mathcal{J}_{V_1,V_2}$

Fix Orlicz functions V_1 and V_2. In this section, we first obtain properties of the function $\mathcal{J} := \mathcal{J}_{V_1,V_2}$ defined in (3.23).

Lemma 3.4.1. *Recall the definitions of the moment map M_V in (3.25) and measures $\mu_{s,t}$ in (3.24). Suppose V_1 and V_2 are two Orlicz functions satisfying Assumption 3.2.10 with $T \in [-\infty, 0]$. Fix $R_1, R_2 > 0$.*

1. *$\mathbb{R}_- \ni s \mapsto M_{V_1}(\mu_{s,0})$ is strictly increasing and there exists a unique $\tau_1(R_1) < 0$ such that $R_1 = M_{V_1}(\mu_{\tau_1(R_1),0})$.*

2. *If $T > -\infty$, then $t \mapsto M_{V_1}(\mu_{0,t})$ is strictly increasing and there exists a unique $\tau_2(R_1) < 0$ such that $R_1 = M_{V_1}(\mu_{0,\tau_2(R_1)})$.*

3. *Define $m_1(R_1) := M_{V_2}(\mu_{0,\tau_2(R_1)})$ and $m_2(R_1) := M_{V_2}(\mu_{\tau_1(R_1),0})$. If $R_2 > m_1(R_1)$ and (R_1, R_2) lies in the interior of the domain of $\mathcal{J}$ then $(s_1, s_2) = \nabla\Lambda(R_1, R_2)$, where $\Lambda := \Lambda_{V_1,V_2}$ is defined in (3.22), attains the supremum of $\mathcal{J}(R_1, R_2)$ in (3.23). Moreover $\partial_1\mathcal{J}(R_1, R_2) < 0$, and if $\partial_2\mathcal{J}(R_1, R_2) > 0$, then $R_2 > m_1(R_1)$.*

Proof. We start by proving property (1). Fix $t < s < 0$ and for $\lambda \in (0, 1)$, by Hölder's inequality, we have

$$\log \int_{\mathbb{R}} e^{(\lambda s+(1-\lambda)t)V_1(x)}dx \le \log\left(\left(\int_{\mathbb{R}} e^{sV_1(x)}dx\right)^{\lambda}\left(\int_{\mathbb{R}} e^{tV_1(x)}dx\right)^{1-\lambda}\right)$$
$$= \lambda \log \int_{\mathbb{R}} e^{sV_1(x)}dx + (1-\lambda)\log \int_{\mathbb{R}} e^{tV_1(x)}dx.$$

Moreover, the inequality is strict because by Definition 3.2.9 of Orlicz functions when $t < s$, $\mathbb{R} \ni x \mapsto e^{sV_1(x)-tV_1(x)}$ is not a constant and hence, the function $\mathbb{R}_- \ni s \mapsto \log \int e^{sV_1(x)}dx$ is strictly convex. $\mathbb{R}_- \ni s \mapsto M_{V_1}(\mu_{s,0})$ is the derivative of $\log \int e^{sV_1(x)}dx$, and thus the map $\mathbb{R}_- \ni s \mapsto M_{V_1}(\mu_{s,0})$ is strictly increasing. To show the existence of $\tau_1 = \tau_1(R_1) < 0$ such that $R_1 = M_{V_1}(\mu_{\tau_1(R_1),0})$, it suffices to show that $s \mapsto R_1 s - \log \int_{\mathbb{R}} e^{sV_1(x)}dx$ tends to $-\infty$ as $s \uparrow 0$ and as $s \to -\infty$ because it then implies by the mean value theorem that there exists τ_1 such that $R_1 = \frac{d}{ds}\log \int e^{sV_1(x)}dx|_{s=\tau_1} = M_{V_1}(\mu_{\tau_1,0})$. Since as $s \uparrow 0$, $-\log \int_{\mathbb{R}} e^{sV_1(x)}dx \to -\infty$, we only need to consider the case $s \to -\infty$. Now, for any $s < 0$ and $\varepsilon > 0$,

$$\begin{aligned}\frac{\log \int_{\mathbb{R}} e^{sV_1(x)}dx}{s} &\le \frac{\log \int_{\varepsilon>|x|>\varepsilon/2} e^{sV_1(x)}dx}{t}\\ &\le \frac{\log \int_{\varepsilon>|x|>\varepsilon/2} e^{sV_1(\varepsilon)}dx}{s}\\ &= \frac{\log(\varepsilon/2)+sV_1(\varepsilon)}{s},\end{aligned}$$

where the second inequality follows because V_1 is convex and symmetric about the origin and is therefore increasing in $\mathbb{R}_+$. Taking first the limit superior as $s \to -\infty$ and then the limit superior as $\epsilon \downarrow 0$, by the continuity of V_1, which is a consequence of V_1 being a convex function on R, we see that $\limsup_{s\to-\infty} \frac{\log \int_{\mathbb{R}} e^{sV_1(x)}dx}{s} \le 0$, and thus

$$\lim_{s\to-\infty} s\left(R - \frac{\log \int_{\mathbb{R}} e^{sV_1(x)}dx}{s}\right) = -\infty.$$

Finally, the uniqueness of such a τ_1 follows from the fact that $\mathbb{R}_- \ni s \mapsto M_{V_1}(\mu_{s,0})$ is strictly increasing. This proves the first property of the lemma.

For the second assertion, first note that

$$\frac{d}{dt}M_{V_1}(\mu_{0,t}) = M_{V_1\cdot V_2}(\mu_{0,t}) - M_{V_1}(\mu_{0,t})M_{V_2}(\mu_{0,t}),$$

which is greater than 0 by the symmetry of $\mu_{0,t}$ and the monotonicity of V_1 and V_2 in $\mathbb{R}_+$. Hence, $t \mapsto M_{V_1}(\mu_{0,t})$ is monotonically increasing. Fix $R_1 > 0$. To show the existence of $\tau_2 = \tau_2(R_1)$ such that $R_1 = M_{V_1}(\mu_{0,\tau_2})$, it suffices to show that $\lim_{t\to-\infty} M_{V_1}(\mu_{0,t}) = 0$ and $\lim_{t\uparrow T} M_{V_1}(\mu_{0,t}) = \infty$. Then the uniqueness of $\tau_2(R_1)$ follows from the monotonicity of $M_{V_1}(\mu_{0,t})$. Now, fix $\varepsilon > 0$. By Assumption 3.2.10, $V_2'(x) > 0$ for $x > 0$. This implies that $V_2(\varepsilon) > 0$ and we may choose $\delta > 0$ such that $V_2(\varepsilon) - V_2(\delta) > 0$ because V_2 is increasing in $\mathbb{R}_+$ by Definition 3.2.9. For $t < T$, we see that

$$\begin{aligned}
M_{V_1}(\mu_{0,t}) &= \frac{\int_{\mathbb{R}} V_1(x)e^{tV_2(x)}dx}{\int_{\mathbb{R}} e^{tV_2(x)}dx} \\
&\le \frac{\int_{|x|<\varepsilon} V_1(x)e^{tV_2(x)}dx}{\int_{\mathbb{R}} e^{tV_2(x)}dx} + \frac{\int_{|x|\ge\varepsilon} V_1(x)e^{tV_2(x)}dx}{\int_{\delta+1>|x|>\delta} e^{tV_2(x)}dx} \\
&\le V_1(\varepsilon) + \frac{\int_{|x|>\varepsilon} V_1(x)e^{(t-T)V_2(\varepsilon)+TV_2(x)}dx}{e^{tV_2(\delta)}} \\
&\le V_1(\varepsilon) + e^{t(V_2(\varepsilon)-V_2(\delta))}e^{-TV_2(\varepsilon)}\int_{|x|>\varepsilon} V_1(x)e^{TV_2(x)}dx,
\end{aligned}$$

where the second inequality follows because V_2 is increasing in $\mathbb{R}_+$. Letting first $t \to -\infty$, we obtain

$$\limsup_{t\to-\infty} M_{V_1}(\mu_{0,t}) \le V_1(\varepsilon).$$

Since $\varepsilon > 0$ is arbitrary, by the continuity of V_1 and the fact that $V_1(0) = 0$, on taking $\varepsilon \to 0$, we obtain $\lim_{t\to-\infty} M_{V_1}(\mu_{0,t}) = 0$.

We next show that $\lim_{t\uparrow T} M_{V_1}(\mu_{0,t}) = \infty$. First, suppose $T < 0$. Then, we have, by Fatou's lemma,

$$\liminf_{t\uparrow T} M_{V_1}(\mu_{0,t}) \geq \liminf_{t\uparrow T} \frac{\int_{\mathbb{R}} V_1(x)e^{tV_2(x)}dx}{\int_{\mathbb{R}} e^{tV_2(x)}dx} = \frac{\liminf_{t\uparrow T} \int_{\mathbb{R}} V_1(x)e^{tV_2(x)}dx}{\int_{\mathbb{R}} e^{TV_2(x)}dx} = \infty,$$

where the last equality follows from the definition of T in Assumption 3.2.10. Next, suppose $T = 0$. For $K > 0$, we have by the monotonicity of V_1 on $\mathbb{R}_+$ that for $t < 0$,

$$\begin{aligned}
\frac{1}{M_{V_1}(\mu_{0,t})} &= \frac{\int_{\mathbb{R}} e^{tV_2(x)}dx}{\int_{\mathbb{R}} V_1(x)e^{tV_2(x)}dx} \\
&\leq \frac{V_1(K)^{-1}\int_{|x|>K} V_1(x)e^{tV_2(x)}dx + \int_{|x|\leq K} e^{tV_2(x)}dx}{\int_{\mathbb{R}} V_1(x)e^{tV_2(x)}dx} \\
&\leq V_1(K)^{-1} + \frac{2e^{tV_2(K)}}{\int_{|x|\leq K} V_1(x)e^{tV_2(x)}dx}.
\end{aligned}$$

Letting $t \uparrow 0$, we obtain $\liminf_{t\uparrow 0} M_{V_1}(\mu_{0,t}) \geq V_1(K)$ by the definition of T in Assumption 3.2.10. Since $K > 0$ is arbitrary and $V_1(x) \to \infty$ as $x \to \infty$ by Assumption 3.2.9, sending $K \to \infty$ yields $\lim_{t\uparrow 0} M_{V_1}(\mu_{0,t}) = \infty$. This proves the second assertion of the lemma.

For the last assertion, note that $\mathcal{J}$ is the Legendre transform of Λ defined in (3.22). Hence, by duality of the Legendre transform, for $R_1, R_2 > 0$, $\nabla\mathcal{J}(R_1, R_2) = (s_1, s_2)$ where $(s_1, s_2) = (s_1(R_1, R_2), s_2(R_1, R_2))$ that attains the supremum in the expression for $\mathcal{J}(R_1, R_2)$ in (3.23). Since $R_2 > m_1(R_1)$, we see that $t_1 = \partial_1\mathcal{J}(R_1, R_2)$ is strictly less than 0 and (t_1, t_2) lies in the interior of the domain of Λ. Since by assumption, (R_1, R_2) also lies in the interior of the domain of $\mathcal{J}$, the duality of Legendre transform implies that $(t_1, t_2) = \nabla\Lambda(R_1, R_2)$. Next, if $t_2 = \partial_2\mathcal{J}(R_1, R_2) > 0$, then we have $t_1 = \partial_1\mathcal{J}(R_1, R_2) < 0$. This in turn implies that $R_2 > m_1(R_1)$. □

Lemma 3.4.2. *Let V_1 and V_2 be two Orlicz functions satisfying Assumption 3.2.10. Recall the definition of $\mathcal{J} := \mathcal{J}_{V_1,V_2}$ in (3.23). Fix R_1, $R_2 \in \mathbb{R}_+$. Let $m_1 := m_1(R_1)$ and $m_2 := m_2(R_1)$*

be defined as in Lemma 3.4.1. Then

1. *the infimum of $\mathcal{J}$ over $[0, R_1] \times [0, R_2]$ is uniquely attained at*

$$(u_1, u_2) = \begin{cases} (R_1, R_2), & \text{if} \quad m_1 < R_2 < m_2 \\ (R_1, m_2), & \text{if} \quad R_2 \geq m_2; \end{cases} \tag{3.54}$$

2. *the infimum of $\mathcal{J}$ over $[0, R_1] \times [R_2, \infty)$ is uniquely attained at*

$$(u_1, u_2) = \begin{cases} (R_1, m_2), & \text{if} \quad R_2 < m_2 \\ (R_1, R_2), & \text{if} \quad R_2 \geq m_2 \text{ and } \mathcal{J}(R_1, R_2) < \infty. \end{cases} \tag{3.55}$$

Furthermore, $\partial_2 \mathcal{J}(R_1, R_2) < 0$ for $m_1 < R_2 < m_2$ and $\partial_2 \mathcal{J}(R_1, m_2) = 0$.

Proof. Recall the definition of Λ in (3.22). Define $A := [0, R_1] \times [0, R_2]$. Let $\mathcal{J}$ achieve its minimum over A at (u_1, u_2), which can be assumed without loss of generality to belong to $D_{\mathcal{J}}$, the domain of $\mathcal{J}$. We claim that the $(u_1, u_2) \in \partial A$. To see why, suppose to the contrary that (u_1, u_2) lies in the interior of A. Then it must be that

$$\nabla \mathcal{J}(u_1, u_2) = 0. \tag{3.56}$$

Since Λ is convex, by duality of the Legendre transform, Λ is the Legendre transform of $\mathcal{J}$. Hence, (3.56) implies that $\Lambda(0, 0) = -\mathcal{J}(u_1, u_2)$. Since $(u_1, u_2) \in D_{\mathcal{J}}$, this implies that $\mathcal{J}(u_1, u_2)$ is finite, which contradict the fact that $\Lambda(0, 0) = \infty$ from (3.22). Hence, this proves the claim and in fact shows that $\nabla \mathcal{J}(u_1, u_2) \neq (0, 0)$.

We now turn to the proof of property (1). Fix R_1, $R_2 > 0$. From the definitions of $\tau_1(R_1)$ and $m_2(R_1)$ given in Lemma 3.4.1, $\nabla \Lambda(\tau_1(R_1), 0) = (R_1, m_2(R_1))$. Duality of the Legendre transform implies that $\nabla\mathcal{J}(R_1, m_2(R_1)) = (\tau_1(R_1), 0)$. For any $y > m_1(R_1)$, define $(t_1(y), t_2(y)) := \nabla\mathcal{J}(R_1, y)$. Then $t_2(m_2(R_1)) = 0$. By the strict convexity of $\mathcal{J}$, $y \mapsto t_2(y)$ is strictly increasing. This implies that $\partial_2\mathcal{J}(R_1, R_2) = t_2(R_2) < 0$ for $m_1(R_1) < R_2 < m_2(R_1)$ and $\partial_2\mathcal{J}(R_1, m_2) = t_2(m_2(R_1)) = 0$, and thus the infimum of $\mathcal{J}$ over A is (y_1, R_2) for some $y_1 \in [0, R_1]$ if $R_2 < m_2(R_1)$ and $(y_1, m_2(R_1))$ for some $y_1 \in [0, R_1]$ if $R_2 \geq m_2(R_1)$. Since by assumption $R_2 > m_1$, we see that $\partial_1\mathcal{J}(y_1, R_2) \leq \partial_1\mathcal{J}(R_1, R_2) < 0$, where the last inequality follows from Lemma 3.4.1. We conclude that the minimizer is (R_1, R_2) if $m_1 < R_2 < m_2(R_1)$ and $(R_1, m_2(R_1))$ if $R_2 \geq m_2(R_1)$.

Next, we turn to the proof of property (2). Define $B := [0, R_1] \times [R_2, \infty)$. From the argument above, we see that the minimizer is attained at (R_1, y) for some $y \in \mathbb{R}_+$. Since $\partial_2\mathcal{J}(R_1, y) \leq 0$ if $y \geq m_2(R_1)$ and the equality holds if and only if $y = m_2(R_1)$, we see that the minimizer is attained at $(R_1, m_2(R_1))$ if $R_2 < m_2(R_1)$ and (R_1, R_2) if $R_2 \geq m_2(R_1)$. $\square$

3.4.1.2 Intersections of Orlicz spheres

Lemma 3.4.3. *Let V_1 and V_2 be two Orlicz functions that satisfy Assumption 3.2.10. Then for any $z_1, z_2 \in \mathbb{R}_+$, the number of intersections of two Orlicz spheres, $C_1 := \{x \in \mathbb{R}^2 : V_1(x_1) + V_1(x_2) = z_1\}$ and $C_2 := \{x \in \mathbb{R}^2 : V_2(x_1) + V_2(x_2) = z_2\}$ is bounded by 8.*

Proof. By the symmetry of Orlicz functions, it suffices to show that there is at most 1 intersection in the region $\{(x_1, x_2) \in \mathbb{R}_+ : x_1 \geq x_2 > 0\}$. Now, suppose the two spheres intersect at a point (x_1^*, x_2^*) with $x_1^* \geq x_2^* > 0$. Then the slopes of C_1 and C_2 at (x_1^*, x_2^*) are $-V_1'(x_1^*)/V_1'(x_2^*)$ and $-V_2'(x_1^*)/V_2'(x_2^*)$, respectively. By the Assumption 3.2.10 that V_1'/V_2' is strictly increasing in $\mathbb{R}_+$, we see that the slope of C_2 at (x_1^*, x_2^*) is greater than that of C_1. This implies that there cannot be two intersections in the region

$\{(x_1, x_2) \in \mathbb{R}_+ : x_1 \geq x_2 > 0\}$. $\square$

3.4.2 Density estimates and Laplace approximations

3.4.2.1 Density estimates for a 2-dimensional vector

For $t := (t_1, t_2) \in \mathbb{R}_- \times \mathbb{R}$. Let $\{X_i\}_{i\in\mathbb{N}}$ be i.i.d. (absolutely) continuous random variables with law μ_t defined in (3.24). Define the 2-dimensional vector

$$S_t^n[V_1, V_2] := \frac{1}{n}\sum_{i=1}^{n} (V_1(X_i), V_2(X_i)) . \tag{3.57}$$

Lemma 3.4.4. *Let V_1 and V_2 be two Orlicz functions that satisfy Assumption 3.2.10. Fix $t := (t_1, t_2) \in \mathbb{R}_- \times \mathbb{R}$. Let $S_t^n = S_t^n[V_1, V_2]$ be defined as in (3.57). Then for each $n > 1$, S_t^n has a density, denoted as $f_{S_t^n}$ and for sufficiently large n, the inverse Fourier transform formula can be applied to $f_{S_t^n}$.*

Proof. Consider the transformation $\mathcal{T} : \mathbb{R}^2 \to \mathbb{R}_+ \times \mathbb{R}_+$ defined by

$$\mathcal{T}(x_1, x_2) := (V_1(x_1) + V_1(x_2), V_2(x_1) + V_2(x_2)) . \tag{3.58}$$

Hence, the Gacobian of T exists almost surely and is equal to the function $\mathcal{G}$ defined in (3.26),

$$\mathcal{G}(x_1, x_2) := \frac{\partial(z_1, z_2)}{\partial(x_1, x_2)} = \left|V_1'(x_1)V_2'(x_2) - V_1'(x_2)V_2'(x_1)\right| .$$

Moreover, Assumption 3.2.10 implies $\mathcal{G}$ is nonzero almost surely, which implies that $f_{S_t^2}$ has a density. Furthermore, this implies $f_{S_t^n}$ has a density for $n > 1$.

Let φ^n be the Fourier transform of $f_{S_t^n}$. Then for $s := (s_1, s_2) \in \mathbb{R}^2$,

$$\begin{aligned}\varphi^2(s_1, s_2) &:= \int_{\mathbb{R}^2} e^{i\langle s,x\rangle} f_{S_t^2}(dx)\\ &= \int_{\mathbb{R}^n} e^{i\sum_{j=1}^2 \langle s,(V_1(x_j),V_2(x_j))\rangle} f(x_1)\cdots f(x_n) dx_1 \cdots dx_n,\end{aligned}$$

where f is the density of X_i and is given by (3.24) to be

$$f(x) = \frac{e^{t_1 V_1(x)+t_2 V_2(x)}}{\int_{\mathbb{R}} e^{t_1 V_1(x)+t_2 V_2(x)} dx}$$

To prove the lemma, it suffices to establish the claim that φ^2 lies in $L^q(\mathbb{R}^2)$ for some $q > 2$ because by Hölder's inequality, the integrand in the last display lies in $L^1(\mathbb{R}^2)$ for all sufficiently large n.

To prove the claim, let $R(\mathcal{T}) \subset \mathbb{R}^2$ be the range of the transformation $\mathcal{T}$. For $z \in R(\mathcal{T})$, the solutions to $z = \mathcal{T}(x)$ coincide with the intersections of two Orlicz spheres $\{x \in \mathbb{R}^2 : V_1(x_1) + V_1(x_2) = z_1\}$ and $\{x \in \mathbb{R}^2 : V_2(x_1) + V_2(x_2) = z_2\}$. Let $\xi^z \subset \mathbb{N} \cup \{0\}$ be the number of solutions to $z = \mathcal{T}(x)$, which satisfies $\xi^z \in \{0, \ldots, 8\}$ due to Lemma 3.4.3, denote the solutions by $x_j^z \in \mathbb{R}^2$, $j = 1, \ldots, \xi^z$, let x_j^z. Under the transformation of $\mathcal{T}$, we have for $s = (s_1, s_2) \in \mathbb{R}^2$,

$$\varphi^2(s_1, s_2) = \int_{R(T)} e^{i\langle s,z\rangle} \sum_{j=1}^{\xi^z} f(x_{j,1}^z) f(x_{j,2}^z) \frac{1}{\mathcal{G}(x_{j,1}^z, x_{j,2}^z)} dz.$$

It suffices to show that $z \mapsto \sum_{j=1}^{\xi^z} f(x_{j,1}^z) f(x_{j,2}^z) \mathcal{G}^{-1}(x_{j,1}^z, x_{j,2}^z)$ lies in $L^p(R(\mathcal{T}))$ for some $p \in (0, 1)$,

$$L_T := \int_{R(T)} \left| \sum_{j=1}^{\xi^z} f(x_{j,1}^z) f(x_{j,2}^z) \frac{1}{\mathcal{G}(x_{j,1}^z, x_{j,2}^z)} \right|^p dz_1 dz_2 < \infty,$$

since then the Hausdorff-Young inequality [30, Theorem 8.21] implies φ^2 lies in $L^q(\mathbb{R}^2)$ for $q = p/(p-1) \in (2, \infty)$.

Since $\sup_{z\in\mathbb{R}^2} \xi^z \le 8$, by the inequality $|a_1 + \cdots + a_n|^p \le n^{p-1}(|a_1|^p + \cdots + |a_n|^p)$, we have that

$$
\begin{aligned}
L_T &\le 8^{p-1} \int_{R(T)} \sum_{j=1}^{\xi^z} \left| f(x_{j,1}^z) f(x_{j,2}^z) \frac{1}{\mathcal{G}(x_{j,1}^z, x_{j,2}^z)} \right|^p dz_1 dz_2 \\
&= 8^{p-1} \int_{\mathbb{R}^2} \left| f(x_1) f(x_2) \right|^p |\mathcal{G}(x_1, x_2)|^{1-p} \, dx_1 dx_2 \\
&= 8^{p-1} \int_{\mathcal{N}} \left| f(x_1) f(x_2) \right|^p \left| \frac{1}{\mathcal{G}(x_1, x_2)} \right|^{p-1} dx_1 dx_2 + 8^{p-1} \int_{\mathcal{N}^c} \left| f(x_1) f(x_2) \right|^p \left| \frac{1}{\mathcal{G}(x_1, x_2)} \right|^{p-1} dx_1 dx_2,
\end{aligned}
$$

where $\mathcal{N}$ is the neighborhood of the origin specified in condition (2) of Assumption 3.2.10. By condition (2) of Assumption 3.2.10, there exist $r_1, r_2 > 0$ such that $1/\mathcal{G}$ lies in $L^{r_1}(\mathcal{N})$ and $L^{r_2}(\mathcal{N}^c)$. Pick $p < 1 + \min\{r_1, r_2\}$. By Hölder's inequality,

$$
\begin{aligned}
L_T \le 8^{p-1} &\left(\int_{\mathcal{N}} \left| f(x_1) f(x_2) \right|^{\frac{pr_1}{r_1+1-p}} dx_1 dx_2 \right)^{\frac{r_1+1-p}{r_1}} \left(\int_{\mathcal{N}} \left| \frac{1}{\mathcal{G}(x_1, x_2)} \right|^{r_1} dx_1 dx_2 \right)^{\frac{p-1}{r_1}} \\
&+ \le 8^{p-1} \left(\int_{\mathcal{N}^c} \left| f(x_1) f(x_2) \right|^{\frac{pr_2}{r_2+1-p}} dx_1 dx_2 \right)^{\frac{r_2+1-p}{r_2}} \left(\int_{\mathcal{N}^c} \left| \frac{1}{\mathcal{G}(x_1, x_2)} \right|^{r_2} dx_1 dx_2 \right)^{\frac{p-1}{r_2}},
\end{aligned}
$$

which is finite because $(x_1, x_2) \mapsto f(x_1) f(x_2)$ lies in $L^r(\mathbb{R}^2)$ for any $r > 0$.

$\square$

3.4.2.2 A Laplace approximation result

Before giving the proof, we state a sharp density estimate for sum of i.i.d. random variables obtained in [6].

Lemma 3.4.5 (Theorem 3.1 and Remark 3.2 [6]). *Let $\{Y_n\}_{n\in\mathbb{N}}$ be a sequence of i.i.d. $\mathbb{R}^d$-valued random vectors whose log moment generating function Λ is finite in a neighborhood of the origin and let D_Λ be its domain. Define $\check{S}^n =: \sum_{i=1}^n Y_i$ and its density by $f_{\check{S}^n}$. Suppose that for some n, $\check{S}^n$ has a density to which the inversion Fourier theorem can be applied. Let I be the Legendre transform of Λ. Define the admissible domain $D_I := \{x \in \mathbb{R}^d : x = \nabla\Lambda(\lambda) \text{ for some } \lambda \in D_\Lambda^\circ\}$. Then for every $x \in D_I$,*

$$f_{\check{S}^n} = \left(\frac{n}{2\pi}\right)^{d/2} \det(\operatorname{Hess} I(x))^{1/2} e^{-nI(x)}(1+o(1)),$$

where the expansion holds uniformly for x in any compact subset contained in D_I.

3.4.2.3 Proof of Theorem 3.2.12

For a measurable set $U \subset \mathbb{R}$, Orlicz function V and $n \in \mathbb{N}$, define

$$B_V^n[U] := \left\{x \in \mathbb{R}^n : \frac{1}{n}\sum_{i=1}^n V(x_i) \subset U\right\}. \tag{3.59}$$

Lemma 3.4.6. *Let V_1 and V_2 be two Orlicz functions that satisfy Assumption 3.2.10. Fix $R_1 > 0$. Let $\Xi \subset \mathbb{R}_+$ be an interval with $\Xi \subset (m_1(R_1), \infty)$, where $m_1(R_1)$ is defined in Lemma 3.4.1. Suppose the infimum of $\mathcal{J} := \mathcal{J}_{V_1,V_2}$, defined in (3.23), over $[0, R_1] \times \Xi$ is attained uniquely at (u_1, u_2), and let U_i be a neighborhood of u_i for $i = 1, 2$. Then*

$$\left|B_{V_1}^n[[0,R_1]\cap U_1] \cap B_{V_2}^n[\Xi \cap U_2]\right| = \int_{\Xi\cap U_2} \frac{1}{2\pi}\frac{\det(\operatorname{Hess}\mathcal{J}(R_1,y_2))^{1/2}}{\left|\partial_1\mathcal{J}(R_1,y_2)\right|} e^{-n\mathcal{J}(R_1,y_2)} dy_2 (1+o(1)).$$

Proof. Fix Orlicz functions V_1, V_2, $t := (t_1, t_2) \in \mathbb{R}_- \times \mathbb{R}$ and let $S_t^n := S_t^n[V_1, V_2]$, defined in (3.57), for $n \in \mathbb{N}$ and $f_{S_t^n}$ be its density, which exists by Lemma 3.4.4. Fix $R_1 > 0$ and U_i as in the lemma. Then the volume of $A^n := B_{V_1}^n[[0,R_1] \cap U_1] \cap B_{V_2}^n[\Xi \cap U_2]$ can be written as

$$
\begin{aligned}
|A^n| &= \int_{A^n} dx \\
&= Z^n_{t_1,t_2} \int_{A^n} e^{-t_1 \sum_{i=1}^n V_1(x_i) - t_2 \sum_{i=1}^n V_2(x_i)} \prod_{i=1}^n \frac{1}{Z_{t_1,t_2}} e^{t_1 V_1(x_i) + t_2 V_2(x_i)} dx_i \\
&= Z^n_{t_1,t_2} \int_{y_1 \in [0,R_1] \cap U_1, y_2 \in \Xi \cap U_2} e^{-nt_1 y_1 - nt_2 y_2} f_{S^n_t}(y_1, y_2) dy_1 dy_2. \qquad (3.60)
\end{aligned}
$$

Recall that the law of X_1 is μ_t defined in (3.24). Define the log moment generating function,

$$\Lambda_t(s_1, s_2) := \log \mathbb{E}\left[e^{s_1 V_1(X_1) + s_2 V_2(X_1)}\right], \quad (s_1, s_2) \in \mathbb{R}^2,$$

and note that from (3.23) that, with Λ defined in (3.22),

$$\Lambda_t(s_1, s_2) = \Lambda(s_1 + t_1, s_2 + t_2) - \log Z_{t_1,t_2}.$$

Hence, by (3.23), it follows that for $(x_1, x_2) \in D_{\mathcal{J}}$,

$$\Lambda_t^*(x_1, x_2) - \log Z_{t_1,t_2} + t_1 x_1 + t_2 x_2 = \mathcal{J}(x_1, x_2), \qquad (3.61)$$

where we set $\mathcal{J} = \mathcal{J}_{V_1,V_2}$.

Since Λ_t is finite near the origin by assumption, using Lemma 3.4.4 to justify the application of Lemma 3.4.5 with $d = 2$, $Y_i = (V_1(X_i), V_2(X_i))$, $\{X_i\}_{n\in\mathbb{N}}$ i.i.d. with law μ_t, using (3.62), which also implies $\mathrm{Hess}\,\mathcal{J} = \mathrm{Hess}\,\Lambda_t^*$, we obtain

$$Z^n_{t_1,t_2} e^{-nt_1 x_1 - nt_2 x_2} f_{S^n_t}(x_1, x_2) = \frac{n}{2\pi} \det\left(\mathrm{Hess}\,\mathcal{J}(x_1, x_2)\right)^{1/2} e^{-n\mathcal{J}(x_1,x_2)}(1 + o(1)), \qquad (3.62)$$

where the estimate is uniform in any compact subset of $D_{\mathcal{J}}$. From (3.60) and (3.62), we then have the following estimate:

$$|A^n| = \int_{[0,R_1]\cap U_1 \times \Xi \cap U_2} \frac{n}{2\pi} \det(\operatorname{Hess} \mathcal{J}(y_1, y_2))^{1/2} e^{-n\mathcal{J}(y_1,y_2)} dy_1 y_2 (1 + o(1)).$$

We now obtain Laplace type asymptotics for the integral in the last display. For $y_2 \in \Xi \cap U_2$, applying Lemma 3.4.2(3) with $R_2 = y_2 > m_1$ by the assumption of Ξ, we have $\partial_1 \mathcal{J}(R_1, y_2) < 0$. Therefore, the minimum of $y_1 \mapsto \mathcal{J}(y_1, y_2)$ over $[0, R_1] \cap U_1$ is achieved at R_1. Hence, applying [83, Theorem 1 of Chapter II] to the inner integral below and setting both α and μ therein to 1, we obtain

$$\begin{aligned} |A^n| &= \int_{\Xi\cap U_2} \left(\int_{[0,R_1]\cap U_1} \frac{n}{2\pi} \det(\operatorname{Hess} \mathcal{J}(y_1, y_2))^{1/2} e^{-n\mathcal{J}(y_1,y_2)} dy_1 \right) dy_2 (1 + o(1)) \\ &= \int_{\Xi\cap U_2} \frac{1}{2\pi} \frac{\det(\operatorname{Hess} \mathcal{J}(R_1, y_2))^{1/2}}{\left| \partial_1 \mathcal{J}(R_1, y_2) \right|} e^{-n\mathcal{J}(R_1,y_2)} dy_2 (1 + o(1)). \end{aligned}$$

□

3.4.3 Proof of Theorem 3.2.12

Proof of Theorem 3.2.12. Fix Orlicz functions V_1, V_2 satisfying Assumption 3.2.10. The first assertion of the theorem follows from properties (1) and (2) of Lemma 3.4.1. We now turn to the second assertion. Let $\mathcal{J} := \mathcal{J}_{V_1,V_2}$. By Lemma 3.4.2, the infimum of $\mathcal{J}$ over $[0, R_1] \times [0, R_2]$ is uniquely attained at a point (u_1, u_2) that satisfies (3.54). For $i = 1, 2$, let U_i be an open neighborhood of u_i. We separate the volume into two parts

$$\left| \bigcap_{i=1}^{2} B^n_{V_i}(R_i) \right| = \left| \bigcap_{i=1}^{2} B^n_{V_i}[[0, R_i] \cap U_i] \right| + \left| \bigcap_{i=1}^{2} B^n_{V_i}[[0, R_i] \cap U_i^c] \right|. \tag{3.63}$$

For the first term in (3.63), recalling (3.59), we have by Lemma 3.4.6 with $\Xi = [0, R_2]$ that

$$\mathcal{V}^n := \left| \bigcap_{i=1}^{2} B^n_{V_i}[[0, R_i] \cap U_i] \right| = \int_{[0,R_2]\cap U_2} \frac{1}{2\pi} \frac{\det(\mathrm{Hess}\, \mathcal{J}(R_1, y_2))^{1/2}}{\left|\partial_1 \mathcal{J}(R_1, y_2)\right|} e^{-n\mathcal{J}(R_1,y_2)} dy_2 (1 + o(1)). \tag{3.64}$$

To obtain the asymptotics of this integral, we discuss three different regimes.

1. Subcritical ($m_1 < R_2 < m_2$): Since $\partial_2 \mathcal{J}(R_1, R_2) < 0$ and (3.4.2), by Lemma 3.4.2, applying [83, Theorem 1 of Chapter II] with both α and μ therein set equal to 1, we obtain

$$\begin{aligned} \mathcal{V}^n &= \frac{n}{2\pi} \frac{1}{n^2} \frac{\det(\mathrm{Hess}\, \mathcal{J}(R_1, R_2))^{1/2}}{|\partial_1 \mathcal{J}(R_1, R_2)|\, |\partial_2 \mathcal{J}(R_1, R_2)|} e^{-n\mathcal{J}(R_1,R_2)} (1 + o(1)) \\ &= \frac{1}{2\pi n} \frac{\det(\mathrm{Hess}\, \mathcal{J}(R_1, R_2))^{1/2}}{|\partial_1 \mathcal{J}(R_1, R_2)|\, |\partial_2 \mathcal{J}(R_1, R_2)|} e^{-n\mathcal{J}(R_1,R_2)} (1 + o(1)). \end{aligned} \tag{3.65}$$

2. Critical ($R_2 = m_2$): In the so-called critical regime, we have instead $\partial_2 \mathcal{J}(R_1, R_2) = 0$ by Lemma 3.4.2. Hence, applying [83, Theorem 1 of Chapter II] to the integral in (3.64), but with α and μ therein replaced by 1 and 2, respectively, we obtain

$$\begin{aligned} \mathcal{V}^n &= \frac{n}{2\pi} \frac{\sqrt{\pi}}{\sqrt{2} n^{3/2}} \frac{\det(\mathrm{Hess}\, \mathcal{J}(R_1, m_2))^{1/2}}{|\partial_1 \mathcal{J}(R_1, m_2)|\, \sqrt{\partial_{22} \mathcal{J}(R_1, m_2)}} e^{-n\mathcal{J}(R_1,m_2)} (1 + o(1)) \\ &= \frac{1}{2\sqrt{2\pi n}} \frac{\det(\mathrm{Hess}\, \mathcal{J}(R_1, m_2))^{1/2}}{|\partial_1 \mathcal{J}(R_1, m_2)|\, \sqrt{\partial_{22} \mathcal{J}(R_1, m_2)}} e^{-n\mathcal{J}(R_1,m_2)} (1 + o(1)). \end{aligned} \tag{3.66}$$

3. Supercritical ($R_2 > m_2$): For the supercritical regime, $\partial_2 \mathcal{J}(R_1, m_2) = 0$ by Lemma 3.4.2. The asymptotic expression of $\mathcal{V}^n$ is twice that of the critical case

because by (3.64),

$$\mathcal{V}^n = \int_{[0,m_2]\cap U_2} \frac{1}{2\pi} \frac{\det(\mathrm{Hess}\,\mathcal{J}(R_1,y_2))^{1/2}}{\left|\partial_1 \mathcal{J}(R_1,y_2)\right|} e^{-n\mathcal{J}(R_1,y_2)} dy_2 (1+o(1)) \tag{3.67}$$

$$+ \int_{[m_2,R_2]\cap U_2} \frac{1}{2\pi} \frac{\det(\mathrm{Hess}\,\mathcal{J}(R_1,y_2))^{1/2}}{\left|\partial_1 \mathcal{J}(R_1,y_2)\right|} e^{-n\mathcal{J}(R_1,y_2)} dy_2 (1+o(1))$$

$$= \frac{n}{2\pi} \frac{2\sqrt{\pi}}{\sqrt{2}n^{3/2}} \frac{\det(\mathrm{Hess}\,\mathcal{J}(R_1,m_2))^{1/2}}{|\partial_1 \mathcal{J}(R_1,m_2)|\,\sqrt{\partial_{22}\mathcal{J}(R_1,m_2)}} e^{-n\mathcal{J}(R_1,m_2)}(1+o(1))$$

$$= \frac{1}{\sqrt{2\pi n}} \frac{\det(\mathrm{Hess}\,\mathcal{J}(R_1,m_2))^{1/2}}{|\partial_1 \mathcal{J}(R_1,m_2)|\,\sqrt{\partial_{22}\mathcal{J}(R_1,m_2)}} e^{-n\mathcal{J}(R_1,m_2)}(1+o(1)), \tag{3.68}$$

where the second equality follows from [83, Theorem 1 of Chapter II] with α and μ therein replaced by 1 and 2, respectively.

Since (u_1,u_2) achieves the unique infimum of $\mathcal{J}$ over $[0,R_1]\times[0,R_2]$ and $u_i \in U_i$ for $i = 1,2$, there exists $\delta > 0$ such that

$$\inf_{x_1\in[0,R_1]\cap U_1^c,\, x_2\in[0,R_2]\cap U_2^c} \mathcal{J}(x_1,x_2) > \mathcal{J}(u_1,u_2) + \delta. \tag{3.69}$$

By [49, Proposition 3.11 and Remark 3.14], we see that the second term in (3.63) satisfies

$$\limsup_{n\to\infty} \frac{1}{n} \log \left| \bigcap_{i=1}^{2} B^n_{V_i}[[0,R_i]\cap U_i^c] \right| < -\mathcal{J}(u_1,u_2) - \delta. \tag{3.70}$$

Combining (3.70) and the estimates in (3.65), (3.66) and (3.68), we see that the second term on the right-hand-side of (3.63) is negligible with respect to $\mathcal{V}^n$ in each regime. Hence, the theorem follows from (3.63), (3.65), (3.66) and (3.68).

$\square$

3.4.4 Volume of Orlicz balls

We start with a sharp estimate of the asymptotic volume of Orlicz balls, which is also obtained in [42]. We provide a different proof in Section 3.6 to make the paper self-contained.

Proposition 3.4.7. *Let V be an Orlicz function and recall the definition of Ψ_V^* from (**??**). Then for $R > 0$, there exists a constant $\tau_R \in \mathbb{R}_-$ that uniquely achieved the supremum of $\Psi_V^*(R)$ in (**??**) and satisfies*

$$R = \frac{\int_{\mathbb{R}} V(x) e^{\tau_R V(x)} dx}{Z_{V,\tau_R}}. \tag{3.71}$$

Moreover, if σ_R^2 is the variance of μ_{V,τ_R}, defined in (3.24), then

$$\left|B_V^n(R)\right| = \frac{1}{\sqrt{2\pi n}\sigma_R \left|\tau_R\right|} e^{-n\Psi_V^*(R)}(1 + o(1)). \tag{3.72}$$

Remark 3.4.8. Note that the form of the sharp estimate (3.72) is very similar to the sharp large deviation estimate in (3.1). Indeed, formally (3.72) can be viewed as saying

$$\left|B_V^n(R)\right| = \left|\left\{x \in \mathbb{R}^n : \sum_{i=1}^n V(x_i) \leq n\right\}\right| \text{``} = \text{''} \mathbb{P}\left(\sum_{i=1}^n V(X_i) < R\right),$$

where "$\{X_i\}_{i\in\mathbb{N}}$ is distributed according to Lebesgue measure". Hence, Ψ_V, defined in (4.15), is the analogue of the log-moment generating function and the tail estimate (3.72) takes a form similar to (3.1).

Combining Proposition 3.4.7 and Theorem 3.2.12, we immediately obtain Corollary 3.2.14

Proof of Corollary 3.2.14. Fix Orlicz functions V_1 and V_2 and let $R_1, R_2 > 0$. Recall the definitions of Λ in (3.22) and let $\tau := \tau_1(R_1)$ be the unique value such that $R_1 = M_{V_1}(\mu_{\tau_1(R_1),0}, 0)$ as in Theorem 3.2.12(1), which by (3.24) and (3.22) implies $R_1 = \partial_1\Lambda(\tau, 0)$. Also note that τ coincides wit the quantity τ_{R_1} in Proposition 3.4.7. On the other hand, from the definition of m_2 in Theorem 3.2.12(2), (3.24) and (3.22), it follows that $m_2(R_1) = M_{V_2}(\mu_{\tau,0}) = \partial_2\Lambda(\tau, 0)$. Therefore, when $R_2 = m_2(R_1)$, it follows that $\nabla\Lambda(\tau, 0) = (R_1, R_2)$ and $(\tau, 0) \in D_\Lambda$ achieves the supremum in (3.23) since Λ is the Legendre transform of $\mathcal{J}$ and by duality of the Legendre transform $\nabla\mathcal{J}(R_1, R_2) = (\tau, 0)$. Hence, we have $\mathcal{J}(R_1, R_2) = \Psi^*_{V_1}(R_1)$ defined in (**??**). Let $\sigma = \sigma_{R_1}$ and τ_{R_1} be defined as in Proposition 3.4.7. Combining Proposition 3.4.7, Theorem 3.2.12(2), and the identities $\tau_{R_1} = \tau$ and $\mathcal{J}(R_1, R_2) = \Psi^*_{V_1}(R_1)$, we obtain

$$\lim_{n\to\infty} \frac{\left|B^n_{V_1}(R_1) \cap B^n_{V_2}(R_2)\right|}{\left|B^n_{V_1}(R_1)\right|} = \frac{\sigma \det(\operatorname{Hess}\mathcal{J}(R_1, R_2))^{1/2}}{2\sqrt{\partial_{22}\mathcal{J}(R_1, R_2)}}.$$

By duality of the Legendre transform, we see that

$$\frac{\partial_{22}\mathcal{J}(R_1, R_2)}{\det(\operatorname{Hess}\mathcal{J}(R_1, R_2))} = \partial_{11}\left(\log \int_{\mathbb{R}} e^{s_1 V_1(x) + s_2 V_2(x)} dx\right)\Bigg|_{(s_1, s_2) = (\tau, 0)} = \sigma^2.$$

The corollary then follows from the last two displays. □

3.5 Examples verifying the assumptions

3.5.1 Proof of Theorem 3.2.11

Proof of Theorem 3.2.11. Fix an Orlicz function V such that (V, x^2) satisfies Assumption 3.2.10. Let $X^n = X^{n,V}$ be uniformly distributed on $B^n_V(1)$ and let f_{Q^n} be the density of

$Q^n := \sum_{i=1}^n (X_i^n)^2/n$. For a measurable set $\Upsilon \subset \mathbb{R}$, define $B_V^n[\Upsilon] := \{x \in \mathbb{R}^n : \frac{1}{n}\sum_{i=1}^n V(x_i) \leq 1, \frac{1}{n}\sum_{i=1}^n x_i^2 \in \Upsilon\}$. Then, we have

$$\mathbb{P}\left(S^n \in \Upsilon\right) = \frac{\left|B_V^n[\Upsilon]\right|}{\left|B_V^n(1)\right|}. \tag{3.73}$$

For $t < 0$, recalling from (3.24) that $Z_{V,t} := \int_{\mathbb{R}} e^{tV(x)}dx$, note that

$$\left|B_V^n[\Upsilon]\right| = \int_{B_V^n[\Upsilon]} dx = Z_{V,t}^n \int_{B_V^n[\Upsilon]} e^{-t\sum_{i=1}^n V(x_i)} \prod_{i=1}^n \frac{1}{Z_{V,t}^n} e^{tV(x_i)} dx_i.$$

If we define $f_{\bar{Q}^n}$ to be the density of $\bar{Q}^n := \frac{1}{n}(\sum_{i=1}^n V(Y_i), \sum_{i=1}^n Y_i^2)$ with $\{Y_i\}_{i\in\mathbb{N}}$ i.i.d. distributed according to $\mu_{V,t}$, defined in (**??**), and $D := \{y \in \mathbb{R}^2 : y_1 \in [0,1], y_2 \in \Upsilon\}$, then we have

$$\left|B_V^n[\Upsilon]\right| = Z_{V,t}^n \int_D e^{-nty_1} f_{\bar{Q}^n}(y)dy. \tag{3.74}$$

Next, define

$$\Lambda_t(s_1, s_2) := \log \mathbb{E}\left[e^{s_1 V(X_1) + s_2 X_1^2}\right]$$

and let Λ_t^* be the Legendre transform of Λ_t. Denoting $\mathcal{J} = \mathcal{J}_{V,x^2}$, by (3.23), for $(x_1, x_2) \in D_{\mathcal{J}}$, we have

$$\begin{aligned}
&\Lambda_t^*(x_1, x_2) - \log Z_{V,t} + tx_1 && (3.75)\\
&\quad = \sup_{s_1\in\mathbb{R}, s_2\in\mathbb{R}} \left\{x_1(s_1+t) + x_2 s_2 - \log Z_{V,t} - \log \int_{\mathbb{R}} e^{s_1 V(x) + s_2 x^2} dx\right\} \\
&\quad = \sup_{s_1+t<0, s_2\in\mathbb{R}} \left\{x_1(s_1+t) + x_2 s_2 - \log \int_{\mathbb{R}} e^{(s_1+t)V(x) + s_2 x^2} \frac{1}{Z_{V,t}} e^{tV(x)} dx\right\} \\
&\quad = \mathcal{J}(x_1, x_2), && (3.76)
\end{aligned}$$

where the second inequality follows since V is superquadratic by Assumption 3.2.10 and hence the integral is infinite whenever $s_1 + t > 0$.

Note that Λ_t is finite near the origin. Lemma 3.4.4 ensures that there exists $N \in \mathbb{N}$ such that for $n \geq N$, the inverse Fourier transform can be applied to $f_{\bar{Q}^n}$. Therefore, we may apply Lemma 3.4.5 with $d = 2$ to conclude that, for x in the admissible domain D_{V,x^2},

$$f_{\bar{Q}^n}(x) = \frac{n}{2\pi} \det\left(\operatorname{Hess} \Lambda_t^*(x)\right)^{1/2} e^{-n\Lambda_t^*(x)}(1 + o(1)),$$

where the estimate is uniform in any compact neighborhood of $x \in D_{V,x^2}$. Recall $D_J := \{x \in \mathbb{R}_+ : (1, x) \in D_{V,x^2}\}$. Now, by (3.73), (3.74) and (3.76),

$$\begin{aligned} f_{Q^n}(x_2) &= \frac{Z_{V,t}^n}{\left|B_V^n(1)\right|} \int_{[0,1]} e^{-ntx_1} f_{\bar{Q}^n}(x_1, x_2) dx_1 \\ &= \frac{1}{\left|B_V^n(1)\right|} \frac{n}{2\pi} \int_{[0,1]} \det\left(\operatorname{Hess} \mathcal{J}(x_1, x_2)\right)^{1/2} e^{-n\mathcal{J}(x_1,x_2)} dx_1 (1 + o(1)). \end{aligned}$$

For fixed $x_2 \in D_J$, the minimum of $x_1 \mapsto \mathcal{J}(x_1, x_2)$ over $[0, 1]$ is attained at $x_1 = 1$ by Lemma 3.4.2. To simplify the integral, apply [83, Theorem 1 of Chapter II], with $s = 0$, $\alpha = 1$ and $\mu = 1$, to obtain for $x \in D_J$,

$$f_{Q^n}(x) = \frac{1}{\left|B_V^n(1)\right|} \frac{n}{2\pi} \frac{1}{n} \frac{\det\left(\operatorname{Hess} \mathcal{J}(1, x)\right)^{1/2}}{|\partial_1 \mathcal{J}(1, x)|} e^{-n\mathcal{J}(1,x)}(1 + o(1)). \tag{3.77}$$

Using Proposition 3.4.7, to get an expression for $\left|B_V^n(1)\right|$, we see that there exist $\sigma_1 > 0$, $\tau_1 < 0$ such that

$$\begin{aligned} f_{Q^n}(x) &= \sqrt{2\pi n}\sigma_1 |\tau_1| \frac{n}{2\pi} \frac{1}{n} \frac{\det\left(\operatorname{Hess} \mathcal{J}(1, x)\right)^{1/2}}{|\partial_1 \mathcal{J}(1, x)|} e^{-n\left(\mathcal{J}(1,x) - \Psi_V^*(1)\right)}(1 + o(1)) \\ &= \sqrt{n} \frac{\sigma_1 |\tau_1|}{\sqrt{2\pi}} \frac{\det\left(\operatorname{Hess} \mathcal{J}(1, x)\right)^{1/2}}{|\partial_1 \mathcal{J}(1, x)|} e^{-n\left(\mathcal{J}(1,x) - \Psi_V^*(1)\right)}(1 + o(1)) \end{aligned}$$

$$= \sqrt{n}\frac{\sigma_1 |\tau_1|}{\sqrt{2\pi}} \frac{\det(\mathrm{Hess}\,\mathcal{J}(1,x))^{1/2}}{|\partial_1 \mathcal{J}(1,x)|} e^{-nJ_V(\sqrt{x})}(1+o(1)),$$

where the last equality follows from (3.27). This proves that the sequence $\{X^{n,V}\}_{n\in\mathbb{N}}$ satisfies Assumption 3.2.1 with $J = J_V$, $\alpha = 1/2$ and h as stated in the theorem.

To prove the second assertion of the theorem, it remains to show that y_a, defined in the theorem, belongs to D_J. Recall the definition of $\bar{J}^{\mathrm{sub}}$ in (3.13) and $\mathcal{D}_a$ in (3.14), observe that

$$\inf_{\mathcal{D}_a} \bar{J}^{\mathrm{sub}}(x,y) = \inf_{y>|a|^2} \left\{ -\frac{1}{2}\log\left(1-\frac{|a|^2}{y}\right) + J(\sqrt{y}) \right\}.$$

Taking the derivative with respect to y of the term being infimized on the righ-hand side, we obtain

$$-\frac{1}{2}\frac{|a|^2}{(y-|a|^2)y} + \frac{1}{2\sqrt{y}}J'(\sqrt{y}).$$

Since $y > |a|^2$, the minimizer y_a satisfies $J'(y_a) > 0$, and thus (3.27) implies $\partial_2 \mathcal{J}_{V,x^2}(1,y_a) > 0$. Applying Lemma 3.4.1(3) with $V_1 = V$, $V_2(x) = x^2$, $R_1 = 1$ and $R_2 = y_a$, it follows that $y_a > m_1(1)$. Hence $(1, y_a) \in D_{V,x^2}$, which implies that $y_a \in D_J$. □

3.6 Sharp volume estimates of Orlicz balls

To obtain the sharp estimate of Orlicz ball. We follow the proof of [9] with a slight modification.

Proof of Proposition 3.4.7. Fix $R > 0$. The existence and uniqueness of $\tau := \tau_R$ follows from the first assertion of Lemma 3.4.1. Since τ achieves the maximum in (**??**), we obtain (3.71). By (**??**) and the definition of $\tau := \tau_R$,

$$\Psi_V^*(R) = R\tau - \log Z_{V,\tau}. \tag{3.78}$$

To calculate the volume of $B_V^n(R)$, recalling the measure $\mu_{V,\tau}$ from (**??**), we write

$$\begin{aligned}\left|B_V^n(R)\right| &= \int_{\left\{x\in\mathbb{R}^n:\frac{1}{n}\sum_{i=1}^n V(x_i)\le R\right\}} dx \\ &= \int_{\left\{x\in\mathbb{R}^n:\frac{1}{n}\sum_{i=1}^n V(x_i)\le R\right\}} Z_{V,\tau}^n e^{-\tau\sum_{i=1}^n V(x_i)} \prod_{i=1}^n \mu_{V,\tau}(dx_i).\end{aligned}$$

This can be viewed as "tilting" the uniform measure on $B_V^n(R)$ by $\otimes^n \mu_{V,\tau}$ restricted to $B_V^n(R)$.

Let $\{Y_i\}_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables distributed according to $\mu_{V,\tau}$ and let σ^2 be the variance of Y_i. Note that (3.71) implies $\mathbb{E}[V(Y_1)] = R$ and for $n \in \mathbb{N}$, define

$$U^n := \frac{\sum_{i=1}^n V(Y_i) - nR}{n^{1/2}\sigma},$$

and let H^n be the cumulative distribution function (cdf) of U^n. Then, integration by parts yields

$$\begin{aligned}\left|B_V^n(R)\right| &= Z_{V,\tau}^n e^{-nR\tau} \int_{(-\infty,0]} e^{-n^{1/2}\sigma\tau x} dH^n(x) \\ &= Z_{V,\tau}^n e^{-nR\tau}\left(e^{-n^{1/2}\sigma\tau x}(H^n(x) - H(0))\Big|_{x=-\infty}^{x=0} + n^{1/2}\sigma\tau \int_{-\infty}^0 e^{-n^{1/2}\sigma\tau x}(H^n(x) - H(0))dx \right) \\ &= Z_{V,\tau}^n e^{-nR\tau} n^{1/2}\sigma\tau \int_{-\infty}^0 e^{-n^{1/2}\sigma\tau x}(H^n(x) - H(0)dx, \end{aligned} \tag{3.79}$$

where the last equation uses the fact that $\lim_{x\to-\infty} e^{-n^{1/2}\sigma\tau x}(H^n(x) - H(0)) = 0$ since $\tau < 0$.

Let Φ be the cdf of a standard normal random variable. Since V is convex and

$\lim_{x\to\infty} V(x) = \infty$, and the density $\mu_{V,tau}$ of Y_1 is absolutely continuous with respect to Lebesgue measure, $V(Y_1)$ is not latticed valued. By [26, Theorem 2]

$$H^n(x) = \widetilde{H}^n(x) + o(n^{-1/2})$$

uniformly in $x \in \mathbb{R}$ and

$$\widetilde{H}^n(x) := \Phi(x) - \frac{\chi_3}{6\sigma^3 n^{1/2}}(x^2-1)\frac{1}{2\pi}e^{-x^2/2},$$

where χ_3 is the third moment of $V(Y_1)$. Since $\tau < 0$, this implies

$$n^{1/2}\sigma\tau \int_{-\infty}^{0} e^{-n^{1/2}\sigma\tau x}(H^n(x) - H(0)dx = \widetilde{\Gamma}_n + o\left(n^{-1/2}\right), \tag{3.80}$$

where

$$\widetilde{\Gamma}_n := n^{1/2}\sigma\tau \int_{-\infty}^{0} e^{-n^{1/2}\sigma\tau x}(\widetilde{H}^n(x) - \widetilde{H}^n(0)dx = \int_{-\infty}^{0} e^{-n^{1/2}\sigma\tau x}(\widetilde{H}^n)'(x)dx + o\left(n^{-1/2}\right),$$

with the last equation follows by integration by parts. Now, the Fourier transform of $e^{-n^{1/2}\sigma\tau x}1_{(-\infty,0)}$ is $-1/(n^{1/2}\sigma\tau - i\xi)$ and that of $(\widetilde{H}^n)'$ is

$$e^{-\xi^2/2}\left(1 + \frac{\chi_3}{6\sigma^2 n^{1/2}}(-i\xi)^3\right).$$

Parseval's formula and the symmetry of the Gaussian density shows that

$$\begin{aligned}
\widetilde{\Gamma}_n &= \frac{1}{2\pi}\int_{\mathbb{R}} \frac{-1}{n^{1/2}\sigma\tau + i\xi}\left(1 + \frac{\chi_3}{6\sigma^2 n^{1/2}}(-i\xi)^3\right)e^{-\xi^2/2}d\xi \\
&= \frac{-1}{2\pi\sqrt{n}\sigma\tau}\int_{\mathbb{R}} \frac{n\sigma^2\tau^2 - i\sqrt{n}\sigma\tau\xi}{n\sigma^2\tau^2 + \xi^2}\left(1 + \frac{\chi_3}{6\sigma^2 n^{1/2}}(-i\xi)^3\right)e^{-\xi^2/2}d\xi \\
&= \frac{-1}{2\pi\sqrt{n}\sigma\tau}\int_{\mathbb{R}} \left(\frac{n\sigma^2\tau^2}{n\sigma^2\tau^2 + \xi^2} + \frac{\tau^2\chi_3}{6\sigma(n\sigma^2\tau^2 + \xi^2)}\xi^4\right)e^{-\xi^2/2}d\xi
\end{aligned}$$

$$= \frac{-1}{\sqrt{2\pi n}\sigma\tau}(1 + o(1)), \tag{3.81}$$

where the third equality follows because $\widetilde{\Gamma}_n$ is real and the last inequality by the dominated convergence theorem.

Combining this with (3.79), (3.80) and (3.81), we finally obtain

$$\left|B_V^n(R)\right| = \frac{-1}{\sqrt{2\pi n}\sigma\tau} Z_{V,\tau}^n e^{-n\tau}(1 + o(1)) = \frac{-1}{\sqrt{2\pi n}\sigma\tau} e^{-n\Psi_V^*(R)}(1 + o(1)).$$

Hence, (3.72) follows on observing $\tau \in \mathbb{R}_-$. □

3.7 Proof of Lemma 3.2.4

Proof of Lemma 3.2.4. We can rewrite the rate function as

$$I_{\mathbf{A}X,k}(x) = \inf_{z > \|x\|_2} \left\{ -\frac{1}{2}\log\left(1 - \frac{\|x\|_2}{z^2}\right) + J(z) \right\} = \hat{I}_{\mathbf{A}X,k}(\|x\|_2).$$

For the second claim, let $r_1 > r_2 > 0$. We have

$$-\frac{1}{2}\log\left(1 - \frac{r_1}{z^2}\right) \geq -\frac{1}{2}\log\left(1 - \frac{r_2}{z^2}\right).$$

Since $\{z > r_1\} \subset \{z > r_2\}$, we then obtain

$$\hat{I}_{\mathbf{A}X,k}(r_1) = \inf_{z > r_1}\left\{-\frac{1}{2}\log\left(1 - \frac{r_1}{z^2}\right) + J(z)\right\} \geq \inf_{z > r_2}\left\{-\frac{1}{2}\log\left(1 - \frac{r_2}{z^2}\right) + J(z)\right\} = \hat{I}_{\mathbf{A}X,k}(r_2).$$

□

Chapter Four

Quenched sharp large deviation estimates for random projections

4.1 Introduction

In this chapter, we first describe analytic sharp large deviation results obtained in [58] in the setting of ℓ_p^n balls and spheres (see Section 4.3). To provide a flavor of the main ingredients of the proofs of sharp large results for projections, we then establish such large deviation estimates for projections of product measures (see Theorem 4.2.3), which provide a refinement of rough large deviation results for product measures obtained in [33]. The refined large deviation results for both product measure and ℓ_p^n balls mentioned above are obtained in the quenched setting, namely conditioned on the sequence of projection directions. It is noteworthy that while the (quenched) large deviation rate function obtained in [34] coincides for balls and spheres, the (quenched) sharp large deviations estimates of [58] distinguish between these different measures. This shows that the prefactor does indeed capture important additional geometric information beyond the rate function. We then use the tilted measure arising in the large deviations analysis to develop importance sampling algorithms for estimating tail probabilities.

In addition to providing a mechanism for distinguishing between high-dimensional distributions, large deviation or tail estimates also provide insight into outliers, which are of increasing importance with the preponderance of high-dimensional data. Going beyond ℓ_p^n balls, Chapter 2 introduced a general condition on sequences of high-dimensional measures called the asymptotic thin shell condition, under which the corresponding sequence of (possibly multidimensional) random projections satisfies an (annealed) LDP. The tail probabilities estimated in the case of ℓ_p^n balls allow one to estimate geometric quantities of relevance such as the fraction of volume of a spherical p-cap of an ℓ_p^n ball. Thus, sharp large deviation estimates for more general sequences of measures and related computational algorithms could potentially be useful more broadly in asymptotic convex geometry.

4.1.1 Prior Work on Sharp Large Deviation Estimates for Sums

The first sharp large deviation results for sums were obtained by Bahadur and Ranga-Rao in [10]. Given a sequence of i.i.d random variables $\{X_n\}_{n\in\mathbb{N}}$, under suitable assumptions they showed that

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq a\right) = \frac{e^{-n\mathbb{I}(a)}}{\sigma_a\tau_a\sqrt{2\pi n}}\left(1+o(1)\right), \tag{4.1}$$

where $\mathbb{I}$ is the Legendre transform of the log moment generating function, and $\sigma_a, \tau_a > 0$ are the variance under the tilted measure and the minimizer that attains the infimum in the Legendre transform $\mathbb{I}(a)$. The main idea behind the proof of [10] is to express the probability on the left-hand side as an expectation with respect to a tilted i.i.d. measure, with marginals that have mean a, then use Parseval's identity to transform the expectation under the tilted product measure and finally invoke quantitative versions of the central limit theorem under the tilted measure to obtain the estimate.

In this chapter, we consider instead the random projection of i.i.d. random variables. Consider a probability space $(\mathbb{P}, \Omega, \mathcal{F})$. Let S^{n-1} denote the unit sphere in $\mathbb{R}^n$ and define $\mathbb{S}$ to be the sequence space of unit vectors in each dimension:

$$\mathbb{S} := \prod_{n=1}^{\infty} S^{n-1}. \tag{4.2}$$

Let Θ be a random element defined on the probability space and take values in the sequence space $\mathbb{S}$, defined in (4.2), let $\Theta^n \in S^{n-1}$ denote the n-th element of that sequence.

Consider a measure σ on the sequence space $\mathbb{S}$ which has the property that for each $n \in \mathbb{N}$, the image measure under the coordinate map, $\Theta = (\Theta^1, \Theta^2, \cdots, \Theta^n, \cdots) \mapsto \Theta^n$, coincides with the unique rotation invariant measure σ^{n-1} on S^{n-1}. Note that we do not assume that the coordinates $\{\Theta^i\}_{i\in\mathbb{N}}$ are independent. For $\theta \in \mathbb{S}$, we will denote $\mathbb{P}_\theta$ to be the probability measure conditioning on the sequence θ. Now the random projection can be reformulated as a randomly weighted sum

$$\frac{1}{n}\sum_{i=1}^{n} X_i^n \left(\sqrt{n}\Theta_i^n\right),$$

where we condition on $\theta^n = (\theta_1^n, \ldots, \theta_n^n)$, a random vector sampled from the unit sphere in $\mathbb{R}^n$. Here we renormalized the random direction so that $\sqrt{n}\theta_i^n$ is typically of order 1. The main difference from [10] is that conditioned on the direction, though the terms in the sum we consider here are still independent, they are no longer identically distributed anymore. The result analogous to (4.1) for such sums is given in Theorem 4.2.3.

Another result for conditional sharp large deviation estimate was previously obtained by Bovier and Meyer in [17] where they looked at i.i.d. weights instead. In [40], Joutard showed a sharp large deviation principle for arbitrary sequences. The proof technique is similar to [17] by adopting [19]. However, our rate functions do not satisfy their assumption (A.2). Some other results for sharp large deviation in different settings can be found in the reference of [40]. A very recent sharp large deviation principle result is obtained for weakly dependent sum [27]. The proof is similar to ours using an Edgeworth expansion to further estimate the tail probability.

As an attempt to go from independent to nonindependent, we also look at random projections of random vectors $X^{(n,p)}$ sampled from uniform measure on ℓ_p^n sphere/ball in $\mathbb{R}^n$. We rescale the projection and consider

$$\frac{1}{n}\sum_{i=1}^{n}\left(n^{1/p}X_i^{(n,p)}\right)\left(n^{1/2}\Theta_i^n\right).$$

The sum above is neither independent nor identically distributed conditioning on the direction θ^n. Also the direction Θ^n is not i.i.d and is different from the assumption in [17]. Instead of directly attacking the sum, we utilize a representation for the measure on ℓ_p^n spheres and calculate the tail measure under a transformation. In particular, we consider a sharp large deviation estimate in multi-dimension, which require a different technique from the 1-d case. Other than obtaining the sharp estimates, we also prove a CLT for a function of the coordinates, which is closely related to the CLT results in [44].

4.1.2 Notation

Given a twice differentiable function $f:\mathbb{R}^d\to\mathbb{R}$, we use Hess f to denote the $d\times d$ Hessian matrix of f. Also, given a $m\times d$ matrix A, let A^T denote its transpose and when $m=d$, let det$A$ denote its determinant. For $q\in\mathbb{N}$, define the function space $\mathbb{L}_q(\mathbb{R}^d)$ to be

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

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 $x=(x_1,\dots,x_n)\in\mathbb{R}^n$,

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

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

$$\mathbb{S}_p^{n-1}:=\{x\in\mathbb{R}^n:\|x\|_{n,p}=1\}\quad\text{and}\quad B_p^n:=\{x\in\mathbb{R}^n:\|x\|_{n,p}\le 1\}.\tag{4.3}$$

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

$$\mu_{n,p}(A) := \frac{\mathrm{vol}([0,1]A)}{\mathrm{vol}(B_p^n)}, \tag{4.4}$$

where $[0,1]A := \{xa \in \mathbb{R}^n : x \in [0,1], a \in A\}$, and 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 σ_n to denote $\mu_{n,2}$.

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

$$\mathcal{W}_p(\nu, \nu') := \inf_{\pi \in \Pi(\nu, \nu')} \int_{\mathbb{R}^2} \left| x - y \right|^p \pi(dx, dy), \quad \nu, \nu' \in \mathcal{P}_p(\mathbb{R}). \tag{4.5}$$

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

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

Definition 4.1.1. Given $m \in \mathbb{N}$, we say that a function $f : \mathbb{R} \to \mathbb{R}$ has polynomial growth of degree m if there exist $T \in \mathbb{R}$ and $C \in (0, \infty)$ such that

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

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

We now recall the definition of the p-Wasserstein distance on probability measures.

Lemma 4.1.2 (Definition 6.8 and Theorem 6.9 of [79])**.** *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. $\mathcal{W}_p(\nu^n, \nu) \to 0$.
2. *For any continuous $\phi : \mathbb{R} \to \mathbb{R}$ that has polynomial growth of degree p*

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

4.2 Tail probabilities for projections of product measures

4.2.1 Analytical results

For each $n \in \mathbb{N}$, let $X^n = (X_1^n, \ldots, X_n^n) \in \mathbb{R}^n$ be a sequence of independent and identically distributed random variables with a common probability distribution $\eta \in \mathcal{P}(\mathbb{R})$ and independent of Θ. For $n \in \mathbb{N}$, let $W^{n,\infty}$ be the normalized scalar projection of X^n along the direction Θ^n, defined as

$$W^{n,\infty} := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i^n \Theta_i^n. \tag{4.6}$$

We now introduce notation to state a quenched large deviation principle (LDP) for the sequence $\{W^{n,\infty}\}$ that was obtained in [33, Theorem 2.4]. Define

$$\Lambda(t) := \log \mathbb{E}\left[e^{tX_1}\right], \quad t \in \mathbb{R}, \tag{4.7}$$

$$\Psi(t) := \int_{\mathbb{R}} \Lambda(tu)\mu_2(du), \quad t \in \mathbb{R}, \tag{4.8}$$

where μ_2 is the standard Gaussian distribution. Let $\mathbb{I}_\infty$ be the Legendre transform of Ψ, defined as follows:

$$\mathbb{I}_\infty(t) := \sup_{s\in\mathbb{R}}\{st - \Psi(s)\}, \quad t \in \mathbb{R}. \tag{4.9}$$

Theorem 4.2.1. [33, Theorem 2.4] *Assume that Λ satisfies the following averaged fourth-moment condition*

$$\forall t \in \mathbb{R}, \quad \int_{\mathbb{R}} |\Lambda(tu)|^4 \, \mu_2(du) < \infty.$$

Then for σ-a.e. θ, under $\mathbb{P}_\theta$ the sequence $\{W^{n,\infty}\}$ satisfies an LDP with the convex, good rate function $\mathbb{I}_\infty$ defined in (4.9).

In the following, we obtain a sharp large deviation estimate of the probability of $W^{n,\infty}$ being larger than some positive number a, which captures not only the exponential rate, but also the pre-factor in the spirit of Bahadur and Ranga-Rao [10].

We say that a function $f : \mathbb{R} \mapsto \mathbb{R}$ has polynomial growth (of degree m) if there exists $m \in \mathbb{N}$ and $T \in \mathbb{R}$ such that

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

We impose the following assumptions:

Assumption 4.2.2. The log moment generating function Λ defined in (4.7) satisfies the following properties.

1. Λ'''' has polynomial growth.

2. The characteristic function of X_1, $\phi(t) := \mathbb{E}\left[e^{itX_1}\right]$ $t \in \mathbb{R}$ satisfies Cramér's condition:

$$\limsup_{|t|\to\infty} \left|\phi(t)\right| < 1.$$

We now present our main theorem.

Theorem 4.2.3. *Suppose Assumptions 4.2.2 holds. Fix $a > \mathbb{E}[X_1]$ such that $\mathbb{I}_\infty(a) < \infty$. Then there exist τ_a, $\{\sigma_n\}_{n\in\mathbb{N}} \in \mathbb{R}_+$, mappings $\{R_a^n\}_{n\in\mathbb{N}}$, $\{c_a^n\}_{n\in\mathbb{N}} : S^{n-1} \mapsto \mathbb{R}$ such that for σ-a.e. θ,*

$$\mathbb{P}_\theta(W^{n,\infty} > a) = \frac{C_a^n(\theta^n)}{\sigma_n \tau_a \sqrt{2\pi n}} e^{-n\mathbb{I}_\infty(a)}(1 + o(1)), \tag{4.10}$$

where

$$C_a^n(\theta^n) := e^{\sqrt{n}R_a^n(\theta^n) + c_a^n(\theta^n)^2}.$$

The proof of Theorem 4.2.3 is given in Section 4.5. We first do a change of measure where under the change of measure, the quenched sum in (4.6) remains a sum of independent random variables. We next show a refined CLT under the change of measure using the Edgeworth expansion. The main differences of our proof from [10] are that we do not have i.i.d. random variables and hence the tilted measure here is slightly different from the usual one in Cramér's theorem and that since (4.6) is a projection onto random directions, we need a strong law of large numbers and a central limit theorem for the direction sequences.

The sharp LDP of the projection onto the $(1, 1, \ldots, 1)$ direction is given in (4.1). To compare with the random projection in Theorem 4.2.3, we see that we have an additional factor $C_a^n(\theta^n)$. This term shows up because the sum we consider in (4.6) is not identically distributed. The exact definition of R_a^n and c_a^n can be found in (4.61) and (4.62). A further analysis of R_a^n and c_a^n is given in Proposition 4.5.6. We see that both terms are zero if the direction sequence were to be $(1, 1, \dots, 1)$. The extra correction terms have a Gaussian fluctuation of θ^n dependency. A similar result of this kind can also be seen in [17], where they considered i.i.d. weights as opposed to our non-independent direction sequences. Despite not having i.i.d. weights, the direction sequence $\{\Theta_j^n\}_{j=1}^n$ is "*asymptotic independent*". Thus, we could possibly use a similar method to [17] to obtain our results. However, in order to extend the results to ℓ_p^n balls instead of only product measures a different approach is taken here and the proof in Section 4.5 could provide intuition to the more complicated proof of ℓ_p^n balls given in [58].

4.2.2 Numerical results

The estimation in Theorem 4.2.3 gives an approximation to the tail probability. However, in the proof we use a tilted measure in (4.55), and it turns out that the measure $\mathbb{Q}^n$ is the correct measure if we want to use importance sampling to approximate the tail probability.

To be more precise, for $a > 0$, let τ_a be the constant in Theorem 4.2.3. Let f be the density of X_1. Given a direction $\theta^n \in S^{n-1}$, consider the following densities of random variables $\tilde{X}_j^n$,

$$f_j^n(x) := e^{\tau_a \sqrt{n}\theta_i^n x - \Lambda(\tau_a \sqrt{n}\theta_i^n)} f(x), \quad \text{for} \quad j = 1, \dots, n.$$

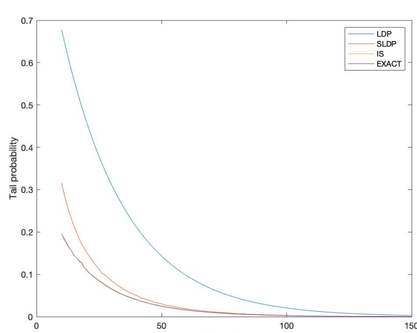
We then calculate the expectation

$$\mathbb{E}\left[1_{\{\frac{1}{\sqrt{n}}\sum_{j=1}^n \tilde{X}_j^n\theta_j^n > a\}} \exp\left(-\tau_a\sqrt{n}\sum_{j=1}^n \tilde{X}_j^n\theta_j^n + \sum_{j=1}^n \Lambda(\tau_a\sqrt{n}\theta_i^n)\right)\right] = \mathbb{P}_\theta(W^{n,\infty} > a).$$

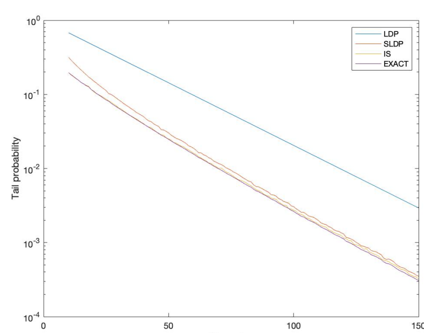
Therefore, we can calculate the empirical mean and get an approximation of the tail probability.

In the following, we use importance sampling with the proposed measure to approximate the tail probability of random projected sum of i.i.d. uniform random variables $U[-1,1]$. We first randomly sample a direction sequence. Then the importance sampling is done with 10^4 iterations for each $n = 10, 11, \ldots, 200$.

Apart from importance sampling, in orange line, we calculate the approximation given in our Theorem 4.2.3, in red line, the large deviation principles without prefactor, in blue line, and also the exact value of the probabilities, in purple line. We do the simulation for small and large values of a. In both cases, the IS and sharp LDP both work very well, and is close to the exact values. For the exact value, we calculate the product of the characteristic functions, and take the inverse transform to get the density. However, the exact calculation fails to work for large a since the probability is too small.



(a) Linear scale.



(b) Log scale.

Figure 4.1: $\mathbb{P}_\theta(W^{n,\infty} > 0.1592)$. We can see that the importance sampling match almost perfectly with the exact value. The sharp large deviation estimate also capture the pre-factor very well.

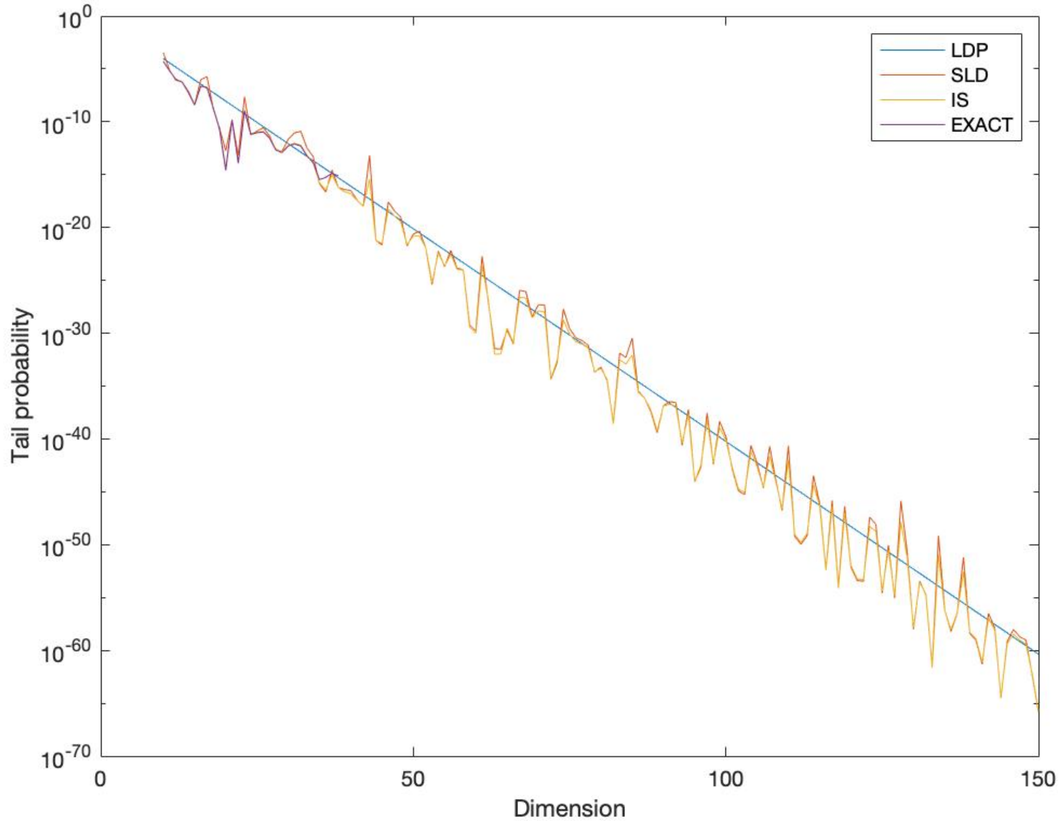


Figure 4.2: $a = 0.6236$. Log scale. For a much bigger a, we still have that the importance sampling and sharp large deviation principle is doing very well. However, the exact calculation fails to work at around $n = 30$.

4.3 Tail probabilities for projections for ℓ_p^n spheres

Fix $p \in (1, \infty)$. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which are defined three independent sequences $\Theta = (\Theta^n)_{n \in \mathbb{N}}$, $X = (X^{(n,p)})_{n \in \mathbb{N}}$ and $\widetilde{\mathcal{X}} = (\widetilde{\mathcal{X}}^{(n,p)})_{n \in \mathbb{N}}$ where Θ 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 each $X^{(n,p)}/\widetilde{\mathcal{X}}^{(n,p)}$ is distributed according to the cone/volume measure $\mu_{n,p}$ on the unit ℓ_p^n, as defined in (4.4). We assume that Θ has distribution σ, where σ 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 σ_n, the unique rotation invariant measure on $\mathbb{S}^{n-1}$. The dependence between the random vectors Θ^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 Θ^n defined as

$$W^{(n,p)} := \frac{n^{1/p}}{n^{1/2}} \sum_{i=1}^{n} X_i^{(n,p)} \Theta_i^n, \tag{4.11}$$

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

$$\widetilde{\mathcal{W}}^{(n,p)} := \frac{n^{1/p}}{n^{1/2}} \sum_{i=1}^{n} \widetilde{\mathcal{X}}_i^{(n,p)} \Theta_i^n. \tag{4.12}$$

First, in Section 4.3.1, we set notation that is used in describing the quenched sharp large deviation estimates. In Section 4.3.2 we recall the quenched LDP for ℓ_p^n spheres of [34] 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 ℓ_p^n spheres. The corresponding results for ℓ_p^n balls are give in Section 4.3.3. Finally, in Section 4.3.4 we provide a brief outline of the proof, and present a more detailed comparison of our results with classical Bahadur-Ranga Rao bounds.

4.3.1 Preliminary notation

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

$$f_p(y) := \frac{1}{2p^{1/p}\Gamma(1 + \frac{1}{p})} e^{-|y|^p/p}, \quad y \in \mathbb{R}, \tag{4.13}$$

where Γ is the Gamma function. For $t_1, t_2 \in \mathbb{R}$ and $t_2 < 1/p$, define

$$\Lambda_p(t_1, t_2) := \log\left(\int_{\mathbb{R}} e^{t_1 y + t_2 |y|^p} \gamma_p(dy)\right), \tag{4.14}$$

and

$$\Psi_p(t_1, t_2) := \int_{\mathbb{R}} \Lambda_p(ut_1, t_2)\gamma_2(du). \tag{4.15}$$

Also, let Ψ_p^* be the Legendre transform of Ψ_p:

$$\Psi_p^*(t_1, t_2) := \sup_{s_1, s_2 \in \mathbb{R}} \{t_1 s_1 + t_2 s_2 - \Psi_p(s_1, s_2)\}, \quad t_1, t_2 \in \mathbb{R}, \tag{4.16}$$

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

$$\mathbb{J}_p := \{(x_1, x_2) \in \mathbb{R}^2 : \Psi_p^*(x_1, x_2) < \infty\}. \tag{4.17}$$

Since by Lemma 5.8 in [34], Λ_p defined in (4.14) is strictly convex on its effective domain, which we denote by $\mathbb{D}_p$, Ψ_p is also strictly convex on $\mathbb{D}_p$. By Lemma 5.9 in [34], Ψ_p is essentially smooth, lower-semicontinuous and hence closed. Therefore by Theorem 26.5 of [72], $\nabla\Psi_p$ is one-to-one and onto from the domain of Ψ_p to $\mathbb{J}_p$. Thus, for each $(x_1, x_2) \in \mathbb{J}_p$ there exists a unique λ_x such that $\nabla\Psi_p(\lambda_x) = x$, which in turn implies that λ_x uniquely achieves the supremum in (4.16), implying that

$$\nabla\Psi_p(\lambda_x) = x, \tag{4.18}$$

and

$$\Psi_p^*(x) = \langle x, \lambda_x \rangle - \Psi_p(\lambda_x). \tag{4.19}$$

Remark 4.3.1. Since Ψ_p is a strictly convex smooth function, the inverse function theorem and (4.18) imply that the mapping $x \mapsto \lambda_x$ is also smooth.

4.3.2 Results on projections of ℓ_p^n spheres

We first state a quenched LDP for the sequence $(W^{(n,p)})_{n\in\mathbb{N}}$. It follows from Theorem 2.5, Lemma 3.1 and Lemma 3.4 of [34] that for σ-a.e. θ, under $\mathbb{P}_\theta$, the sequence $(W^{(n,p)})_{n\in\mathbb{N}}$ satisfies an LDP with (speed n and) a quasiconvex good rate function

$$\mathbb{I}_p(t) = \inf_{\tau_1\in\mathbb{R},\tau_2>0:\tau_1\tau_2^{-1/p}=t} \Psi_p^*(\tau_1,\tau_2), \tag{4.20}$$

where recall that a quasiconvex function is a function whose level sets are convex. Note that the rate function is insensitive to the projection directions, in the sense that it is the same for σ-a.e. θ.

We show in the following lemma that the infimum in (4.20) is attained uniquely at $(t,1)$, yielding a simpler form for the rate function that shows that it is strictly convex and has a unique minimizer. The latter is a crucial property for both obtaining sharp large deviation estimates and developing importance sampling algorithms.

Lemma 4.3.2. *For $p \in (1,\infty)$ and $a > 0$ such that $\Psi_p^*(a,1) < \infty$,*

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

The proof of Lemma 4.3.2 is relegated to Section 4.9; when combined with Theorem 2.5, Lemma 3.1 and Lemma 3.4 of [34], it yields the following simpler form of the quenched LDP.

Theorem 4.3.3. *Fix $p \in (1,\infty)$. For σ-a.e. θ, under $\mathbb{P}_\theta$, the sequence $(W^{(n,p)})_{n\in\mathbb{N}}$ satisfies an*

LDP with the strictly convex, symmetric, good rate function $\mathbb{I}_p$ given by

$$\mathbb{I}_p(a) := \Psi_p^*(a,1) = \sup_{s_1,s_2\in\mathbb{R}} \left\{ as_1 + s_2 - \Psi_p(s_1,s_2) \right\}. \tag{4.21}$$

We now introduce notation to state the sharp large deviation estimate for $W^{(n,p)}$. Recall the definitions of Ψ_p, Ψ_p^*, $\mathbb{J}_p$ and λ_x from Section 4.3.1 and for $x \in \mathbb{J}_p$ define $\mathcal{H}_x = \mathcal{H}_{p,x}$ by

$$\mathcal{H}_{p,x} := (\operatorname{Hess}\Psi_p)(\lambda_x), \tag{4.22}$$

where we suppress the dependence on p from λ_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 (4.21), that is,

$$\Psi_p^*(a,1) = a\lambda_{a,1} + \lambda_{a,2} - \Psi_p(\lambda_{a,1}, \lambda_{a,2}), \tag{4.23}$$

and

$$\mathcal{H}_a := \left(\operatorname{Hess}\Psi_p\right)(\lambda_a). \tag{4.24}$$

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

$$\xi_a^2 := \langle \mathcal{H}_a\lambda_a, \lambda_a\rangle \det \mathcal{H}_a, \tag{4.25}$$

$$\kappa_a^2 := 1 - \frac{(\lambda_{a,1}^2 + \lambda_{a,2}^2)^{3/2} p(p-1)a}{\left|\lambda_{a,2}^2(\mathcal{H}_a^{-1})_{11} - 2\lambda_{a,1}\lambda_{a,2}(\mathcal{H}_a^{-1})_{12} + \lambda_{a,1}^2(\mathcal{H}_a^{-1})_{22}\right| (a^2+p^2)^{3/2}}. \tag{4.26}$$

Remark 4.3.4. Although it is not obvious that the right-hand side of (4.26) is positive, this will be apparent from the proof of Theorem 4.3.5.

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

$$\begin{aligned} \ell_a(x) &:= \Lambda_p(x\lambda_{a,1}, \lambda_{a,2}), \\ \ell_{a,1}(x) &:= x\partial_1\Lambda_p(x\lambda_{a,1}, \lambda_{a,2}), \\ \ell_{a,2}(x) &:= \partial_2\Lambda_p(x\lambda_{a,1}, \lambda_{a,2}). \end{aligned} \tag{4.27}$$

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

We are now ready to state the sharp large deviation estimate for scaled projections of ℓ_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 4.3.5. *Fix $p \in (1, \infty)$ and $a > 0$ such that $\mathbb{I}_p(a) < \infty$. Then the following statements hold with constants $\xi_a = \xi_{p,a}$ and $\kappa_a = \kappa_{p,a}$ defined as in (4.25) and (4.26), respectively:*

(i) For $n \in \mathbb{N}$, there exist mappings $R_a^n = R_{p,a}^n : \mathbb{S}^{n-1} \to \mathbb{R}$ and $c_a^n = c_{p,a}^n : \mathbb{S}^{n-1} \to \mathbb{R}^2$ such that for σ-a.e. θ,

$$\mathbb{P}_\theta\left(W^{(n,p)} > a\right) = \frac{C_a^n(\theta^n)}{\kappa_a \xi_a \sqrt{2\pi n}} e^{-n\mathbb{I}_p(a) + \sqrt{n}R_a^n(\theta^n)}(1 + o(1)), \tag{4.28}$$

where

$$C_a^n(\theta^n) := \exp\left(\left\|\mathcal{H}_a^{-1/2} c_a^n(\theta^n)\right\|^2\right). \tag{4.29}$$

(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 $n \in \mathbb{N}$,

$$(R_a^n(\Theta^n), c_a^n(\Theta^n)) \overset{(d)}{=} \left(r_n + \frac{1}{\sqrt{n}} s_n + o\left(\frac{1}{\sqrt{n}}\right), (t_{n,1} + o(1), t_{n,2} + o(1))\right), \tag{4.30}$$

and as $n \to \infty$,

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

where

$$\begin{aligned}(\mathfrak{R}, \mathfrak{S}, \mathfrak{T}_1, \mathfrak{T}_2) &:= \\ &\left(\widetilde{\mathfrak{A}} - \frac{1}{2}\mathbb{E}[\ell_a'(Z)Z]\widetilde{\mathfrak{D}}, \frac{1}{8}\mathbb{E}[\ell_a''(Z)Z^2]\widetilde{\mathfrak{D}}^2, \widetilde{\mathfrak{E}} - \frac{1}{2}\mathbb{E}\left[\ell_{a,1}'(Z)Z\right]\widetilde{\mathfrak{D}}, \widetilde{\mathfrak{G}} - \frac{1}{2}\mathbb{E}\left[\ell_{a,2}'(Z)Z\right]\widetilde{\mathfrak{D}}\right),\end{aligned}$$

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

$$\begin{pmatrix} \mathrm{Cov}(\ell_a(Z), \ell_a(Z)) & \mathrm{Cov}(\ell_a(Z), Z^2) & \mathrm{Cov}(\ell_a(Z), \ell_{a,1}(Z)) & \mathrm{Cov}(\ell_a(Z), \ell_{a,2}(Z)) \\ \mathrm{Cov}(Z^2, \ell_a(Z)) & \mathrm{Cov}(Z^2, Z^2) & \mathrm{Cov}(Z^2, \ell_{a,1}(Z)) & \mathrm{Cov}(Z^2, \ell_{a,2}(Z)) \\ \mathrm{Cov}(\ell_{a,1}(Z), f_a(Z)) & \mathrm{Cov}(\ell_{a,1}(Z), Z^2) & \mathrm{Cov}(\ell_{a,1}(Z), \ell_{a,1}(Z)) & \mathrm{Cov}(\ell_{a,1}(Z), \ell_{a,2}(Z)) \\ \mathrm{Cov}(\ell_{a,2}(Z), \ell_a(Z)) & \mathrm{Cov}(\ell_{a,2}(Z), Z^2) & \mathrm{Cov}(\ell_{a,2}(Z), \ell_{a,1}(Z)) & \mathrm{Cov}(\ell_{a,2}(Z), \ell_{a,2}(Z)) \end{pmatrix}. \tag{4.31}$$

An outline of the proof of Theorem 4.3.5 is given in Section 4.3.4, with full details provided in Sections 4.6.5 and 4.6.6; the precise definitions of the functions c_a^n and R_a^n are given in (4.83) and (4.84), respectively.

Remark 4.3.6. We will refer to the term $\frac{C_a^n(\theta^n)}{\kappa_a \xi_a \sqrt{2\pi n}} e^{\sqrt{n} R_a^n(\theta^n)}$ in (4.28) as the "prefactor" since it is of lower order than $e^{-n\mathbb{I}_p(a)}$, as follows from the statement of the LDP. In addition, it follows from (4.29)-(4.30) that (in distribution) $R_a^n(\theta^n)$ and $C_a^n(\theta^n)$ are both of order 1; see also Lemma 4.6.8 for further estimates. Further insight into the form of the prefactor can be found in Section 4.3.4; see Remark 4.3.14.

As mentioned above, the leading order term in (4.28) that depends on θ is $\sqrt{n} R_{p,a}^n(\theta^n)$.

The following proposition describes the additional geometric information contained in this term beyond what is available in the rate function $\mathbb{I}_p$ which characterizes the asymptotic exponential decay rate of the projection under $\mathbb{P}_\theta$ for σ-a.e. θ.

Proposition 4.3.7. *Fix $p \in (1, \infty)$, $a > 0$ such that $\mathbb{I}_p(a) < \infty$ and let $R^n_{p,a}$ be the mapping in Theorem 4.3.5. Then*

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

where e_j is defined to be the basis vector in $\mathbb{R}^n$.

As a corollary, combining the two parts of Theorem 4.3.5, we obtain an alternative expression for the distribution of the conditioned tail probability:

Corollary 4.3.8. *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 4.3.5 (ii), and that of $\mathcal{H}_a$ from (4.24). Then*

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

where

$$M_n := \exp\left(s_n + \left\|\mathcal{H}_a^{-1/2} t_n\right\|^2\right). \tag{4.32}$$

Moreover, as $n \to \infty$,

$$(M_n, r_n) \Rightarrow \left(\exp\left(\mathfrak{S} + \left\|\mathcal{H}_a^{-1/2}\mathfrak{T}\right\|^2\right), \mathfrak{R}\right), \tag{4.33}$$

where $(\mathfrak{R}, \mathfrak{S}, \mathfrak{T}_1, \mathfrak{T}_2)$ *is defined as in Theorem 4.3.5(ii).*

Proof. By (4.28), (4.29) and (4.30), the tail probability can be written as

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

since $\exp(o(1)) = 1 + o(1)$. Also, from the relation (4.32), 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 4.3.5(ii) to obtain the joint convergence stated in (4.33). □

Remark 4.3.9. In Theorem 4.3.5 and Corollary 4.3.8 we only consider values $p \in (1, \infty)$ because when $p \in (0, 1)$, when the ℓ_p^n balls are no longer convex, the existence of even an LDP has not been established, and when $p = 1$ the quenched LDP has a different form that is not universal in the sense that it only exists when the sequence of projection directions satisfies a certain additional condition, and in that case, the speed of the LDP is $n/\sqrt{\log n}$ rather than n, and the associated rate function depends on a certain asymptotic function of the projection directions (see [34, Theorem 2.6]).

4.3.3 Results on projections of ℓ_p^n balls

Next, we state the corresponding sharp large deviation results for balls. For $p \in (1, \infty)$ and $a > 0$, recall that $\lambda_{a,1}$ is the first coordinate of the maximizer λ_a in the expression for $\Psi_p^*(a)$ in (4.19), and $\mathcal{H}_a$ is as defined in (4.24), and define the positive constant $\gamma_a = \gamma_{p,a}$

via the relation

$$\gamma_a^2 := \lambda_{a,1}^2(1 + a\lambda_{a,1})^2 \det \mathcal{H}_a \left| -\frac{a(p-1)}{p^2}\lambda_{a,1} + \frac{2a}{p}(\mathcal{H}_a)^{-1}_{12} + (\mathcal{H}_a)^{-1}_{22} + \frac{a^2}{p^2}(\mathcal{H}_a)^{-1}_{11} \right|. \tag{4.34}$$

Theorem 4.3.10. *Fix $p \in (1, \infty)$ and $a > 0$ such that $\mathbb{I}_p(a) < \infty$. Then for $n \in \mathbb{N}$,*

$$\mathbb{P}_\theta\left(\widetilde{\mathcal{W}}^{(n,p)} > a\right) = \frac{C_a^n(\theta^n)}{\gamma_a \sqrt{2\pi n}} e^{-n\mathbb{I}_p(a) + \sqrt{n}R_a^n(\theta^n)}(1 + o(1)), \tag{4.35}$$

where C_a^n is defined in (4.29), $\gamma_a = \gamma_{p,a}$ is the constant defined in (4.34), and R_a^n, c_a^n are the constants given in Theorem 4.3.5.

Remark 4.3.11. (i) Note that the tail probability in (4.35) has a geometric interpretation, in that it characterizes the volumes of spherical caps (at level a) of ℓ_p^n balls along the direction θ^n.

(ii) While it follows from the results of [34] (see Theorem 4.3.3) that the LDPs for random projections of ℓ_p^n balls and spheres coincide (in the sense that they have the same speed and rate function), we see from (4.28) and (4.35) that although the two prefactors have a similar form, their actual values differ significantly as borne out by the numerical computations in [57]). Thus, the sharp large deviation estimates obtained here are sufficiently refined to distinguish these two objects.

Similar to Corollary 4.3.8, we have the following immediate corollary for balls:

Corollary 4.3.12. *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 4.3.8. Then*

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

where (4.32) and (4.33) hold.

Remark 4.3.13. Proposition 4.3.7 shows how the sharp large deviation estimates reflect the difference in the geometry of ℓ_p^n spheres for $p \in (1,2)$ and $p > 2$. This motivates obtaining sharp large deviation estimates for projections of more general high-dimensional objects to uncover new geometric information about these objects. Theorem 4.3.5 and Theorem 4.3.10 show that the sharp estimate for the ℓ_p^n ball differs from that for the sphere, thus further emphasizing the usefulness of sharp large deviation estimates over LDPs.

4.3.4 A reformulation and the proof outline

Fix $p \in (1,\infty)$. As mentioned in the introduction, one of the reasons the estimate (4.28) is challenging to establish is that $W^{(n,p)}$ is a randomly weighted sum of random variables that are not independent, where 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 two-dimensional random vector $\bar{S}^{(n,p)}$ 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-th Gaussian random variables $(Y_i^{(p)})_{i\in\mathbb{N}}$, independent of Θ, 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 (4.13). Then, it follows from [74, Lemma 1] (see also a statement of this property at the bottom of p. 548 in [16]) that

$$X^{(n,p)} \overset{(d)}{=} \frac{Y^{(n,p)}}{\left\|Y^{(n,p)}\right\|_{n,p}}, \qquad n \in \mathbb{N}, \tag{4.36}$$

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

$$\bar{S}^{(n,p)} := \frac{1}{n} \sum_{j=1}^{n} \left(\sqrt{n} \Theta_j^n Y_j^{(p)}, \left|Y_j^{(p)}\right|^p \right). \tag{4.37}$$

In view of (4.11) and the independence of $X^{(n,p)}$, (4.36), and Θ, for $a > 0$ and $\theta \in \mathbb{S}$, we may rewrite the tail probability on the left-hand side of (4.28) as

$$\begin{aligned} \mathbb{P}_\theta\left(W^{(n,p)} > a\right) &= \mathbb{P}\left(\frac{n^{1/p}}{n} \sum_{j=1}^{n} \frac{\sqrt{n}\theta_j^n Y_j^{(p)}}{\left\|Y^{(n,p)}\right\|_{n,p}} > a \right) \\ &= \mathbb{P}\left(\frac{1}{n} \sum_{j=1}^{n} \sqrt{n}\theta_j^n Y_j^{(p)} > a \left(\frac{1}{n} \sum_{j=1}^{n} \left|Y_j^{(p)}\right|^p \right)^{1/p} \right) \\ &= \mathbb{P}_\theta\left(\bar{S}^{(n,p)} \in \bar{D}_{p,a}\right), \end{aligned} \tag{4.38}$$

where

$$\bar{D}_{p,a} := \left\{ (x,y) \in \mathbb{R}^2 : y > 0, x > a y^{1/p} \right\}. \tag{4.39}$$

On the other hand, again from [74, Lemma 1], we also have an equivalent representation for $\widetilde{\mathcal{X}}^{(n,p)}$:

$$\widetilde{\mathcal{X}}^{(n,p)} \overset{(d)}{=} {}^{1/n} \frac{Y^{(n,p)}}{\left\|Y^{(n,p)}\right\|_{n,p}}, \qquad n \in \mathbb{N}, \tag{4.40}$$

where 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

$$\bar{\mathcal{S}}^{(n,p)} := \left(\frac{1}{n} \sum_{j=1}^{n} \sqrt{n} \Theta_j^n Y_j^{(p)}, \frac{1}{n} \sum_{j=1}^{n} \left| Y_j^{(p)} \right|^p, {}^{1/n} \right) = \left(\bar{S}^{(n,p)}, {}^{1/n} \right). \tag{4.41}$$

From the equivalent representation (4.40), for $a > 0$ and $\theta \in \mathbb{S}$, we may rewrite the tail probability of $\widetilde{\mathcal{W}}^{(n,p)}$ as

$$\begin{aligned}
\mathbb{P}_\theta \left(\widetilde{\mathcal{W}}^{(n,p)} > a \right) &= \mathbb{P} \left(\frac{n^{1/p}}{n} \sum_{j=1}^{n} {}^{1/n} \frac{\sqrt{n} \theta_j^n Y_j^{(p)}}{\left\| Y^{(n,p)} \right\|_{n,p}} > a \right) \\
&= \mathbb{P} \left({}^{1/n} \frac{1}{n} \sum_{j=1}^{n} \sqrt{n} \theta_j^n Y_j^{(p)} > a \left(\frac{1}{n} \sum_{j=1}^{n} \left| Y_j^{(p)} \right|^p \right)^{1/p} \right) \\
&= \mathbb{P}_\theta \left(\bar{\mathcal{S}}^{(n,p)} \in \bar{\mathcal{D}}_{p,a} \right),
\end{aligned} \tag{4.42}$$

where

$$\bar{\mathcal{D}}_{p,a} := \left\{ (x_1, x_2, y) \in \mathbb{R}^3 : 1 \geq y \geq 0, x_2 > 0, x_1 y > a x_2^{1/p} \right\}. \tag{4.43}$$

While several results on sharp large deviations in multiple dimensions have been obtained (see, e.g., [6, 41] as well as [11] for a comprehensive list of references), none of these cover the cases of interest in (4.38). In particular, the work [6] 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, and further, the results of [41] also do not apply since Assumption (A.2) of [41] therein is not satisfied here, in particular since we have an additional $\sqrt{n}$ factor in the exponent of (4.28) compared with [41, Equation (3)]. Instead, our proof proceeds by first exploiting quantitative asymptotic independence results of the weights $(\Theta_i^n)_{j=1,\ldots,n}$ obtained in Section 4.4, and combining them with new asymptotic estimates for certain Laplace-type integrals stated in Section 4.6.

Remark 4.3.14. Comparing the estimate in (4.28) with the sharp large deviation estimate for the projection of an i.i.d. sum onto the vector $\mathfrak{I}^n = (1, 1, \ldots, 1)/\sqrt{n}$ given in (4.1), we see that ξ_a here plays a role similar to $\sigma_a\tau_a$ in (4.1). On the other hand, the additional constant κ_a in (4.28) arises due to the geometry of the domain $\bar{D}_{p,a}$ defined in (4.39) and the fact that we obtain this estimate by reformulating it in terms of a a two-dimensional sharp large deviations problem. From a technical point of view, the additional θ^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 θ^n. From their exact definitions given in (4.84) and (4.83), 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 $\theta^n = \mathfrak{I}^n = (1, 1, \ldots, 1)/\sqrt{n}$.

4.3.5 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 (4.4), 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 4.3.5. For $a > 0$, fix $p \in (1, \infty)$ and recall the constant λ_a defined in (4.23). Also, recall the definition of the density f_p in (4.13). Given $n \in \mathbb{N}$ let $\widetilde{Y}^{(n,p)} := (\widetilde{Y}_1^{(n,p)}, \ldots, \widetilde{Y}_n^{(n,p)})$, where $\widetilde{Y}_j^{(n,p)}$, $j = 1, \ldots, n$, are random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$ that are independent under $\mathbb{P}_\theta$ for each $\theta \in \mathbb{S}$

such that $\widetilde{Y}_j^{(n,p)}$ has density

$$\widetilde{f}_{p,j}^{n}(y) := \exp\left(\left\langle \lambda_a, \left(\sqrt{n}\theta_j^n y, |y|^p\right)\right\rangle - \Lambda_p\left(\sqrt{n}\theta_i^n \lambda_{a,1}, \lambda_{a,2}\right)\right) f_p(y), \quad y \in \mathbb{R}, \tag{4.44}$$

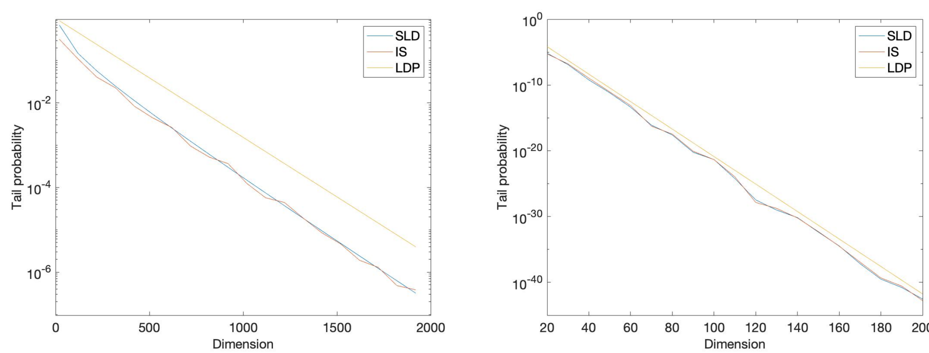
where we suppress the explicit dependence of $\widetilde{f}_{p,j}^{n}$ on θ^n. Also define

$$\widetilde{W}^{(n,p)} := \frac{n^{1/p}}{n^{1/2}} \sum_{j=1}^{n} \frac{\widetilde{Y}_j^{(n,p)} \Theta_j^n}{\left\|\widetilde{Y}^{(n,p)}\right\|_{n,p}}. \tag{4.45}$$

In view of (4.44) and (4.45), it then follows that

$$\mathbb{P}_\theta(W^{(n,p)} > a) = \mathbb{E}_\theta\left[1_{\{\widetilde{W}^{(n,p)}>a\}} \prod_{j=1}^{n} \exp\left(-\left\langle \lambda_a, \left(\sqrt{n}\theta_j^n \widetilde{Y}_j^{(n,p)}, \left|\widetilde{Y}_j^{(n,p)}\right|^p\right)\right\rangle + \Lambda_p\left(\sqrt{n}\theta_i^n \lambda_{a,1}, \lambda_{a,2}\right)\right)\right]. \tag{4.46}$$

The IS algorithm estimates the tail probability on the left-hand side of (4.46) by first sampling a direction θ^n according to σ_n and sampling from i.i.d. copies of the vector $\widetilde{Y}^{(n,p)} := (\widetilde{Y}_1^p, \ldots, \widetilde{Y}_n^p)$, independently of the θ^n sample, to approximate the expectation on the right-hand side of (4.46) by a standard Monte Carlo estimate.



(a) $a = 0.1$. (b) $a = 0.7$.

Figure 4.3: 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 4.3.6) 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 4.1 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 4.3 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 4.3(B) and Table 4.1, 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 4.3(A) and Table 4.2, the SLD estimate appears to achieve the same accuracy only for much larger n, which reflects the dependence of the $o(1)$ term in (4.28) on a.

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

Table 4.1: 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	6.9193×10^{-1}	3.2004×10^{-1}	116.20%	$[2.5636 \times 10^{-1}, 3.8372 \times 10^{-1}]$
420	1.1317×10^{-2}	8.3597×10^{-3}	35.38%	$[5.6085 \times 10^{-3}, 1.1110 \times 10^{-2}]$
820	6.0651×10^{-4}	5.2198×10^{-4}	16.19%	$[3.2412 \times 10^{-4}, 7.1985 \times 10^{-4}]$
1220	3.7235×10^{-5}	4.4306×10^{-5}	15.96%	$[2.8312 \times 10^{-5}, 6.0299 \times 10^{-5}]$

Table 4.2: Estimates of $P_\theta(W^{(n,p)} > 0.1)$ for $p = 3$. The sample size for IS is 100.

Finally, we also ran simulations for different realizations θ of the direction sequence Θ. We see from Figure 4.4 that different direction sequences result in fluctuations around the basic sharp large deviation estimate, which ignores the θ^n-dependent terms in the prefactor in (4.28), i.e., $e^{-n\mathbb{I}_p(a)}/\kappa_a \xi_a \sqrt{2\pi n}$. As shown in Theorem 4.3.5(ii), these fluctuations converge in distribution to functionals of a multidimensional Gaussian vector with an explicit covariance matrix.

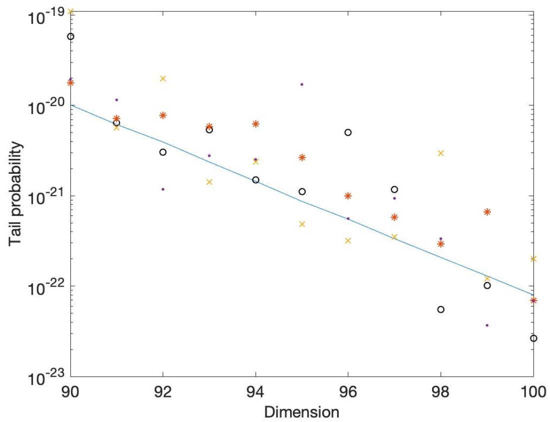


Figure 4.4: Log scale plot of estimates of $P_\theta(W^{(n,p)} > 0.7)$ vs. dimension for $p = 3$. Solid line is $e^{-n\mathbb{I}_p(a)}/\kappa_a\xi_a\sqrt{2\pi n}$. Scatter points are SLD estimates for different direction sequences.

4.4 Asymptotic Independence Results for the Weights

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

$$L^n_\theta := \frac{1}{n}\sum_{i=1}^{n}\delta_{\sqrt{n}\theta^n_i}. \tag{4.47}$$

We first recall a strong law of large numbers for $(L^n_\theta)_{n\in\mathbb{N}}$ that was established in [33, Lemma 5.11]. Recall that γ_2 denotes the standard normal distribution.

Lemma 4.4.1 (Lemma 5.11 of [33]). *For $p \in (1,\infty)$, for σ-a.e. $\theta \in \mathbb{S}$,*

$$\mathcal{W}_p\left(L^n_\theta, \gamma_2\right) \to 0, \quad \textit{as } n \to \infty.$$

Next, we establish a central limit theorem refinement of Lemma 4.4.1. Given an i.i.d. array $(Z^n = (Z^n_j, j = 1\dots,n))_{n\in\mathbb{N}}$ of standard normal random variables, for any twice continuously differentiable function ϕ, define

$$\hat{s}_n(\phi) := \sum_{j=1}^{n} \frac{\phi''(Z^n_j)}{2}\left(\frac{\sqrt{n}Z^n_j}{\|Z^n\|_{n,2}} - Z^n_j\right)^2, \tag{4.48}$$

and set

$$\hat{r}_n(\phi) := \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[\phi(Z^n_j) - \int_{\mathbb{R}}\phi(x)\gamma_2(dx) + \phi'(Z^n_j)\left(\frac{\sqrt{n}Z^n_j}{\|Z^n\|_{n,2}} - Z^n_j\right)\right]. \tag{4.49}$$

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 $F : \mathbb{R} \to \mathbb{R}$.

Lemma 4.4.2. *Given a thrice continuously differentiable function $F : \mathbb{R} \to \mathbb{R}$ and two twice continuously differentiable functions $G_1, G_2 : \mathbb{R} \to \mathbb{R}$ such that F''', G_1'' and G_2'' have polynomial growth in the sense of Definition 4.1.1, we have the following expansion,*

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

where $\hat{s}_n$ and $\hat{r}_n$ are as defined in (4.48) and (4.49), and as $n \to \infty$,

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

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

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

where Z *is a standard normal random variable.*

This result is similar in spirit to [43, Theorem 1.1], which establishes a central limit theorem for the sequence of q-norms of $\sqrt{n}\Theta^n$, $n \in \mathbb{N}$. In Lemma 4.4.2, we obtain 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 4.4.2 is deferred to Section 4.10.

4.5 Proof for product measures

4.5.1 Preliminary results

Define the empirical measure of the direction θ^n, L^n_θ, as in (4.47).

We first define the Gärtner-Ellis cumulant functions. For $\theta^n \in S^{n-1}$ and $t \in \mathbb{R}$ define

$$\Psi_{n,\theta^n}(t) := \frac{1}{n} \log \mathbb{E}_\theta \left[e^{nW^{n,\infty} t} \right]. \tag{4.50}$$

By the definition of $W^{n,\infty}$ in (4.6), and the i.i.d. nature of the sequence $\{X_i^n\}$, we have

$$\begin{aligned}\Psi_{n,\theta^n}(t) &= \frac{1}{n}\log\prod_{j=1}^{n}\mathbb{E}_\theta\left[e^{\sqrt{n}X_i^n\theta_i^n t}\right]\\ &= \frac{1}{n}\sum_{i=1}^{n}\Lambda(\sqrt{n}\theta_i^n t)\\ &= \int \Lambda(tu)L_\theta^n(du).\end{aligned}\tag{4.51}$$

Recall from Assumption 4.2.2 that Λ has polynomial growth. Therefore, we may apply Lemma 4.4.1 and the definition of Ψ in (4.15) to obtain the following corollary:

Corollary 4.5.1. *For σ-a.e. θ, for every $t \in \mathbb{R}$,*

$$\Psi_{n,\theta^n}(t) = \int \Lambda(tu)L_\theta^n(du) \to \int \Lambda(tu)\mu_2(du) = \Psi(t). \tag{4.52}$$

By [33, Lemma 3.8], the function Ψ is differentiable on $\mathbb{R}$. The strict convexity of Ψ follows from the strict convexity of Λ. Note that Cramér condition in Assumption 4.2.2 rules out the case for X_1 being a point mass. Therefore we may pick τ_a to be the unique number that attains the supremum in the Legendre transform

$$\mathbb{I}_\infty(a) = a\tau_a - \Psi(\tau_a). \tag{4.53}$$

From the smoothness, the supremum is attained at derivative being 0 and thus we have

$$\Psi'(\tau_a) = a. \tag{4.54}$$

By the strict convexity, Ψ' is strictly increasing. Since $\Psi'(0) = \mathbb{E}[X_1]$ and $a > \mathbb{E}[X_1]$, τ_a is strictly positive.

4.5.2 A change of measure

For each $n \in \mathbb{N}$, let $\mathbb{Q}^n$ be a new measure on $\mathbb{R}^n$ that is absolutely continuous with respect to $\mathbb{P}_\theta$, with the likelihood ratio

$$\frac{d\mathbb{Q}^n}{d\mathbb{P}_\theta} := e^{\tau_a n W^{n,\infty} - n\Psi_n(\tau_a)}. \tag{4.55}$$

Note that $\mathbb{Q}^n$ is a probability measure because

$$\mathbb{E}_\theta\left[e^{\tau_a n W^{n,\infty}}\right] = e^{n\Psi_n(\tau_a)}.$$

Now, define

$$U^n := \frac{\sqrt{n}(W^{n,\infty} - a_n)}{\sigma_n}, \tag{4.56}$$

where

$$a_n := \mathbb{E}_{\mathbb{Q}^n}\left[W^{n,\infty}\right] \quad \text{and} \quad \sigma_n := n\mathrm{Var}_{\mathbb{Q}^n}\left[W^{n,\infty}\right] \tag{4.57}$$

so that U^n has zero mean and unit variance under $\mathbb{Q}^n$.

Define

$$H^n(x) := \mathbb{Q}^n(U^n \le x), \tag{4.58}$$

and

$$\chi_n := \mathbb{E}_{\mathbb{Q}^n}\left[|\sigma_n U^n|^3\right]. \tag{4.59}$$

The proof of Theorem 4.2.3 relies crucially on the following refined CLT under the change of measure. Its proof is deferred to Section 4.5.5. Let H be the cumulative distribution function of a standard Gaussian random variable.

Proposition 4.5.2. *Let $W^{n,\infty}$, a_n, σ_n, χ_n and H^n be defined as above. Then for σ-a.e. θ, the limit of $\{\sigma_n\}$ exists,*

$$\sigma_a := \lim_{n\to\infty} \sigma_n > 0.$$

Moreover, for σ-a.e. θ,

$$\sup_{x\in\mathbb{R}} \left| H^n(x) - \widetilde{H}^n(x) \right| = o\left(\frac{1}{\sqrt{n}}\right), \tag{4.60}$$

where $\widetilde{H}^n$ is defined to be

$$\widetilde{H}^n(x) := H(x) - \frac{\chi_n}{6\sigma_n^3\sqrt{n}}(x^2-1)\frac{1}{\sqrt{2\pi}}e^{-x^2/2},$$

and the sequence $\{\chi_n/\sigma_n^3\}_{n\in\mathbb{N}}$ is uniformly bounded:

$$\sup_{n\in\mathbb{N}} \frac{\chi_n}{\sigma_n^3} < \infty.$$

4.5.3 Proof of Theorem 4.2.3

Fix $a > 0$, τ_a as in equation (4.53), and recall the definition of a_n and σ_n^2 in (4.57). For $n \in \mathbb{N}$ define

$$R_a^n(\theta^n) := \sqrt{n}\left(\Psi_n(\tau_a) - \Psi(\tau_a)\right), \tag{4.61}$$

$$c_a^n(\theta^n) := \frac{\sqrt{n}(a-a_n)}{\sigma_n}. \tag{4.62}$$

For conciseness, we will omit the dependency on θ^n in R_a^n and c_a^n in the rest of the proof.

Claim 4.5.3. We first claim that for $\theta \in S^{n-1}$,

$$\mathbb{P}_\theta(W^{n,\infty} > a) = K_n e^{-n\mathbb{I}_\infty(a)}, \tag{4.63}$$

where

$$K_n := e^{\sqrt{n}\tau_a\sigma_n c_a^n + \sqrt{n}R_a^n} \int_{(c_a^n,\infty)} e^{-\tau_a\sigma_n\sqrt{n}x} dH^n(x).$$

Proof of Claim 1. First, to identify the values a_n and σ_n explicitly, we calculate the mean and variance of $W^{n,\infty}$ under $\mathbb{Q}^n$, by differentiating the cumulant generating function to obtain

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}^n}\left[W^{n,\infty}\right] &= \frac{d}{dt}\int e^{tW^{n,\infty}} d\mathbb{Q}^n\bigg|_{t=0} \\
&= e^{-n\Psi_n(\tau_a)}\frac{d}{dt}\int e^{(t+\tau_a n)W^{n,\infty}} d\mathbb{P}_\theta\bigg|_{t=0} \\
&= \Psi_n'(\tau_a).
\end{aligned}$$

Similarly, the variance of $W^{n,\infty}$ satisfies

$$\mathrm{Var}_{\mathbb{Q}^n}(W^{n,\infty}) = \frac{\Psi_n''(\tau_a)}{n}. \tag{4.64}$$

Therefore,

$$a_n = \Psi_n'(\tau_a), \quad \text{and} \quad \sigma_n^2 = \Psi_n''(\tau_a).$$

By Assumption 4.2.2, $u \mapsto u\Lambda'(tu)$ has polynomial growth.Similar to Corollary 4.52, by

Lemma 4.4.1 and equation (4.54), the expectation converges

$$\Psi_n'(\tau_a) \to \Psi'(\tau_a) = a.$$

By (4.55) and (4.56), we have

$$\begin{aligned}
\mathbb{P}_\theta(W^{n,\infty} > a) &= \int_{\{W^{n,\infty}>a\}} d\mathbb{P}_\theta \\
&= \int_{\{W^{n,\infty}>a\}} e^{n\Psi_n(\tau_a)-\tau_a n W^{n,\infty}} d\mathbb{Q}^n \\
&= e^{n(\Psi_n(\tau_a)-\tau_a a_n)} \int_{\{W^{n,\infty}>a\}} e^{-\tau_a\sigma_n\sqrt{n}\frac{\sqrt{n}(W^{n,\infty}-a_n)}{\sigma_n}} d\mathbb{Q}^n \\
&= e^{n(\Psi_n(\tau_a)-\tau_a a_n)} \int_{(c_a^n,\infty)} e^{-\tau_a\sigma_n\sqrt{n}x} dH^n(x) \\
&= e^{n\left(\Psi(\tau_a)-\tau_a a+\tau_a(a-a_n)+R_a^n/\sqrt{n}\right)} \int_{(c_a^n,\infty)} e^{-\tau_a\sigma_n\sqrt{n}x} dH^n(x) \\
&= K_n e^{-n\mathbb{I}_\infty(a)},
\end{aligned}$$

where we use the definitions of c_a^n, R_a^n and (4.53). □

We next apply Proposition 4.5.2 to obtain a further expansion of the integral.

Claim 4.5.4.

$$e^{\sqrt{n}\tau_a\sigma_n c_a^n} \int_{(c_a^n,\infty)} e^{-\tau_a\sigma_n\sqrt{n}x} dH^n(x) = e^{\sqrt{n}\tau_a\sigma_n c_a^n} \int_{(c_a^n,\infty)} e^{-\tau_a\sigma_n\sqrt{n}x} d\widetilde{H}^n(x) + o(n^{-1/2}). \tag{4.65}$$

Proof of Claim 2. Using integration by parts for the integral in (4.63), we see that

$$\begin{aligned}
\int_{(c_a^n,\infty)} e^{-\tau_a\sigma_n\sqrt{n}x} dH^n(x) &= \tau_a\sigma_n\sqrt{n}\int_{c_a^n}^{\infty} e^{-\tau_a\sigma_n\sqrt{n}x}(H^n(x)-H^n(c_a^n))dx \\
&= \tau_a\sigma_n\sqrt{n}\int_{c_a^n}^{\infty} e^{-\tau_a\sigma_n\sqrt{n}x}(\widetilde{H}^n(x)-\widetilde{H}^n(c_a^n))dx
\end{aligned}$$

$$
\begin{aligned}
&+ \tau_a \sigma_n \sqrt{n} \int_{c_a^n}^{\infty} e^{-\tau_a \sigma_n \sqrt{n} x} (H^n(x) - \widetilde{H}^n(x)) dx \\
&+ \tau_a \sigma_n \sqrt{n} \int_{c_a^n}^{\infty} e^{-\tau_a \sigma_n \sqrt{n} x} (\widetilde{H}^n(c_a^n) - H^n(c_a^n)) dx.
\end{aligned}
$$

From (4.60) and using integration by parts again, we can further deduce that

$$
\begin{aligned}
\int_{(c_a^n,\infty)} e^{-\tau_a \sigma_n \sqrt{n} x} dH^n(x) &= \tau_a \sigma_n \sqrt{n} \int_{c_a^n}^{\infty} e^{-\tau_a \sigma_n \sqrt{n} x} (\widetilde{H}^n(x) - \widetilde{H}^n(c_a^n)) dx \\
&\quad + 2o(n^{-1/2}) \tau_a \sigma_n \sqrt{n} \int_{c_a^n}^{\infty} e^{-\tau_a \sigma_n \sqrt{n} x} dx \\
&= \int_{(c_a^n,\infty)} e^{-\tau_a \sigma_n \sqrt{n} x} d\widetilde{H}^n(x) + 2o(n^{-1/2}) e^{-\tau_a \sigma_n \sqrt{n} c_a^n}.
\end{aligned}
$$

The claim follows by multiplying both sides by the factor $e^{\sqrt{n}\tau_a \sigma_n c_a^n}$. □

Let Φ be the characteristic function of a standard Gaussian random variable. In the next claim, we use Fourier transform to estimate the integral in (4.65).

Claim 4.5.5.

$$
e^{\sqrt{n}\tau_a \sigma_n c_a^n} \int_{(c_a^n,\infty)} e^{-\tau_a \sigma_n \sqrt{n} x} d\widetilde{H}^n(x) = \frac{1}{2\pi} \int \frac{\cos(t c_a^n) \tau_a \sigma_n \sqrt{n}}{t^2 + \sigma_n^2 \tau_a^2 n} \Phi(t) dt + O(n^{-1}). \tag{4.66}
$$

Proof of Claim 3. The Fourier transform of $\frac{1}{\sqrt{2\pi}}\Phi$ is Φ, and the Fourier transform of $e^{-\tau_a \sigma_n \sqrt{n} x} 1_{(c_a^n,\infty)}$ is

$$
\begin{aligned}
\int_{-\infty}^{\infty} e^{-\tau_a \sigma_n \sqrt{n} x} 1_{(c_a^n,\infty)} e^{ixt} dx &= \left. \frac{e^{itx - \tau_a \sigma_n \sqrt{n} x}}{it - \tau_a \sigma_n \sqrt{n}} \right|_{x=c_a^n}^{\infty} \\
&= \frac{-1}{(it - \tau_a \sigma_n \sqrt{n})} e^{itc_a^n - \tau_a \sigma_n \sqrt{n} c_a^n},
\end{aligned}
$$

where the last equality above uses the positivity of $\tau_a \sigma_n$. Also the Fourier transform of

$d\widetilde{H}^n$ is

$$\int_{\mathbb{R}} e^{ixt} d\widetilde{H}^n(x) = \Phi(t)\left[1 + \frac{\chi_n}{6\sigma_n^3\sqrt{n}}(-it)^3\right].$$

For a complex number z, denote its complex conjugate as $\bar{z}$. Using the last two displays and Parseval's formula, we have

$$\begin{aligned} e^{\sqrt{n}\tau_a\sigma_n c_a^n}\int_{(c_a^n,\infty)} e^{-\tau_a\sigma_n\sqrt{n}x} d\widetilde{H}^n(x) &= \frac{1}{2\pi}\int_{\mathbb{R}} \overline{\frac{-e^{-itc_a^n}}{it-\tau_a\sigma_n\sqrt{n}}}\Phi(t)\left[1 + \frac{\chi_n}{6\sigma_n^3\sqrt{n}}(-it)^3\right]dt \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} \frac{e^{itc_a^n}}{it+\tau_a\sigma_n\sqrt{n}}\Phi(t)\left[1 + \frac{\chi_n}{6\sigma_n^3\sqrt{n}}(-it)^3\right]dt. \qquad (4.67) \end{aligned}$$

We first bound the second term in the integral on the right-hand side of (4.67) as follows:

$$\begin{aligned} \left|\frac{1}{2\pi}\int_{\mathbb{R}} \frac{e^{itc_a^n}}{it+\tau_a\sigma_n\sqrt{n}}\Phi(t)\frac{\chi_n}{6\sigma_n^3\sqrt{n}}(-it)^3 dt\right| &\le \frac{1}{2\pi}\int_{\mathbb{R}}\left|\frac{e^{itc_a^n}}{it+\tau_a\sigma_n\sqrt{n}}\Phi(t)\frac{\chi_n}{6\sigma_n^3\sqrt{n}}(-it)^3\right|dt \\ &= \frac{1}{2\pi n}\int_{\mathbb{R}} \frac{\sqrt{n}}{\sqrt{t^2+\tau_a^2\sigma_n^2 n}}\left|\frac{\chi_n}{6\sigma_n^3}\right| t^3\Phi(t)dt \\ &= O(n^{-1}), \qquad (4.68) \end{aligned}$$

by the uniform bound of $\{\chi_n/\sigma_n^3\}$ proved in Lemma 4.5.2 and the dominated convergence theorem.

Since the the left-hand side of (4.67) is a real number, the first term in the integral on the right-hand side takes the form

$$\begin{aligned} \frac{1}{2\pi}\int \frac{e^{itc_a^n}}{it+\tau_a\sigma_n\sqrt{n}}\Phi(t)dt &= \frac{1}{2\pi}\int \mathrm{Re}\left\{\frac{e^{itc_a^n}}{it+\tau_a\sigma_n\sqrt{n}}\right\}\Phi(t)dt \\ &= \frac{1}{2\pi}\int \frac{\cos(tc_a^n)\tau_a\sigma_n\sqrt{n} + t\sin(tc_a^n)}{t^2+\sigma_n^2\tau_a^2 n}\Phi(t)dt \end{aligned}$$

$$= \frac{1}{2\pi} \int \frac{\cos(tc_a^n)\tau_a \sigma_n \sqrt{n}}{t^2 + \sigma_n^2 \tau_a^2 n} \Phi(t) dt + O(n^{-1}), \tag{4.69}$$

where the last equality follows from the convergence of σ_n to σ_a and the dominated convergence theorem.

From the two estimates (4.68) and (4.69), the claim is proved. □

Combining all the claims, we conclude that

$$\begin{aligned} \mathbb{P}_\theta(W^{n,\infty} > a) &= \frac{e^{-n\mathbb{I}_\infty(a) + nR_a^n(\theta^n)}}{\sigma_n \tau_a \sqrt{2\pi n}} \left(\frac{1}{\sqrt{2\pi}} \int \frac{\cos(tc_a^n(\theta^n))\tau_a^2 \sigma_n^2 n}{t^2 + \sigma_n^2 \tau_a^2 n} \Phi(t) dt + o(1) \right) \\ &= \frac{e^{-n\mathbb{I}_\infty(a) + nR_a^n(\theta^n)}}{\sigma_n \tau_a \sqrt{2\pi n}} \left(\frac{1}{\sqrt{2\pi}} \int \cos(tc_a^n(\theta^n)) \Phi(t) dt + o(1) \right) \\ &= \frac{e^{-n\mathbb{I}_\infty(a) + nR_a^n(\theta^n) + c_a^n(\theta^n)^2}}{\sigma_n \tau_a \sqrt{2\pi n}} (1 + o(1)). \end{aligned} \tag{4.70}$$

4.5.4 A central limit theorem result

To identify the order of $R_a^n(\Theta^n)$ and $c_a^n(\Theta^n)$.

Proposition 4.5.6. *There exist sequences of random variables $\{r_n := r_a^n\}_{n\in\mathbb{N}}$, $\{s_n := s_a^n\}_{n\in\mathbb{N}}$ and $\{t_n := t_a^n\}_{n\in\mathbb{N}}$ such that for $n \in \mathbb{N}$,*

$$(R_a^n(\Theta^n), \sigma_n c_a^n(\Theta^n)) \overset{(d)}{=} \left(r_n + \frac{1}{\sqrt{n}} s_n + o\left(\frac{1}{\sqrt{n}}\right), t_n + o(1) \right).$$

Further, let $f(x) := \Lambda(\tau_a x)$. We have the following convergence

$$(r_n, s_n, t_n) \Rightarrow \left(\tilde{A} - \frac{1}{2}\mathbb{E}[f'(Z)Z]\tilde{D}, \frac{1}{8}\mathbb{E}[f''(Z)Z^2]\tilde{D}^2, \frac{1}{\tau_a}\left(\tilde{B} - \frac{1}{2}\left(\mathbb{E}[f'(Z)Z] + \mathbb{E}[f''(Z)Z^2] \right) \tilde{D} \right) \right),$$

where Z is a standard Gaussian random variable and $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$ is jointly Gaussian with mean 0 and covariance matrix Σ_∞ defined to be the covariance matrix of the vector $(f(Z), f(Z), f''(Z)Z^2, Z^3)$.

Combining Theorem 4.2.3 and Proposition 4.5.6, we immediately obtain the following corollary, which is of a similar flavor to the case for ℓ^p spheres and balls of Corollaries 2.5 and 2.9 in [58].

Corollary 4.5.7. *Suppose Assumptions 4.2.2 holds. Fix $a > \mathbb{E}[X_1]$ such that $\mathbb{I}_\infty(a) < \infty$. Assuming the constants defined in Theorem 4.2.3 and Proposition 4.5.6, we have*

$$\mathbb{P}_\Theta(W^{n,\infty} > a) \stackrel{(d)}{=} \frac{e^{s_n + t_n^2/\sigma_n^2}}{\sigma_n \tau_a \sqrt{2\pi n}} e^{-n\mathbb{I}_\infty(a) + \sqrt{n} r_n}(1 + o(1)).$$

With Lemma 4.4.2, we need only to identify the corresponding functions F and G in Proposition 4.5.6.

Proof of Proposition 4.5.6. Let $F(x) := f(x) = \Lambda(\tau_a x)$ with Λ as in (4.7). Define $G(x) := x\Lambda'(\tau_a x)$. Then F and G are smooth functions by the smoothness of moment generating functions. Also recall from (4.61), (4.51), and (4.8) that the expression for R_a^n is

$$\begin{aligned} R_a^n(\Theta^n) = \sqrt{n}\,(\Psi_n(\tau_a) - \Psi(\tau_a)) &= \sqrt{n}\left(\int \Lambda(\tau_a u) L_{n,\Theta}(du) - \int \Lambda(\tau_a u)\mu_2(du)\right) \\ &= \sqrt{n}\left(\int F(u) L_{n,\Theta}(du) - \int F(u)\mu_2(du)\right), \end{aligned} \tag{4.71}$$

and from (4.62), (4.57) and the fact that $\Psi'(\tau_a) = a$ from (4.54) that c_a^n is given as

$$-c_a^n = \frac{\sqrt{n}(a_n - a)}{\sigma_n} = \frac{1}{\sigma_n}\sqrt{n}\left(\int u\Lambda'(\tau_a u) L_{n,\Theta}(du) - \int u\Lambda'(\tau_a u)\mu_2(du)\right)$$

$$= \frac{1}{\sigma_n}\sqrt{n}\left(\int G(u)L_{n,\Theta}(du) - \int G(u)\mu_2(du)\right).$$

Note that we need Assumption 4.2.2 to interchange differentiation and integration for $\Psi'(\tau_a)$.

By Assumption 4.2.2, F and G satisfy the assumption in the lemma, and therefore we get the expansion. Moreover, observe that $G'(x) = xG'(x)/\tau_a$ and $G''(x) = (F'(x) + xF''(x))/\tau_a$. We may further simplify the expression for t_n and obtain

$$t_n \Rightarrow \frac{1}{\tau_a}\left(\tilde{B} - \frac{1}{2}\left(\mathbb{E}[F'(Z)Z] + \mathbb{E}[F''(Z)Z^2]\right)\tilde{D}\right).$$

□

4.5.5 A refined CLT under the change of measure

In this section, we prove a Berry-Esseen bound for U^n. We apply the Edgeworth expansion in Theorem 4.3 and Corollary 4.4 of [7].

Define

$$Y_j^n := \sqrt{n}(X_j^n\theta_j^n - \mathbb{E}_{\mathbb{Q}^n}[X_j^n\theta_j^n]).$$

We start with a lemma to show that the assumptions in the corollary is satisfied.

Lemma 4.5.8. *Suppose Assumption 4.2.2 holds. Then for σ-a.e. θ,*

$$\lim_{n\to\infty}\sigma_n = \mathbb{E}\left[Z^2\Lambda''(\tau_a Z)\right] > 0,$$

where Z denotes a standard Gaussian random variable. Moreover, for σ-a.e. θ,

$$\sup_n \frac{\chi_n}{\sigma_n^3} < \infty, \quad \textit{and}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \mathbb{E}_{\mathbb{Q}^n}\left[\left| Y_j^n \right|^3 1_{\left\{ \left| Y_j^n \right| > \epsilon \sqrt{n} \right\}} \right] \to 0. \tag{4.72}$$

Proof. For σ-a.e. θ, by the independence of $\{X_j^n\}_{j=1,\ldots,n}$ under $\mathbb{Q}^n$, (4.6) and (4.57),

$$\frac{1}{n} \sum_{j=1}^{n} \mathrm{Var}_{\mathbb{Q}^n}(Y_j^n) = n\mathrm{Var}_{\mathbb{Q}^n}(W^{n,\infty}) = \sigma_n^2.$$

We now identify the limit of the sequence $\{\sigma_n^2\}_{n\in\mathbb{N}}$. Since $\sigma_n^2 = \Psi_n''(\tau_a)$, by (4.51) and taking derivative, we obtain

$$\begin{aligned} \sigma_n^2 &= \frac{1}{n} \sum_{j=1}^{n} (\sqrt{n}\theta_j^n)^2 \Lambda''(\sqrt{n}\theta_j^n \tau_a) \\ &= \int u^2 \Lambda''(\tau_a u) L_\theta^n(du). \end{aligned}$$

Note that by Assumption 4.2.2, the function $u^2\Lambda''(\tau_a u)$ has polynomial growth. By Lemma 4.4.1, we see that for σ-a.e. θ, as n tends to infinity,

$$\sigma_n^2 = \int u^2 \Lambda''(\tau_a u) L_\theta^n(du) \to \int u^2 \Lambda''(\tau_a u) \mu_2(du) = \mathbb{E}\left[Z^2 \Lambda''(\tau_a Z) \right], \tag{4.73}$$

where Z denotes a standard Gaussian random variable. The variance is zero only when $\Lambda'' \equiv 0$, in which case X_1 is a point mass. However, this violates the Cramér condition in Assumption 4.2.2. Therefore, the variance is strictly positive.

Define $\rho_k(n)$ for $k = 3, 4$ to be

$$\rho_k(n) := \frac{1}{n} \sum_{j=1}^{n} \mathbb{E}_{\mathbb{Q}^n} \left[\left| Y_j^n \right|^k \right].$$

Suppose Assumption 4.2.2 holds. From the relation between cumulants and central moments, we have

$$\rho_4(n) = 3\sigma_n^2 + \frac{1}{n} \sum_{i=1}^{n} \left(\Lambda(\tau_a \sqrt{n} \theta_i^n) \right)'''' . \tag{4.74}$$

The first term converges as is shown in (4.73). By Assumption 4.2.2, the second term also converges from Lemma 4.4.1. Therefore $\rho_4(n)$ is uniformly bounded in n.

By the Cauchy-Schwartz inequality and the concavity of $x \mapsto x^{3/4}$, we have

$$\begin{aligned}
\rho_3(n) &\leq \frac{1}{n} \sum_{j=1}^{n} \left(\mathbb{E}_{\mathbb{Q}^n} [\left| Y_j^n \right|^4] \right)^{3/4} \\
&\leq \left(\frac{1}{n} \sum_{j=1}^{n} \mathbb{E}_{\mathbb{Q}^n} [\left| Y_j^n \right|^4] \right)^{3/4} \\
&= \left(\rho_4(n) \right)^{3/4} .
\end{aligned}$$

The term on the right is uniformly bounded, and therefore $\rho_3(n)$ is also uniformly bounded in n.

Combining the convergence of $\{\sigma_n\}_{n \in \mathbb{N}}$, (4.73), and the positivity of σ_a^2, with the boundedness of third absolute moment, we conclude that the coefficient χ_n / σ_n^3 is uniformly bounded.

Next, we show that for $\epsilon > 0$, (4.72) holds for σ-a.e. θ. On the set $\left\{ \left| Y_j^n \right| > \epsilon \sqrt{n} \right\}$ we have

$$\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}_{\mathbb{Q}^n}\left[\left|Y_j^n\right|^3 1_{\left\{\left|Y_j^n\right|>\epsilon\sqrt{n}\right\}}\right] \leq \frac{1}{\epsilon\sqrt{n}n}\sum_{i=1}^{n}\mathbb{E}_{\mathbb{Q}^n}\left[\left|Y_j^n\right|^4 1_{\left\{\left|Y_j^n\right|>\epsilon\sqrt{n}\right\}}\right]$$
$$\leq \frac{1}{\epsilon\sqrt{n}n}\sum_{i=1}^{n}\mathbb{E}_{\mathbb{Q}^n}\left[\left|Y_j^n\right|^4\right]$$
$$= \frac{1}{\epsilon\sqrt{n}}\rho_4(n).$$

Again by the uniformly boundedness of $\rho_4(n)$, the term on the right-hand side converges to 0. Thus, the limit (4.72) holds. □

Proof of Proposition 4.5.2. The first and last assertions follow immediately from Lemma 4.5.8. Next let $\phi_{Y_j^n}(t)$ be the characteristic function of Y_j^n under $\mathbb{Q}^n$,

$$\phi_{Y_j^n}(t) := \mathbb{E}_{\mathbb{Q}^n}\left[e^{itY_j^n}\right].$$

Then we claim that

$$\lim_{|t|\to\infty}\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\left|\phi_{Y_j^n}(t)\right| = 0.$$

To see why the claim is true, by definition of the change of measure in (4.55), we have

$$\frac{1}{n}\sum_{j=1}^{n}\left|\phi_{Y_j^n}(t)\right| = \frac{1}{n}\sum_{j=1}^{n}\left|\mathbb{E}_{\mathbb{Q}^n}\left[e^{itY_j^n}\right]\right|$$
$$= \frac{1}{n}\sum_{j=1}^{n}\left|\frac{\mathbb{E}\left[e^{it\sqrt{n}X_j^n\theta_j^n+\tau_a\sqrt{n}X_j^n\theta_j^n}\right]}{\mathbb{E}\left[e^{\tau_a\sqrt{n}X_j^n\theta_j^n}\right]}\right|$$
$$= \frac{1}{n}\sum_{j=1}^{n}\left|e^{\Lambda((it+\tau_a)\sqrt{n}\theta_j^n)-\Lambda(\tau_a\sqrt{n}\theta_j^n)}\right|$$

$$= \int \left| e^{\Lambda((it+\tau_a)u) - \Lambda(\tau_a u)} \right| L_\theta^n(du).$$

Since $\phi_{Y_j^n}$ is a characteristic function, the norm of the characteristic function is smaller than 1 for any real number t. We can apply the convergence of $L_{n,\theta}$ to μ_2 from Lemma 4.4.1 and the continuity of

$$z \mapsto \left| e^{\Lambda((it+\tau_a)z) - \Lambda(\tau_a z)} \right| \quad \text{for} \quad z \in \mathbb{R},$$

to conclude that for σ-a.e. θ,

$$\frac{1}{n} \sum_{j=1}^{n} \left| \phi_{Y_j^n}(t) \right| \to \mathbb{E} \left| e^{\Lambda((it+\tau_a)Z) - \Lambda(\tau_a Z)} \right|. \tag{4.75}$$

By Assumption 4.2.2, the characteristic function satisfies the Cramér condition,

$$\limsup_{|t| \to \infty} \left| e^{\Lambda((it+\tau_a)x)} \right| < 1, \text{ for } x \in \mathbb{R}/\{0\}.$$

Note that the term in the expectation on the right-hand side of (4.75) is again bounded by 1 almost surely. Hence, by the dominated convergence theorem, we have

$$\limsup_{|t| \to \infty} \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} \left| \phi_{Y_j^n}(t) \right| < 1.$$

Now, we can apply Corollary 4.4 in [7] to Y_j^n, and use (4.56) to obtain

$$H^n(x) = \mathbb{Q}^n(U^n \leq x) = \mathbb{Q}^n \left(\frac{\sum_{j=1}^n Y_j^n}{\sqrt{n}\sigma_n} \leq x \right).$$

Thus, we conclude from the corollary that for σ-a.e. θ,

$$\sup_{x\in\mathbb{R}} \left|H^n(x) - \widetilde{H}^n(x)\right| = o\left(\frac{1}{\sqrt{n}}\right).$$

□

4.6 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 4.3.4 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 (4.13), 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 (4.38), we then have

$$\mathbb{P}_\theta\left(W^{(n,p)} > a\right) = \mathbb{P}_\theta\left(\bar{S}^{(n,p)} \in \bar{D}_{p,a}\right) = \int_{\bar{D}_a} \bar{h}_\theta^n(x, y)dxdy, \tag{4.76}$$

where $\bar{D}_a = \bar{D}_{p,a}$ is the domain defined in (4.39).

Remark 4.6.1. Note that $\bar{h}_\theta^n$ depends on θ only through θ^n. For notational simplicity throughout we will adopt the convention that for quantities that depend on both n and θ^n, we will use a superscript n to denote the former dependence and a subscript θ instead of θ^n to denote the dependence on θ^n.

The key ingredients required to estimate the tail probability in (4.76) are an asymptotic expansion for the joint density $\bar{h}_\theta^n$ in Proposition 4.6.4 of Section 4.6.2, a multidimensional generalized Laplace approximation stated in Proposition 4.6.7 in Section 4.6.3, and a certain estimate that justifies the application of this Laplace approximation that is stated in Lemma 4.6.8 of Section 4.6.4. The proofs of Proposition 4.6.4 is somewhat involved and hence deferred to Section 4.8. Instead, these results are first used in Sections 4.6.5 and 4.6.6 to prove Theorem 4.3.5. We first state a preliminary result in Section 4.6.1.

4.6.1 Estimates on the joint logarithmic moment generating function

We obtain an estimate on the growth of the log moment generating function Λ_p of $(Y_j, |Y_j|^p)$ defined in (4.14), which will be useful in the subsequent discussion. The following expression was established in [33, Lemma 5.7]:

$$\Lambda_p(t_1, t_2) = -\frac{1}{p}\log(1 - pt_2) + \log M_{\gamma_p}\left(\frac{t_1}{(1 - pt_2)^{1/p})}\right), \tag{4.77}$$

for

$$(t_1, t_2) \in \mathbb{D}_p := \left\{(t_1, t_2) \in \mathbb{R}^2 : t_2 < 1/p\right\}, \tag{4.78}$$

where

$$M_{\gamma_p}(t) := \mathbb{E}\left[e^{tY_j}\right], \quad t \in \mathbb{R}, \tag{4.79}$$

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

Lemma 4.6.2. *For $1 < p < \infty$, let M_{γ_p} and Λ_p be as defined in (4.79) and (4.14), respectively. Then for every $k \in \mathbb{N} \cup \{0\}$,*

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

exists and has at most polynomial growth, in the sense of Definition 4.1.1. Therefore, for

$j, k \in \mathbb{N} \cup \{0\}$, and any $t_2 < 1/p$,

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

has at most polynomial growth.

The proof of Lemma 4.6.2 involves conceptually straightforward (though detailed) estimates, and is thus deferred to Section 4.11.

4.6.2 An asymptotic expansion for the joint density

The main result of this section is Proposition 4.6.4, which provides an asymptotic expansion for the joint density $\bar{h}^n_\theta$ of the two-dimensional random vector $\bar{S}^n$ under $\mathbb{P}_\theta$. To state the result, for $n \in \mathbb{N}$, define

$$\bar{V}^n_j := (\sqrt{n}\Theta^n_j Y_j, \left|Y_j\right|^p), \qquad j = 1, \ldots, n. \tag{4.80}$$

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

$$\Phi_p(t_1, t_2) := \mathbb{E}\left[e^{t_1 Y_j + t_2 |Y_j|^p}\right]. \tag{4.81}$$

The observation $|e^{t_1 Y_j + t_2 |Y_j|^p}| = e^{\mathrm{Re}\{t_1\} Y_j + \mathrm{Re}\{t_2\} |Y_j|^p}$ shows that Φ_p is finite precisely when $\mathrm{Re}\{t_2\} < 1/p$, or equivalently, $(\mathrm{Re}\{t_1\}, \mathrm{Re}\{t_2\})$ lies in $\mathbb{D}_p$, the effective domain of Λ_p defined in (4.78). For $t = (t_1, t_2) \in \mathbb{D}_p$ and $\theta \in \mathbb{S}$, also define

$$\Psi^n_{p,\theta}(t) := \frac{1}{n} \sum_{j=1}^n \log \Phi_p(\sqrt{n}\theta^n_j t_1, t_2) = \int_{\mathbb{R}} \log \Phi_p(u t_1, t_2) L^n_\theta(du), \tag{4.82}$$

where L_θ^n is the empirical measure of the coordinates of $\sqrt{n}\theta^n$, as defined in (4.47).

Remark 4.6.3. 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 4.6.2. Hence, Lemma 4.4.1 shows that for every $t = (t_1,t_2) \in \mathbb{D}_p$ and σ-a.e. θ, as $n \to \infty$,

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

where the last equality holds by the definition of Ψ_p given in (4.15).

Next, recall the definition of $\mathbb{J}_p$ from (4.17) and for $x \in \mathbb{J}_p$, also define for $\theta \in \mathbb{S}$,

$$c_x^n(\theta^n) := \sqrt{n}\nabla\left(\Psi_{p,\theta}^n(\lambda_x) - \Psi_p(\lambda_x)\right), \qquad \mathcal{H}_x^n(\theta^n) := \mathrm{Hess}\,\Psi_{p,\theta}^n(\lambda_x), \tag{4.83}$$

and

$$R_x^n(\theta^n) := \sqrt{n}(\Psi_{p,\theta}^n(\lambda_x) - \Psi_p(\lambda_x)), \tag{4.84}$$

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 indeed depend on θ only through θ^n (see Remark 4.6.1).

For $a > 0$, with the same abuse of notation used for $\mathcal{H}_a$ earlier, 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 4.8.2 that $c_x^n(\theta^n)$ and $\mathcal{H}_x^n(\theta^n)$ are the mean 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 (4.80), under a certain quenched tilted measure; see (4.134) and (4.135).

Proposition 4.6.4. *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* (4.15), (4.16), (4.17) *and* (4.82)*, 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* (4.22), (4.83) *and* (4.84)*, respectively. Then for σ-a.e. θ,*

$$\bar{h}_\theta^n(x) = \frac{n}{2\pi}\bar{g}_\theta^n(x)e^{-n\Psi_p^*(x)}(1 + o(1)), \tag{4.85}$$

and the expansion in (4.85) *is uniform on any compact subset of* $\mathbb{J}_p$*, where* $\bar{g}^n_\theta$ *is the smooth function defined by*

$$\bar{g}^n_\theta(x) := \det \mathcal{H}_x^{-1/2} e^{\sqrt{n}R^n_x(\theta^n)} e^{\left\|\mathcal{H}_x^{-1/2} c^n_x(\theta^n)\right\|^2}. \tag{4.86}$$

Section 4.8 is devoted to establishing Proposition 4.6.4, with the final proof given in Section 4.8.4.

4.6.3 A multidimensional generalized Laplace approximation

Recall from (4.76) that the tail probability of interest is obtained by integrating the function $\bar{h}^n_\theta$ specified in Proposition 4.6.4 over the domain $\bar{D}_a$ defined in (4.39). Due to the additional dependence of n in $\bar{g}^n_\theta$, we cannot directly apply conventional Laplace approximations such as those in [15, Chapter 8] or [83, Chapter V]. Instead, in Propositions 4.6.6 and 4.6.7, we first establish a generalization of multidimensional Laplace approximations that can accommodate such n-dependent terms

Definition 4.6.5. Given $m, d \in \mathbb{N}$, and a bounded domain $D \subset \mathbb{R}^{m+d}$, we say that the sequence $h^n : \mathbb{R}^{m+d} \to \mathbb{R}$, $n \in \mathbb{N}$, admits a (f, x^*, α, g^n)-representation if for each $n \in \mathbb{N}$,

$$h^n(x) = g^n(x)e^{-nf(x)}, \quad x \in D,$$

where

1. f is a nonnegative function that is smooth in D and achieves its minimum on $\mathrm{cl}(D)$, the closure of D, at a unique point x^*,

2. there exists $\alpha \in (0, 1)$ such that for each $n \in \mathbb{N}$ sufficiently large, $g^n(x) = \exp(r^n(x))$

is smooth with

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

We start with a multidimensional extension of a result in [66].

Proposition 4.6.6. *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 $h^n : \mathbb{R}^{m+d} \to \mathbb{R}$, $n \in \mathbb{N}$, admits a (f, x^*, α, g^n)-representation with $x^* = (0, 0, \ldots, 0)$. Then we have the following asymptotic expansion:*

$$\int_D h^n(x)dx = \frac{(2\pi)^{m/2}}{n^{d+m/2}} \frac{g^n(x^*)}{\prod_{i=1}^d \partial_{m+d} f(x^*) \sqrt{\prod_{j=1}^m \left|\partial^2_{j,j} f(x^*)\right|}} e^{-nf(x^*)}(1 + o(1)). \tag{4.87}$$

The proof is deferred to Section 4.13. We now obtain an alternative representation for this integral. Recall the definition of Weingarten maps, for example, from [6, Section 4]. 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 $L_x : T_x(\mathcal{D}) \to T_x(\mathcal{D})$ where $L_x(v) := \partial_v N_x$ and ∂_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.

Proposition 4.6.7. *Let $\mathcal{D} \subset \mathbb{R}^d$ be a bounded domain whose boundary is a differentiable $(d-1)$-dimensional hypersurface. Let $h^n : \mathbb{R}^{m+d} \to \mathbb{R}$, $n \in \mathbb{N}$, be a sequence of functions admitting a (f, x^*, α, g^n)-representation in $\mathcal{D}$ in the sense of Definition 4.6.5. Then*

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

where for $i = 1, 2$, L_i is the Weingarten map at $x^* \in \partial D$ of the surface C_i, where

$$C_1 := \{y : f(y) = f(x^*)\} \qquad \textit{and} \qquad C_2 := \partial D.$$

Proof. The proof will make use of arguments in [15] as well as a result from [6]. Let h^n and (f, x^*, α, g^n) be as in the statement of the lemma. We first consider the special case $g^n \equiv 1$, when the result can be deduced via classical Laplace asymptotics. Specifically, let $\tilde{h}^n := e^{-nf(x)}$ and

$$\widetilde{\mathcal{I}}^n := \int_{\mathcal{D}} \tilde{h}^n(x)dx = \int_{\mathcal{D}} e^{-nf(x)}.$$

Then $\widetilde{\mathcal{I}}^n$ coincides with the integral in [15, Equation (8.3.63)] when λ, n, ϕ 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 $\mathcal{G} : \mathcal{N}_x \to \mathcal{U}$ 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 $\mathcal{D}$, [15, Equation (8.3.63)] gives the following estimate

$$\int_{\mathcal{D}} e^{-nf(x)}dx = \frac{(2\pi)^{(d-1)/2} \left|\det(\mathcal{J}_*^T \mathcal{J}_*)\right|^{1/2}}{n^{(d+1)/2} |\det \mathrm{Hess}(f \circ \mathcal{G}(x^*))|^{1/2} |\nabla f(x^*)|} e^{-nf(x^*)}(1 + o(1)). \tag{4.88}$$

Next, to further simplify the expression in the last display, by [6, Equations (4.5) and (4.6)], it follows on identifying $DG(0)$, I and A therein with $\mathcal{J}_*$, f and $\mathrm{Hess}(f \circ \mathcal{G}(x^*))$, respectively, that

$$\frac{\left|\det(\mathcal{J}_*^T \mathcal{J}_*)\right|^{1/2}}{|\det \mathrm{Hess}(f \circ \mathcal{G}(x^*))|^{1/2} |\nabla f(x^*)|} = \frac{\det(L_1^{-1}(L_1 - L_2))^{-1/2}}{\langle \mathrm{Hess}\, f^{-1}(x^*) \nabla f(x^*), \nabla f(x^*)\rangle^{1/2}}, \tag{4.89}$$

with L_1, L_2 as in the lemma. (Note that there is an erroneous additional factor of $\sqrt{2\pi n}$ in the denominator of equation Equation (4.6) in [6], which we have corrected.) The proposition in the case $g^n \equiv 1$ then follows on combining the last two displays.

On the other hand, let $\mathcal{N}$ be a neighborhood of x^*. Since $\partial\mathcal{D}$ is a differentiable $(d-1)$-dimensional hypersurface, there exists an one-to-one continuously differentiable transformation $F : \mathcal{N} \to \tilde{\mathcal{N}} \subset \mathbb{R} \times \mathbb{R}_+^{d-1}$ such that F maps x^* to the origin. Setting $J_F(x) \in \mathbb{R}^d \times \mathbb{R}^d$ to be the Jacobian matrix of F at x. We can write

$$\widetilde{\mathcal{I}}^n := \int_{\mathcal{D}} e^{-nf(x)} dx = \int_{F(\mathcal{D})} |\det J_F(x)|\, e^{-nf(F^{-1}(x))} dx.$$

Then Proposition 4.6.6 with m, d g^n and f therein replaced with 1, $d-1$, $|\det J_F(x)|$ and $f \circ F^{-1}$, respectively, implies there exists a constant $C' = C'(F, \mathcal{D}, f) \in (0, \infty)$ such that

$$\widetilde{\mathcal{I}}^n = \frac{(2\pi)^{(d-1)/2} C'}{n^{(d+1)/2}} e^{-nf(x^*)}(1 + o(1)). \tag{4.90}$$

Comparing (4.88), (4.89) and (4.90), we see that

$$C' = \frac{\det(L_1^{-1}(L_1 - L_2))^{-1/2}}{\langle \operatorname{Hess} f^{-1}(x^*) \nabla f(x^*), \nabla f(x^*) \rangle^{1/2}}. \tag{4.91}$$

Now consider the case of general g^n and let $\mathcal{N}$ be a neighborhood of x^*. By the assumption in Definition 4.6.5, 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 F, we again have $\left|r^n(F^{-1}(x))\right| \leq Cn^\alpha \|x\|_2$. Hence, under the same transformation F, we may apply Proposition 4.6.6 and (4.91) to obtain

$$\begin{aligned}
\mathcal{I}^n &= \frac{(2\pi)^{(d-1)/2} C'}{n^{(d+1)/2}} g^n(x^*) e^{-nf(x^*)}(1 + o(1)) \\
&= \frac{(2\pi)^{(d-1)/2} \det(L_1^{-1}(L_1 - L_2))^{-1/2}}{n^{(d+1)/2} \langle \operatorname{Hess} f^{-1}(x^*) \nabla f(x^*), \nabla f(x^*) \rangle^{1/2}} g^n(x^*) e^{-nf(x^*)}(1 + o(1)).
\end{aligned}$$

□

4.6.4 Continuity estimates for terms in the prefactor

In order to apply Proposition 4.6.7 to the expression for $\bar{h}^n_\theta$ given in (4.85)–(4.86), we need to verify that $\bar{h}^n_\theta$ satisfies Definition 4.6.5. The following lemma will be useful in verifying property (2) of Definition 4.6.5.

Lemma 4.6.8. *Fix $p \in (1,\infty)$. For every $y \in \mathbb{J}_p$ and $\alpha \in (1/2,1)$, for $\varepsilon > 0$ with $B_\varepsilon(y) \subset \mathbb{J}_p$, there exists $\Xi = \Xi(y,\alpha,\varepsilon) \in (0,\infty)$, for σ-a.e. θ, there exists $N = N(y,\alpha,\varepsilon,\theta) \in \mathbb{N}$ such that for $x \in B_\varepsilon(y)$ and $n \geq N$,*

$$\left| \sqrt{n} R^n_x(\theta^n) - \sqrt{n} R^n_y(\theta^n) \right| \leq \Xi n^\alpha \left\| x - y \right\|_2, \tag{4.92}$$

and

$$\left| \left\| \mathcal{H}_x^{-1/2} c^n_x(\theta^n) \right\|_2^2 - \left\| \mathcal{H}_y^{-1/2} c^n_y(\theta^n) \right\|_2^2 \right| \leq \Xi n^{\alpha - 1/2} \left\| x - y \right\|_2. \tag{4.93}$$

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 (4.14), (4.81) and the fact that Y_j is a p-th Gaussian random variable, it follows that $\Lambda_p = \log \Phi_p$ on $\mathbb{D}_p$, where $\mathbb{D}_p$ is the domain of Φ_p defined in (4.78). When combined with (4.84), (4.82) and (4.15), this shows that

$$\sqrt{n} R^n_x(\theta^n) = \sum_{j=1}^n \log \Phi_p(\sqrt{n}\theta^n_j \lambda_{x,1}, \lambda_{x,2}) - \mathbb{E}\left[\log \Phi_p(Z\lambda_{x,1}, \lambda_{x,2}) \right],$$

where Z is a standard normal random variable. Since $x \mapsto \lambda_x$ is smooth by Remark 4.3.1, there exists $C' \in (0,\infty)$ such that $\left\| \lambda_x - \lambda_y \right\| \leq C' \left\| x - y \right\|$ for $x \in B_\varepsilon(y)$. Therefore, to show (4.92), it suffices to show that given any fixed $s = (s_1, s_2) \in \mathbb{D}_p$, for $\varepsilon' > 0$ with $B_{\varepsilon'}(s) \subset \mathbb{D}_p$, there exist $C' \in (0, \infty)$ and a random integer $N = N(s, \varepsilon')$ such that $\mathbb{P}$-almost surely,

$$\left| \sum_{j=1}^{n} \left(\mathcal{K}_s(\sqrt{n}\Theta_j^n t_1, t_2) - \mathbb{E}\left[\mathcal{K}_s(Z t_1, t_2) \right] \right) \right| \leq C n^{\alpha} \|t\|, \quad \text{for} \quad \|t\| < \varepsilon' \quad \text{and } n \geq N, \tag{4.94}$$

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

$$\mathcal{K}_s(t_1, t_2) := \log \Phi_p((s_1 + t_1), (s_2 + t_2)) - \log \Phi_p(s_1, s_2). \tag{4.95}$$

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 $Z^{(n)} := (Z_1, \ldots, Z_n)$, we have (e.g. see Section 4.3.4 or [74, Lemma 1]),

$$\left(\Theta_1^n, \ldots, \Theta_n^n \right) \stackrel{(d)}{=} \frac{(Z_1, \ldots, Z_n)}{\left\| Z^{(n)} \right\|}. \tag{4.96}$$

In view of this, we first state the following lemma.

Lemma 4.6.9. *With the notation above, fix $\alpha \in (1/2, 1]$. Let $\mathbb{D} = \mathbb{R} \times (-\infty, T)$ and let $\mathcal{K} : \mathbb{D} \to \mathbb{R}$ 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, Definition 4.1.1. Then for $\varepsilon > 0$ with $B_\varepsilon(0) \subset \mathbb{D}$, there exist $C \in (0, \infty)$ and a random integer N on $(\Omega', \mathcal{F}', \mathbb{P}')$ such that $\mathbb{P}'$-almost surely*

$$\left| \sum_{j=1}^{n} \left(\mathcal{K}\left(\frac{\sqrt{n} Z_j}{\left\| Z^{(n)} \right\|} t_1, t_2 \right) - \mathbb{E}\left[\mathcal{K}(Z t_1, t_2) \right] \right) \right| \leq C n^{\alpha} \|t\|, \quad \textit{for} \quad \|t\| < \varepsilon \quad \textit{and } n \geq N. \tag{4.97}$$

We first show how this lemma can be used to prove Lemma 4.6.4, and only present the proof of Lemma 4.6.9 in Section 4.14.

Proof of Lemma 4.6.8. We begin with (4.92). First, note from (4.96) that a statement about $Z^{(n)}$ holds $\mathbb{P}'$-almost surely if and only if the same statement with $Z^{(n)}$ replaced by Θ^n holds $\mathbb{P}$-almost surely. Since Φ_p is finite near the origin, Φ_p is smooth on its domain $\mathbb{D}_p$ and hence, for any $s \in \mathbb{D}_p$, the functional $\mathcal{K}_s$ from (4.95) is twice continuously differentiable on its domain $\mathbb{D}_{p,s}$. Since $\log \Phi_p = \Lambda_p$ on $\mathbb{D}_p$, from the expression in (4.95), Lemma 4.6.2 implies 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 4.6.9 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 exists $C \in (0, \infty)$ and a random integer $N = N(s, \alpha, \varepsilon)$ such that (4.97) holds $\mathbb{P}'$-almost surely. Since $\sigma = P \circ (\Theta^{(n)})^{-1}$, the aforementioned equivalence between statements about $Z^{(n)}$ and Θ^n, this implies (4.94) holds and thus, (4.92) also holds.

We now turn to the proof of (4.93). For $i = 1, 2$, and $(t_1, t_2) \in \mathbb{D}_{p,s}$, define

$$\mathcal{K}_{s,i}(t_1, t_2) := \partial_i \log \Phi_p((s_1 + t_1), (s_2 + t_2)) - \partial_i \log \Phi_p(s_1, s_2). \tag{4.98}$$

Note that for $i = 1, 2$, by the smoothness of Φ_p, $\mathcal{K}_{s,i}$ from (4.95) 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 $\log \Phi_p = \Lambda_p$ on $\mathbb{D}_p$ by Lemma 4.6.2. Thus, for $i = 1, 2$, $\mathcal{K}_{s,i}$ satisfies the assumption in Lemma 4.6.9. Hence, (4.83), (4.82), (4.96), Lemma 4.6.9 and the equivalence between statements about $Z^{(n)}$ and Θ^n, imply that for every $y \in \mathbb{J}_p$ and $\alpha \in (1/2, 1]$, for $\varepsilon > 0$ with $B_\varepsilon(y) \subset \mathbb{J}_p$, there exists $\Xi = \Xi(y, \varepsilon, \alpha) \in (0, \infty)$, for σ-a.e. θ, there exists $N = N(y, \alpha, \varepsilon, \theta) \in \mathbb{N}$ such that for $x \in B_\varepsilon(y)$ and $n \geq N$,

$$\left| \sqrt{n} c^n_{i,x}(\theta^n) - \sqrt{n} c^n_{i,y}(\theta^n) \right| \leq \Xi n^\alpha \left\| x - y \right\|_2, \quad i = 1, 2, \tag{4.99}$$

where $c_x^n = (c_{1,x}^n, c_{2,x}^n)$. Now consider instead for $i = 1, 2$, $(t_1, t_2) \in \mathbb{D}$,

$$\mathcal{K}'_{s,i}(t_1, t_2) := \partial_i \log \Phi_p(t_1, t_2). \tag{4.100}$$

By the smoothness of $x \mapsto \lambda_x$ in Remark 4.3.1, there exist $C' \in (0, \infty)$ such that $\|\lambda_x\| \le C'$ for $x \in B_{\varepsilon'}(y)$. Hence, (4.100) and Lemma 4.6.9 imply that for every $y \in \mathbb{J}_p$ and $\alpha \in (1/2, 1]$, there exists $\Xi' = \Xi'(y, \varepsilon, \alpha) \in (0, \infty)$, for σ-a.e. θ, there exists $N' = N'(y, \alpha, \varepsilon, \theta) \in \mathbb{N}$ such that for $x \in B_\varepsilon(y)$ and $n \ge N'$,

$$\left| \sqrt{n} c_{i,x}^n(\theta^n) \right| \le \Xi' n^\alpha \|\lambda_x\|_2 \le C' \Xi' n^\alpha, \quad i = 1, 2. \tag{4.101}$$

Recall the definition of $\mathcal{H}_x$ in (4.22). The following relation

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

implies that the left-hand side of (4.93) can be written as

$$\left| \left\| \mathcal{H}_x^{-1/2} c_x^n(\theta^n) \right\|_2^2 - \left\| \mathcal{H}_y^{-1/2} c_y^n(\theta^n) \right\|_2^2 \right| \le \sum_{i=1}^{2} A_i^n B_i^n, \tag{4.102}$$

where for $i = 1, 2$,

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

Recall the definitions of Ψ_p and λ_x in (4.15) and (4.18), respectively. By the smoothness of Ψ_p and λ_x in Remark 4.3.1, $x \mapsto \mathcal{H}_x^{-1/2}$ is also smooth and there exists $C'' \in (0,\infty)$ such that $|(\mathcal{H}_x^{-1/2})_{ij} - (\mathcal{H}_y^{-1/2})_{ij}| \leq C'' \left\|x-y\right\|$ and $|(\mathcal{H}_x^{-1/2})_{ij}| \leq C''$ for $x \in B_\varepsilon(y)$ and $i,j=1,2$. Hence, by (4.99) and (4.101), for σ-a.e. θ, $x \in B_\varepsilon(y)$ and $n \geq \max\{N,N'\}$,

$$\begin{aligned} A_i^n &\leq \sum_{j=1}^{2} \left|((\mathcal{H}_x^{-1/2})_{ij} c_{j,x}^n(\theta^n) - (\mathcal{H}_y^{-1/2})_{ij} c_{j,y}^n(\theta^n))\right| \\ &\leq \left|(\mathcal{H}_x^{-1/2})_{ij} - (\mathcal{H}_y^{-1/2})_{ij}\right| \sum_{j=1}^{2} \left|c_{j,x}^n(\theta^n)\right| + \left|(\mathcal{H}_x^{-1/2})_{ij}\right| \sum_{j=1}^{2} \left|c_{j,x}^n(\theta^n) - c_{j,y}^n(\theta^n))\right| \\ &\leq 2(C''C'\Xi' + C''\Xi) n^{\alpha-1/2} \left\|x-y\right\|_2 . \end{aligned} \tag{4.103}$$

Next, by (4.101), for σ-a.e. θ, $x \in B_\varepsilon(y)$ and $n \geq N'$,

$$B_i^n \leq 2C''C'\Xi' n^{\alpha-1/2}. \tag{4.104}$$

Therefore, (4.93) follows from (4.102)-(4.104).

□

4.6.5 Proof of Theorem 4.3.5(i)

We are now ready to prove the main estimate (4.28). 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 (4.39). By (4.20), Lemma 4.3.2 and the fact that $\mathbb{I}_p(a)$ is increasing for $a \in \mathbb{R}_+$,

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

Hence, the infimum of Ψ_p^* in the closure $\mathrm{cl}(\bar{D}_a)$ of $\bar{D}_a$ is attained at $a^* := (a, 1)$. Moreover due to (4.21), the assumption $\mathbb{I}_p(a) < \infty$ implies $\Psi_p^*(a, 1) < \infty$, and hence, $a^* = (a, 1) \in \mathbb{J}_p$, defined in (4.17). Further, by (4.39), a^* is a point on the boundary $\partial\bar{D}_a$ of $\bar{D}_a$ where the boundary is smooth. Let U be any open neighborhood of a^* such that $U \subset \mathbb{R}^2_> := \{(x, y) \in \mathbb{R}^2 : x > 0, y > 0\}$, 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 (4.76) into two parts:

$$\mathbb{P}_\theta \left(\bar{S}^n \in \bar{D}_a\right) = \mathbb{P}_\theta \left(\bar{S}^n \in \bar{D}_a \cap U\right) + \mathbb{P}_\theta \left(\bar{S}^n \in \bar{D}_a \cap U^c\right). \tag{4.105}$$

The proof will proceed in two steps. In the key first step, we will estimate the first term on the right-hand side of (4.105) by integrating the estimate of the density $\bar{h}^n_\theta$ of $\bar{S}^n$ obtained in Proposition 4.6.4 over the domain $\bar{D}_a \cap U$, and then analyze the asymptotics of the resulting Laplace type integral, as $n \to \infty$ using Proposition 4.6.4, Proposition 4.6.7 and Lemma 4.6.8. 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 (4.105) is negligible.

Step 1. Using the expressions for $\bar{h}^n_\theta$ and $\bar{g}^n_\theta$ from (4.85) and (4.86), respectively, and the fact that the domain $\bar{D}_a \cap U$ is bounded, we have for σ-a.e. θ,

$$\mathbb{P}_\theta \left(\bar{S}^n \in \bar{D}_a \cap U\right) = \int_{\bar{D}_a \cap U} \bar{h}^n_\theta(x)dx = \frac{n}{2\pi}\mathcal{I}^n_\theta(1 + o(1)), \tag{4.106}$$

where

$$\mathcal{I}^n_\theta := \int_{\bar{D}_a \cap U} \bar{g}^n_\theta(x)e^{-n\Psi_p^*(x)}dx. \tag{4.107}$$

To apply Proposition 4.6.7 to (4.107), replace d, g^n, f, x^* and $\mathcal{D}$ with 2, $\bar{g}^n_\theta$, Ψ_p^*, a^* and $\bar{D}_a \cap U$, respectively, for $\theta \in \mathbb{S}$, and note that then $\mathcal{I}^n$ corresponds to $\mathcal{I}^n_\theta$ of (4.107). To see that the assumptions of the lemma are satisfied, note that the domain $\bar{D}_a \cap U$ is bounded with smooth boundary, $\bar{g}^n_\theta$ is smooth by Lemma 4.6.4. Further, note that by (4.15), (4.77) and Lemma 4.6.2, Ψ_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 [84, Section III.D], it follows that Ψ^*_p is twice differentiable at the point a^*, which lies in $\mathbb{J}_p$, the effective domain of Ψ^*_p and achieves its minimum uniquely at $a^* \in \partial(\bar{D}_a \cap U)$. The nonnegativity of Ψ^*_p follows since Ψ^*_p is a rate function. Thus, $f = \Psi^*_p$ and a^* satisfy property (1) of Definition 4.6.5 for $D = \mathcal{D}_a \cap U$. On the other hand, from (4.86), let $\bar{g}^n_\theta = \exp(r^n(x))$ with

$$r^n_\theta(x) := \log\det \mathcal{H}_x^{-1/2} + \sqrt{n} R^n_x(\theta^n) + \left\| \mathcal{H}_x^{-1/2} c^n_x(\theta^n) \right\|^2 .$$

Lemma 4.6.8 and the smoothness of $x \mapsto \log \mathcal{H}_x^{-1/2}$, which follows from Remark 4.3.1 and (4.22), imply that for a sufficiently small neighborhood of x^*, for any $\alpha \in (1/2, 1)$, there exists a finite random variable N such that for σ-a.e. θ, for all $n \geq N$, for σ-a.e. θ, r^n_θ satisfies property (2) of Definition 4.6.5. Thus, $\bar{h}^n_\theta$ admits a $(\Psi^*_p, a^*, \alpha, \bar{g}^n_\theta)$ representation and we may apply Proposition 4.6.7. Now, also note by the duality of the Legendre transform, and the definition of $\lambda_{a,j}$ in (4.23), we have

$$\partial_j \Psi^*_p(a^*) = \lambda_{a,j}, \quad \text{for} \quad j = 1, 2,$$

and $(\operatorname{Hess} \Psi^*_p)(a^*)$ is the inverse of $(\operatorname{Hess} \Psi_p)(\lambda_a)$. Since $(\operatorname{Hess} \Psi_p)(\lambda_a) = \mathcal{H}_a$ by (4.24), we conclude from Proposition 4.6.7 that

$$\mathcal{I}^n_\theta = \frac{(2\pi)^{1/2}}{n^{3/2}} \frac{(L_{a,1}^{-1}(L_{a,1} - L_{a,2}))^{-1/2}}{\langle \mathcal{H}_a \lambda_a, \lambda_a \rangle^{1/2}} \bar{g}^n_\theta(a^*) e^{-n\Psi^*_p(a^*)}(1 + o(1)), \tag{4.108}$$

where $L_{a,1}$ and $L_{a,2}$ are the Weingarten maps of the curves $C_1 := \{x \in \mathbb{R}^2 : \Psi^*_p(x) = \Psi^*_p(a, 1)\}$ and $C_2 := \{x \in \mathbb{R}^2 : x_1 = a x_2^{1/p}\}$, evaluated at $a^* = (a, 1)$.

To further simplify (4.108), note that it follows from [6, Example 4.3] 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 $T : \mathbb{R}^2 \to \mathbb{R}$, 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 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

$$L_{a,1} = \frac{\left|\lambda_{a,2}^2(\mathcal{H}_a^{-1})_{11} - 2\lambda_{a,1}\lambda_{a,2}(\mathcal{H}_a^{-1})_{12} + \lambda_{a,1}^2(\mathcal{H}_a^{-1})_{22}\right|}{(\lambda_{a,1}^2 + \lambda_{a,2}^2)^{3/2}}. \tag{4.109}$$

On the other hand, the curvature of the graph of a function $y = \widetilde{T}(x)$ at the point $(x, \widetilde{T}(x))$ for sufficiently smooth $\widetilde{T} : \mathbb{R} \to \mathbb{R}$ is given by $\left|\widetilde{T}''(x)\right| / (1 + (\widetilde{T}')^2(x))^{3/2}$. Recalling the definition of $\bar{D}_a$ from (4.39), we can apply this with $\widetilde{T}(x) = (x/a)^p$ to compute the curvature of $C_2 = \partial \bar{D}_a$ at a^* as:

$$L_{a,2} = \frac{p(p-1)a}{(a^2 + p^2)^{3/2}}. \tag{4.110}$$

Substituting these calculations back into the expressions (4.106), (4.107) and (4.108), and recalling the definition of ξ_a and κ_a from (4.25) and (4.26), we conclude that for σ-a.e. θ,

$$\mathbb{P}_\theta\left(\bar{S}^n \in \bar{D}_a \cap U\right) = \frac{C_a^n(\theta^n)}{\kappa_a \xi_a \sqrt{2\pi n}} e^{-n\mathbb{I}_p(a) + \sqrt{n} R_a^n(\theta^n)}(1 + o(1)). \tag{4.111}$$

Step 2. We now turn to the second term in (4.105). Note that by the continuity of Ψ_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 4.3.2 of the (quenched) large deviation principle for $\bar{S}^n$ established in [33, Proposition 5.3], Ψ_p^* achieves its unique minimum in $\bar{D}_a$ at $a^* = (a, 1)$. Thus, for σ-a.e. θ,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\theta \left(\bar{S}^n \in \bar{D}_a \cap U^c\right) \leq -\Psi_p^*(a^*) - \eta, \tag{4.112}$$

which shows that the term in (4.112) is negligible with respect to (4.111).

When combined, (4.29), (4.84), (4.86), (4.105), (4.111) and (4.112) together yield (4.28). This completes the proof of Theorem (i).

4.6.6 Proof of Theorem 4.3.5(ii)

We start by obtaining expansions for $R_a^n(\Theta^n)$ and $c_a^n(\Theta^n)$. First, note that the functions ℓ_a, $\ell_{a,1}$ and $\ell_{a,2}$ defined in (4.27) and their derivatives up to second order (for $\ell_{a,1}$ and $\ell_{a,2}$) and third order (for ℓ_a) are continuous and have at most polynomial growth by Lemma 4.6.2. 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 (4.49) and (4.48), respectively, we can apply (4.83), (4.84) and Lemma 4.4.2 to obtain

$$\begin{aligned}
R_a^n(\Theta^n) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\ell_a(\sqrt{n}\Theta_i^n) - \mathbb{E}[\ell_a(Z)] \right) \\
&\overset{(d)}{=} r_n + \frac{1}{\sqrt{n}} s_n + o\left(\frac{1}{\sqrt{n}}\right), \\
c_a^n(\Theta^n) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} \ell_{a,1}(\sqrt{n}\Theta_i^n) - \mathbb{E}[\ell_{a,1}(Z)] \\ \ell_{a,2}(\sqrt{n}\Theta_i^n) - \mathbb{E}[\ell_{a,2}(Z)] \end{pmatrix} \\
&\overset{(d)}{=} \begin{pmatrix} t_{n,1} \\ t_{n,2} \end{pmatrix} + o(1).
\end{aligned}$$

Moreover, Lemma 4.4.2 also shows that we have the convergence

$$\begin{aligned}
&(r_n, s_n, t_{n,1}, t_{n,2}) \Rightarrow \\
&\quad \left(\widetilde{\mathfrak{A}} - \frac{1}{2}\mathbb{E}[\ell_a'(Z)Z]\widetilde{\mathfrak{D}}, \frac{1}{8}\mathbb{E}[\ell_a''(Z)Z^2]\widetilde{\mathfrak{D}}^2, \widetilde{\mathfrak{E}} - \frac{1}{2}\mathbb{E}\left[\ell_{a,1}'(Z)Z\right]\widetilde{\mathfrak{D}}, \widetilde{\mathfrak{G}} - \frac{1}{2}\mathbb{E}\left[\ell_{a,2}'(Z)Z\right]\widetilde{\mathfrak{D}} \right),
\end{aligned}$$

where $(\widetilde{\mathfrak{A}}, \widetilde{\mathfrak{D}}, \widetilde{\mathfrak{E}}, \widetilde{\mathfrak{G}})$ is jointly Gaussian with mean 0 and covariance matrix (4.31).

4.7 Proof of the sharp large deviation estimate for balls

Fix $p \in (1, \infty)$ and $a > 0$ such that $\mathbb{I}_p(a) < \infty$. From the definitions in Section 4.3.4, specifically (4.42), note that we have the following expression for the tail probability of projections of ℓ_p^n balls:

$$\mathbb{P}_\theta\left(\widetilde{\mathcal{W}}^{(n,p)} > a\right) = \int_{\bar{\mathcal{D}}_a} \bar{\mathfrak{h}}_\theta^n(x_1, x_2, y)\,dx_1 dx_2 dy, \tag{4.113}$$

where for $\theta \in \mathbb{S}$, $\bar{\mathfrak{h}}^n_\theta(x_1, x_2, y)$ is the density under $\mathbb{P}_\theta$ of the random vector $\bar{\mathcal{S}}^{(n,p)} = (\bar{S}^{(n,p)}, {}^{1/n})$, defined in (4.41), and $\bar{\mathcal{D}}_a := \bar{\mathcal{D}}_{p,a} \subset \mathbb{R}^3$ is the domain defined in (4.43). By the independence of and $Y^{(n,p)}$, for $x \in \mathbb{R}^2$ and $y \in (0, 1]$, $\bar{\mathfrak{h}}^n_\theta(x_1, x_2, y)$ is the product of $\bar{h}^n_\theta(x_1, x_2)$, the density of $\bar{S}^{(n,p)}$ under $\mathbb{P}_\theta$ evaluated at (x_1, x_2), and the density of ${}^{1/n}$ at y, which is equal to $\frac{n}{y}e^{n \log y}$. Hence, by Proposition 4.6.4, we have the following uniform estimate of $\bar{\mathfrak{h}}^n_\theta$: for σ a.e. θ,

$$\bar{\mathfrak{h}}^n_\theta(x_1, x_2, y) = \frac{n^2}{2\pi}\frac{1}{y}\bar{g}^n_\theta(x_1, x_2)e^{-nF(x_1,x_2,y)}(1 + o(1)), \quad (x_1, x_2) \in \mathbb{R}^2, \quad y \in (0, 1], \tag{4.114}$$

where $\bar{g}^n_\theta$ is as defined in (4.86), and

$$F(x, y) := \Psi^*_p(x) - \log y, \quad x = (x_1, x_2) \in \mathbb{R}^2, y \in (0, 1]. \tag{4.115}$$

Thus, as in Section 4.6.5, the integral (4.113) 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

$$\arg \min_{(x_1,x_2,y)\in\bar{\mathcal{D}}_a} F(x_1, x_2, y) = (a, 1, 1), \tag{4.116}$$

where the last equality follows from Lemma 4.3.2 and [34, Lemma 3.2]. However, in this case, the boundary of the domain $\bar{\mathcal{D}}_a$ is not smooth at the point $(a, 1, 1)$, and so the same method, in particular Lemma 4.6.7, cannot be applied. Instead, in the next section, we use Proposition 4.6.6.

4.7.1 Proof of Theorem 4.3.10

We are now ready to prove Theorem 4.3.10 by combining the observations at the beginning of the section with Proposition 4.6.6, in a manner similar to the proof for spheres in Section 4.6.5.

Proof of Theorem 4.3.10. 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 $\bar{\mathcal{S}}^n$ can be expressed as in (4.113) and (4.114), and recall the assertion in (4.116) that the minimum of the function F in (4.115) on $\bar{\mathcal{D}}_a$ is attained at $(a, 1, 1)$. Thus, for any open neighborhood U of $(a, 1, 1)$, we split the probability into two parts. Fix $\theta \in \mathbb{S}$. Then

$$\mathbb{P}_\theta\left(\bar{\mathcal{S}}^n \in \bar{\mathcal{D}}_a\right) = \mathbb{P}_\theta\left(\bar{\mathcal{S}}^n \in \bar{\mathcal{D}}_a \cap U\right) + \mathbb{P}_\theta\left(\bar{\mathcal{S}}^n \in \bar{\mathcal{D}}_a \cap U^c\right). \tag{4.117}$$

For the first term in (4.117), we have the following estimate from (4.113) and (4.114):

$$\begin{aligned}\mathbb{P}_\theta\left(\bar{\mathcal{S}}^n \in \bar{\mathcal{D}}_a \cap U\right) &= \frac{n^2}{2\pi}\int_{\bar{\mathcal{D}}_a \cap U} \frac{1}{y}\bar{g}^n_\theta(x_1, x_2)e^{-nF(x_1,x_2,y)}dx_1dx_2dy, \\ &= \frac{n^2}{2\pi}\int_{\bar{\mathcal{D}}_a \cap U} \tilde{g}^n_\theta(x_1, x_2, y)e^{-nF(x_1,x_2,y)}dx_1dx_2dy, \end{aligned} \tag{4.118}$$

where

$$\tilde{g}^n_\theta(x_1, x_2, y) = \frac{1}{y}\bar{g}^n_\theta(x_1, x_2), \tag{4.119}$$

with $\bar{g}^n_\theta$ as defined in (4.85) and $F(x_1, x_2, y) = \Psi^*_p(x_1, x_2) - \log y$ as in (4.115).

The bulk of the proof is devoted to the asymptotics of the Laplace type integral in (4.118). In order to apply Proposition 4.6.6, we first do a change of variables to transform the domain of integration. Let $\mathfrak{T} : \bar{\mathcal{D}}_a \to \mathbb{R}^3$ be the mapping that takes (x_1, x_2, y) to $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$ such that

$$\mathcal{X} = x_1 y - a x_2^{1/p}, \quad \mathcal{Y} = 1 - y, \quad \mathcal{Z} = x_2 - 1. \tag{4.120}$$

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

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

and $\mathfrak{T}$ maps the minimizer $(a, 1, 1)$ of F to $(0, 0, 0)$. Hence, under the transformation $\mathfrak{T}$, setting $\tilde{U} := \mathfrak{T}(U)$, we write (4.118) as

$$\mathbb{P}_\theta\left(\bar{\mathcal{S}}^n \in \bar{\mathcal{D}}_a \cap U\right) = \frac{n^2}{2\pi} \int_{\widetilde{\mathcal{D}}_a \cap \widetilde{U}} \tilde{g}_\theta^n \circ \mathfrak{T}^{-1}(\mathcal{X}, \mathcal{Y}, \mathcal{Z}) e^{-nF \circ \mathfrak{T}^{-1}(\mathcal{X}, \mathcal{Y}, \mathcal{Z})} d\mathcal{X} d\mathcal{Y} d\mathcal{Z}. \tag{4.121}$$

Let $v_{ijk} := \partial_1^i \partial_2^j \partial_3^k F(a, 1, 1)$. Then, from (4.120), we have

$$\begin{aligned}
\frac{\partial F}{\partial \mathcal{X}}(0,0,0) &= v_{100} \frac{\partial x_1}{\partial \mathcal{X}} + v_{010} \frac{\partial x_2}{\partial \mathcal{X}} + v_{001} \frac{\partial y}{\partial \mathcal{X}}\bigg|_{(0,0,0)} \\
&= v_{100} \frac{1}{1 - \mathcal{Y}}\bigg|_{(0,0,0)} \\
&= v_{100}; \\
\frac{\partial F}{\partial \mathcal{Y}}(0,0,0) &= v_{100} \frac{\mathcal{X} + a(1 + \mathcal{Z})^{1/p}}{(1 - \mathcal{Y})^2} - v_{001}\bigg|_{(0,0,0)} \\
&= a v_{100} - v_{001}; \\
\frac{\partial F}{\partial \mathcal{Z}}(0,0,0) &= v_{100} \frac{a}{p} \frac{(1 + \mathcal{Z})^{(1-p)/p}}{1 - \mathcal{Y}} + v_{010}\bigg|_{(0,0,0)} \\
&= 0;
\end{aligned}$$

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

Combining (4.115) with the duality of the Legendre transform, simple calculations yield

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

We apply Lemma 4.6.6 to the transformed integral (4.121) with g^n, f, x^* and D therein replaced with $\tilde{g}^n_\theta \circ \mathfrak{T}^{-1}$, $F \circ \mathfrak{T}^{-1}$, $(0,0,0)$ and $\widetilde{\mathcal{D}}_a$. The assumptions of the Proposition then follows from Lemma 4.6.8. Moreover, F, defined in (4.115), and so $F \circ \mathfrak{T}^{-1}$, is smooth in a neighborhood of $(0,0,0)$. Expanding the integrand using (4.119), (4.86), (4.21) and the identity $\mathfrak{T}^{-1}(0,0,0) = (a,1,1)$, substituting the expressions for $\partial F/\partial \mathcal{X}$, $\partial F/\partial \mathcal{Y}$ and $\partial^2 F/\partial \mathcal{Z}^2$ obtained above, and recalling the definition of γ_a in (4.34), we obtain

$$
\begin{aligned}
\mathbb{P}_\theta\left(\bar{\mathcal{S}}^n \in \bar{\mathcal{D}}_a \cap U\right) &= \frac{n^2}{2\pi} \times \frac{\sqrt{2\pi}}{n^{5/2}}\frac{1}{\gamma_a}\tilde{g}^n_\theta \circ \mathfrak{T}^{-1}(0,0,0)e^{-nF\circ\mathfrak{T}^{-1}(0,0,0)}(1+o(1)) \\
&= \frac{n^2}{2\pi} \times \frac{\sqrt{2\pi}}{n^{5/2}}\frac{1}{\gamma_a}\bar{g}^n_\theta(a^*)e^{-n\Psi_p^*(a^*)}(1+o(1)) \\
&= \frac{C^n_a(\theta^n)}{\gamma_a\sqrt{2\pi n}}e^{-n\mathbb{I}_p(a)+\sqrt{n}R^n_a(\theta^n)}(1+o(1)). \qquad (4.122)
\end{aligned}
$$

For the second term in (4.117), as in the proof of ℓ_p^n spheres in (4.112), one can invoke the quenched large deviation principle for $\bar{\mathcal{S}}^n$ established in [34, Proposition 5.3] along with the fact that the rate function has a unique minimum, as proved in Lemma 4.3.2 to show that it is negligible with respect to (4.122). Thus, (4.42), (4.117) and (4.122), together, yield (4.35). $\square$

4.8 The joint density estimate

This section is devoted to the proof of the density estimate obtained in Proposition 4.6.4. As usual, throughout fix $p \in (1, \infty)$. In Section 4.8.1 an identity for the joint density is established in terms of an integral. This integral is then shown in Section 4.8.2 to admit an alternative representation as an expectation with respect to a tilted measure. The latter representation is used in Section 4.8.3 to obtain certain asymptotic estimates. These results are finally combined in Section 4.8.4 to prove Proposition 4.6.4.

4.8.1 An integral representation for the joint density

Lemma 4.8.1. *Fix $n \in \mathbb{N}$ and $\theta \in \mathbb{S}$, and recall the definitions of Ψ_p, $\mathbb{J}_p$, λ_x, Φ_p and $\Psi_{p,\theta}^n$ in* (4.15), (4.17), (4.19), (4.81) *and* (4.82)*, respectively, and recall that $\bar{h}_\theta^n$ is the density, under $\mathbb{P}_\theta$, of $\bar{S}^n$ defined in* (4.37)*. Then for all sufficiently large n, and $x \in \mathbb{J}_p$, the following identity holds,*

$$\bar{h}_\theta^n(x) = \left(\frac{n}{2\pi}\right)^2 e^{-n\Psi_p^*(x)} e^{n(\Psi_{p,\theta}^n(\lambda_x) - \Psi_p(\lambda_x))} \mathcal{I}_\theta^n(x), \tag{4.123}$$

where

$$\mathcal{I}_\theta^n(x) := \int_{\mathbb{R}^2} e^{-i\langle t, nx\rangle} \prod_{j=1}^n \frac{\Phi_p(\sqrt{n}\theta_j^n(\lambda_{x,1}+it_1), \lambda_{x,2}+it_2)}{\Phi_p(\sqrt{n}\theta_j^n\lambda_{x,1}, \lambda_{x,2})} dt. \tag{4.124}$$

Moreover, there exists $s > 1$ such that $(t_1, t_2) \mapsto (\prod_{j=1}^n \Phi_p(\sqrt{n}\theta_j^n(\lambda_1+it_1), \lambda_2+it_2))^{s/n}$ lies in $\mathbb{L}_1(\mathbb{R}^2)$ for all sufficiently large n.

Proof. Let $\mathbb{D}_p$ be as in (4.78), 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 (4.80) and for $\theta \in \mathbb{S}$, 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 $Y_1, \ldots, Y_n$ are i.i.d. with density f_p defined in (4.13) and the finiteness follows because $\lambda \in \mathbb{D}_p$ and thus $\lambda_2 < 1/p$. Then the Fourier transform of the integrable function $x \mapsto e^{\langle\lambda, x\rangle}\bar{l}_\theta^n(x)$ is given as follows: for $t \in \mathbb{R}^2$,

$$\begin{aligned}\int_{\mathbb{R}^2} e^{\langle\lambda+it, y\rangle}\bar{l}_\theta^n(y)dy &= \mathbb{E}_\theta\left[e^{\langle\lambda+it, \sum_{j=1}^n \bar{V}_j^n\rangle}\right] \\ &= \prod_{j=1}^n \mathbb{E}_\theta\left[e^{\langle\lambda+it, \bar{V}_j^n\rangle}\right] \\ &= \prod_{j=1}^n \Phi_p(\sqrt{n}\theta_j^n(\lambda_1+it_1), \lambda_2+it_2).\end{aligned} \tag{4.125}$$

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

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 $j, k \in \{1, \ldots, n\}$, $j \neq k$, we have*

$$K_\theta^{n,j,k}(\lambda, t) := \int_{\mathbb{R}^2} \left| \Phi_p(\sqrt{n}\theta_k^n(\lambda_1 + it_1), \lambda_2 + it_2)\Phi_p(\sqrt{n}\theta_j^n(\lambda_1 + it_1), \lambda_2 + it_2) \right|^s dt < \infty. \tag{4.126}$$

We defer the proof of the claim, first showing how the lemma follows from the claim. By Hölder's inequality, the claim and the boundedness of the moment generating function Φ_p, the product on the right-hand side of (4.125) lies in $\mathbb{L}_1(\mathbb{R}^2)$ for all sufficiently large n, which proves the second assertion of the lemma. We may then apply the inverse Fourier transform formula and obtain, for all sufficiently large n,

$$\bar{l}_\theta^n(x) = \left(\frac{1}{2\pi}\right)^2 \int_{\mathbb{R}^2} e^{-\langle \lambda + it, x\rangle} \prod_{j=1}^n \Phi_p(\sqrt{n}\theta_j^n(\lambda_1 + it_1), \lambda_2 + it_2) dt. \tag{4.127}$$

(Note that it is in precisely to ensure this application of the inverse Fourier transform formula that we have the stipulation in the lemma that n be sufficiently large.) Next, recall that for any $x \in \mathbb{J}_p$, $\lambda = \lambda_x$ is chosen so that (4.19) is satisfied. Also, by (4.37) and (4.80), we have

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

Hence, using (4.127), (4.19) and (4.82) we see that the density $\bar{h}_\theta^n$ of $\bar{S}^n$ under $\mathbb{P}_\theta$ is given by

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

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

for $x \in \mathbb{J}_p$. Since the right-hand side coincides with the definition of $\bar{h}_\theta^n$ given in (4.123) and (4.124), this proves the first part of the lemma given the claim. 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 let $\bar{v} = \bar{v}_\theta^{n,j,k}$ be as in the remark after the claim 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 [30, 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 (4.80) and (4.81), this is equivalent to saying that (4.126) holds with $s = 1 + 1/r > 0$.

To this end, we start by establishing an expression for $\bar{v}$. Note from (4.80) 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 $T := T^{n,j,k}$ is the transformation given by, for $(y_1, y_2) \in \mathbb{R}^2$,

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

Note from the definition that T is differentiable. Given $(z_1,z_2) \in \mathbb{R}^2$, we solve for $(z_1,z_2) = T(y_1,y_2)$. We see that z_2 should be positive for there to be a solution. On the other hand, 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

$\{y \in \mathbb{R}^2 : z_2 = \left|y_1\right|^p + \left|y_2\right|^p\}$. The solution of $(z_1, z_2) = T(y_1, y_2)$ is the intersection of the two curves above, which are a line and an ℓ_p^2 ball, respectively, Hence, there are 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 points such that $T(y^+) = T(y^-) = (z_1, z_2)$. Then by the change of variables formula and the differentiability of T, we may write the density $\bar{v}$ as

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

where $|\partial(y_1, y_2)/\partial(z_1, z_2)|$ is the Jacobian of the transformation T. A simple calculation yields

$$\left|\frac{\partial(y_1, y_2)}{\partial(z_1, z_2)}\right| = \frac{1}{\sqrt{n}p}\left|\frac{1}{\theta_2\mathrm{sgn}(y_1)\left|y_1\right|^{p-1} - \theta_1\mathrm{sgn}(y_2)\left|y_2\right|^{p-1}}\right|, \tag{4.128}$$

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

For $r > 0$, we see that

$$\begin{aligned} &\int_{\mathbb{R}^2}\left|e^{\lambda_1 z_1+\lambda_2 z_2}\bar{v}(z_1, z_2)\right|^{1+r}dz_1dz_2 \\ &\quad= \int_{\mathbb{R}^2}\left|e^{\lambda_1 z_1+\lambda_2 z_2 - z_2/p}\right|^{1+r}\left(\left|\frac{\partial(y_1^+, y_2^+)}{\partial(z_1, z_2)}\right| + \left|\frac{\partial(y_1^-, y_2^-)}{\partial(z_1, z_2)}\right|\right)^{1+r}1_{\{z_2>0,|z_1|<M(z_2)\}}dz_1dz_2 \\ &\quad\le 2^r\int_{\mathbb{R}^2}\left|e^{\lambda_1 z_1+\lambda_2 z_2 - z_2/p}\right|^{1+r}\left(\left|\frac{\partial(y_1^+, y_2^+)}{\partial(z_1, z_2)}\right|^{1+r} + \left|\frac{\partial(y_1^-, y_2^-)}{\partial(z_1, z_2)}\right|^{1+r}\right)1_{\{z_2>0,|z_1|<M(z_2)\}}dz_1dz_2 \\ &\quad= 2^r\int_{\mathbb{R}^2}\left|e^{\lambda_1\sqrt{n}(\theta_1 y_1+\theta_2 y_2)+\left(\lambda_2-\frac{1}{p}\right)\left(\left|y_1\right|^p+\left|y_2\right|^p\right)}\right|^{1+r}\left|\frac{\partial(y_1, y_2)}{\partial(z_1, z_2)}\right|^r dy_1dy_2,\end{aligned}$$

where the inequality follows from $(a+b)^{1+r} \le 2^r(a^{1+r}+b^{1+r})$ for $a, b \in \mathbb{R}_+$.

Let $\mathcal{N} \subset \mathbb{R}^2$ be a neighborhood of the origin. Then

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

Since $p \in (1, \infty)$ and $x \in \mathbb{J}_p$ implies $p\lambda_2 = p\lambda_{x,2} < 1$, it follows that $e^{\lambda_1 \sqrt{n}(\theta_1 y_1 + \theta_2 y_2) + \left(\lambda_2 - \frac{1}{p}\right)\left(|y_1|^p + |y_2|^p\right)}$ lies in $\mathbb{L}_r(\mathbb{R}^2)$ for any $r > 0$. Moreover, by (4.128), there exists $r_1 > 0$ small enough such that the Jacobian lies in $\mathbb{L}_{r_1}(\mathcal{N})$. On the other hand, there exists $0 < r_2 < \infty$ large enough such that the Jacobian 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.

□

4.8.2 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 (4.124) using a change of measure. From Section 4.3.4, 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 $\mathbb{I}_p(a) < \infty$, recall the definition of $\lambda = \lambda_a$ from (4.19). Fix $n \in \mathbb{N}$, consider a "tilted" measure $\widetilde{\mathbb{P}}^n = \widetilde{\mathbb{P}}^{n,a}$ on $(\Omega, \mathcal{F})$ such that the (marginal) distribution of Θ^n remains unchanged but conditioned on $\Theta = \theta \in \mathbb{S}$, $\{Y_j^n, j = 1, \ldots, n\}$ are still independent, but not identically distributed, with Y_j^n having density $\widetilde{f}_j^n = \widetilde{f}_{\theta,j}^{n,a}$ given by

$$\widetilde{f}_j^n(y) := \exp\left(\left\langle \lambda_a, \left(\sqrt{n}\theta_j^n y, |y|^p\right)\right\rangle - \Lambda_p\left(\sqrt{n}\theta_i^n \lambda_1, \lambda_2\right)\right) f_p(y), \quad y \in \mathbb{R}, \tag{4.129}$$

with Λ_p as defined in (4.14) and as before we omit the explicit dependence and other quantities of $\widetilde{f}_j^n$ on p and a. For $\theta \in \mathbb{S}$, denote by $\widetilde{\mathbb{P}}_\theta^n$ and $\widetilde{\mathbb{E}}_\theta^n$ the probability and the expectation taken with respect to $\widetilde{\mathbb{P}}^n$, conditioned on θ, and likewise, let $\widetilde{\mathrm{Var}}_\theta^n(\cdot)$ and $\widetilde{\mathrm{Cov}}_\theta^n(\cdot,\cdot)$ denote the conditional variance and conditional covariance, respectively, under $\widetilde{\mathbb{P}}_\theta^n$.

Recall from (4.14) and (4.81) that $\Lambda_p(t) = \log \Phi_p(t)$ for $t \in \mathbb{R}^2$. Then, by (4.80), (4.81) and (4.129), it follows that for $j = 1, \dots, n$ and $\beta = (\beta_1, \beta_2) \in \mathbb{R}^2$,

$$\widetilde{\mathbb{E}}_\theta^n\left[e^{\langle \beta, \bar{V}_j^n\rangle}\right] = \frac{\Phi_p(\sqrt{n}\theta_j^n(\beta_1+\lambda_1), \beta_2+\lambda_2)}{\Phi_p(\sqrt{n}\theta_j^n\lambda_1, \lambda_2)}, \tag{4.130}$$

and hence,

$$\widetilde{\mathbb{E}}_\theta^n\left[\bar{V}_j^n\right] = \nabla_\beta\, \widetilde{\mathbb{E}}_\theta^n\left[e^{\langle \beta, \bar{V}_j^n\rangle}\right]\Big|_{\beta=(0,0)} = \nabla \log \Phi_p(\sqrt{n}\theta_j^n\lambda_1, \lambda_2). \tag{4.131}$$

Denoting $\bar{V}_j^n = \left(\bar{V}_{j,1}^n, \bar{V}_{j,2}^n\right)$, by (4.130), we also have for $k, l = 1, 2$,

$$\begin{aligned}
\widetilde{\mathrm{Cov}}_\theta^n\left(\bar{V}_{j,k}^n, \bar{V}_{j,l}^n\right) &= \widetilde{\mathbb{E}}_\theta^n\left[\bar{V}_{j,k}^n \bar{V}_{j,l}^n\right] - \widetilde{\mathbb{E}}_\theta^n\left[\bar{V}_{j,k}^n\right]\widetilde{\mathbb{E}}_\theta^n\left[\bar{V}_{j,l}^n\right] \\
&= \partial^2_{\beta_k,\beta_l}\, \widetilde{\mathbb{E}}_\theta^n\left[e^{\langle \beta, \bar{V}_j^n\rangle}\right]\Big|_{\beta=(0,0)} - \widetilde{\mathbb{E}}_\theta^n\left[\bar{V}_{j,k}^n\right]\widetilde{\mathbb{E}}_\theta^n\left[\bar{V}_{j,l}^n\right] \\
&= \partial^2_{k,l} \log \Phi_p(\sqrt{n}\theta_j^n\lambda_1, \lambda_2).
\end{aligned} \tag{4.132}$$

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

$$\widehat{V}_x^n := \frac{1}{\sqrt{n}} \sum_{j=1}^n \left(\bar{V}_j^n - x\right). \tag{4.133}$$

Lemma 4.8.2. *For $x \in \mathbb{J}_p$ and $\theta \in \mathbb{S}$, recall the definitions of $\bar{V}_j^n$, Φ_p, c_x^n, $\mathcal{H}_x^n$ and $\widehat{V}_x^n$ given in (4.80) (4.81), (4.83) and (4.133). Then*

$$c_x^n(\theta^n) = \widetilde{\mathbb{E}}_\theta^n\left[\widehat{V}_x^n\right], \tag{4.134}$$

$$\langle \mathcal{H}_x^n(\theta^n)\, t, t\rangle = \widetilde{\mathrm{Var}}_\theta^n\left(\left\langle t, \widehat{V}_x^n\right\rangle\right), \quad \text{for all } t \in \mathbb{R}^2. \tag{4.135}$$

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

$$\hat{\mu}_{x,\theta}^n(t) := \widetilde{\mathbb{E}}_\theta^n\left[e^{i\left\langle t, \sqrt{n}\widehat{V}_x^n\right\rangle}\right] = e^{-i\langle t, nx\rangle} \prod_{j=1}^n \frac{\Phi_p(\sqrt{n}\theta_j^n(\lambda_{x,1} + it_1), \lambda_{x,2} + it_2)}{\Phi_p(\sqrt{n}\theta_j^n \lambda_{x,1}, \lambda_{x,2})}. \tag{4.136}$$

Furthermore, for σ-a.e. θ, as $n \to \infty$, $\mathcal{H}_x^n(\theta^n)$ converges to the quantity $\mathcal{H}_x$ defined in (4.22).

Proof. We fix $\theta \in \mathbb{S}$ and x in the domain $\mathbb{J}_p$ of Ψ_p^* defined in (4.17) and omit the subscript x from λ_x for notational simplicity. By (4.131), (4.133), the definition of $\Psi_{p,\theta}^n$ in (4.82) and (4.18), we have,

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

When combined with (4.83), this proves (4.134). Similarly, by the independence of $\bar{V}_j^n, j = 1, \ldots, n$, under $\widetilde{\mathbb{P}}_\theta^n$, (4.132), the definition of $\Psi_{p,\theta}^n$ in (4.82) and the definition of $\mathcal{H}_x^n$ in (4.83), it follows that

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

which proves (4.135). Also, by the definitions of $\hat{\mu}^n_{x,\theta}$ and $\widehat{V}^n_x$ in (4.136) and (4.133), respectively, the independence of $\bar{V}^n_j$, $j = 1, \dots, n$, under $\widetilde{\mathbb{P}}^n_\theta$ and the relation (4.130), it follows that for $t \in \mathbb{R}^2$,

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

which proves (4.136).

It only remains to establish the convergence stated in the last assertion of the lemma. By (4.83) and (4.82), it follows that for each $i, j = 1, 2$, there exists $\alpha, \beta \in \mathbb{N}$ such that the entry $(\mathcal{H}^n_x(\theta^n))_{ij}$ can be written as

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

Since, the moment generating function Φ_p is smooth, 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, ϕ has polynomial growth by Lemma 4.6.2. Since Lemma 4.4.1 implies that $\mathcal{W}_p\left(L^n_\theta, \gamma_2\right) \to 0$ as $n \to \infty$, it follows that

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

where from Lemma 4.1.2(2) that, as n tends to infinity, the last equality follows by the definition of $\mathcal{H}_x$ in (4.22). $\square$

4.8.3 Estimates of the integrand

Lemma 4.8.3. *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 (4.133) and (4.80), respectively. There exist constants $\widetilde{C} < \infty$ such that for all $n \in \mathbb{N}$ and for σ-a.e. θ,*

$$\frac{1}{n}\sum_{j=1}^{n} \widetilde{\mathbb{E}}_\theta^n\left[\left\|\bar{V}_j^n - \widetilde{\mathbb{E}}_\theta^n[\bar{V}_j^n]\right\|^3\right] < \widetilde{C}, \quad \frac{1}{n}\sum_{j=1}^{n} \widetilde{\mathbb{E}}_\theta^n\left[\left\|\bar{V}_j^n - \widetilde{\mathbb{E}}_\theta^n[\bar{V}_j^n]\right\|^4\right] < \widetilde{C}, \tag{4.137}$$

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

$$\widetilde{\mathbb{E}}_\theta^n\left[\left\|\widehat{V}_x^n - \widetilde{\mathbb{E}}_\theta^n[\widehat{V}_x^n]\right\|^3\right] < \widetilde{C}. \tag{4.138}$$

Proof. Due to the following standard inequalities, $(a^2+b^2)^{3/2} \le C'(|a|^3+|b|^3)$ and $\frac{1}{n}\sum_{j=1}^n |a_j|^3 \le (\frac{1}{n}\sum_{j=1}^n |a_j|^4)^{3/4}$, to show (4.137) it suffices to show the boundedness of

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

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

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n} \widetilde{\mathbb{E}}_\theta^n\left[\left(\bar{V}_{j,1}^n - \widetilde{\mathbb{E}}_\theta^n\left[\bar{V}_{j,1}^n\right]\right)^4\right] \\
&\quad = \frac{3}{n}\sum_{j=1}^{n} \widetilde{\mathbb{E}}_\theta^n\left[\left(\bar{V}_{j,1}^n - \widetilde{\mathbb{E}}_\theta^n\left[\bar{V}_{j,1}^n\right]\right)^2\right] + \int_{\mathbb{R}} \partial_1^4(\log \Phi_p(u\lambda_{x,1}, \lambda_{x,2}))L_\theta^n(du) \\
&\quad = 3\widetilde{\mathrm{Var}}_\theta^n\left(\widehat{V}_{x,1}^n\right) + \int_{\mathbb{R}} \partial_1^4(\log \Phi_p(u\lambda_{x,1}, \lambda_{x,2}))L_\theta^n(du).
\end{aligned} \tag{4.139}$$

Now, by (4.135), $\widetilde{\mathrm{Var}}_\theta^n(\widehat{V}_{x,1}^n) = (\mathcal{H}_x^n(\theta^n))_{11}$ and so by the last assertion of Lemma 4.8.2, for σ-a.e. θ, as $n \to \infty$, $\widetilde{\mathrm{Var}}_\theta^n(\widehat{V}_{x,1}^n)$ converges to $(\mathcal{H}_x)_{11}$. Also, since the function $\mathbb{R} \ni\mapsto$

$\partial_1^4(\log \Phi_p(u\lambda_{x,1}, \lambda_{x,2}))$ is continuous and has polynomial growth (the latter by Lemma 4.6.2), Lemma 4.4.1 and Lemma 4.1.2(2) together show that for σ-a.e. θ, the second term on the right-hand-side of (4.139) also has a finite limit as $n \to \infty$. Therefore, for σ-a.e. θ, the sum of the two terms is uniformly bounded.

Next, we deal with the second inequality. By (4.133), we have

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

By Jensen's inequality, we further obtain

$$\begin{aligned}
&\widetilde{\mathbb{E}}_\theta^n \left[\left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n \left(\bar{V}_j^n - \widetilde{\mathbb{E}}_\theta^n[\bar{V}_j^n] \right) \right\|^3 \right] \\
&\quad \le \left(\widetilde{\mathbb{E}}_\theta^n \left[\left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n \left(\bar{V}_j^n - \widetilde{\mathbb{E}}_\theta^n[\bar{V}_j^n] \right) \right\|^4 \right] \right)^{3/4} \\
&\quad \le \left(\frac{2}{n^2} \widetilde{\mathbb{E}}_\theta^n \left[\left(\sum_{j=1}^n \left(\bar{V}_{j,1}^n - \widetilde{\mathbb{E}}_\theta^n[\bar{V}_{j,1}^n] \right) \right)^4 \right] + \frac{2}{n^2} \widetilde{\mathbb{E}}_\theta^n \left[\left(\sum_{j=1}^n \left(\bar{V}_{j,2}^n - \widetilde{\mathbb{E}}_\theta^n[\bar{V}_{j,2}^n] \right) \right)^4 \right] \right)^{3/4}.
\end{aligned}$$

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

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

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

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

$$\leq \frac{1}{n^2}\sum_{j=1}^{n}\widetilde{\mathbb{E}}_\theta^n\left[\left(\bar{V}_{j,m}^n - \widetilde{\mathbb{E}}_\theta^n[\bar{V}_{j,m}^n]\right)^4\right] + 6\left(\frac{2}{n}\sum_{j=1}^{n}\widetilde{\mathbb{E}}_\theta^n\left[\left(\bar{V}_{j,m}^n - \widetilde{\mathbb{E}}_\theta^n[\bar{V}_{j,m}^n]\right)^4\right]\right)^2$$

which is bounded above by (4.137). This proves (4.138). □

Lemma 4.8.4. *Fix $x \in \mathbb{J}_p$ and recall the definitions of $\mathcal{H}_x$ Φ_p, c_x^n, $\mathcal{H}_x^n$, $\widehat{V}_x^n$ and $\hat{\mu}_{x,\theta}^n$ given in (4.22), (4.81), (4.83) (4.133) and (4.136), respectively. Then for σ-a.e. θ and every neighborhood $U \subset \mathbb{R}^2$ of the origin, there exist a neighborhood $\widetilde{U}$ of x and a constant $C \in (0,1)$ such that for all sufficiently large n,*

$$\sup_{t\in U^c}\left|\hat{\mu}_{y,\theta}^n(t)\right|^{1/n} < C, \quad y \in \widetilde{U}. \tag{4.140}$$

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

$$\left|\hat{\mu}_{y,\theta}^n\left(\frac{t}{\sqrt{n}}\right)e^{-itc_y^n(\theta^n)}\right| \leq \exp\left(-\frac{1}{2}\left\langle(\mathcal{H}_y - \varepsilon I)t, t\right\rangle\right), \quad y \in \widetilde{U}, \quad t \in U. \tag{4.141}$$

Proof. We omit the subscript x of λ_x for notational simplicity. Now, for $\theta \in \mathbb{S}$, and $t \in \mathbb{R}^2$, the relation (4.130) yields the inequality

$$\left|\frac{\Phi_p(\sqrt{n}\theta_j^n(\lambda_1 + it_1), \lambda_2 + it_2)}{\Phi_p(\sqrt{n}\theta_j^n\lambda_1, \lambda_2)}\right| = \left|\widetilde{\mathbb{E}}_\theta^n\left[e^{i\langle t, \bar{V}_j^n\rangle}\right]\right| \leq \widetilde{\mathbb{E}}_\theta^n\left[\left|e^{i\langle t, \bar{V}_j^n\rangle}\right|\right] \leq 1. \tag{4.142}$$

Noting from (4.81) 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 [30, Theorem 8.22] to obtain

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

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

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

Since Φ_p is a moment generating function which converges to 0 at infinity, Φ_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 $K^{-1} \leq \left|\sqrt{n}\theta_j^n\right| \leq K$ 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

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

Combining this with (4.136) yields the inequality

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

Since $\frac{1}{n}\sum_{j=1}^n 1_{\{K^{-1} \leq \left|\sqrt{n}\theta_j^n\right| \leq K\}} = L_\theta^n([K^{-1}, K] \setminus \{0\})$ whose limit, as $n \to \infty$, is dominated by $c_K := \gamma_2\left([K^{-1}, K]\right) > 0$ due to Lemma 4.4.1, we have for σ-a.e. θ,

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

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

$$\sup_{t \in U^c} \left|\hat{\mu}_{x,\theta}^n(t)\right|^{1/n} < C. \tag{4.143}$$

Since Φ_p is uniformly continuous in λ_x by definition and λ_x is a smooth function of x by the inverse function theorem applied to (4.19), we may choose a neighborhood $\widetilde{U}$ of

x such that for $y \in \widetilde{U}$,

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

i.e., for σ-a.e. θ and all sufficiently large n (possibly depending on θ), (4.140) holds.

Next, note that by (4.136) and (4.134), for $t \in \mathbb{R}^2$,

$$\hat{\mu}^n_{x,\theta}\left(\frac{t}{\sqrt{n}}\right) e^{-i\langle t, c^n_x(\theta^n)\rangle} = \widetilde{\mathbb{E}}^n_\theta\left[e^{i\left\langle t, \widehat{V}^n_x - \widetilde{\mathbb{E}}^n_\theta[\widehat{V}^n_x]\right\rangle}\right].$$

Thus, for $\theta \in \mathbb{S}$, by (4.135) and [25, Lemma 3.3.7], we have the following expansion:

$$\begin{aligned}\left| \hat{\mu}^n_{x,\theta}\left(\frac{t}{\sqrt{n}}\right) e^{-i\langle t, c^n_x(\theta^n)\rangle} - 1 + \frac{1}{2}\langle \mathcal{H}^n_x(\theta^n)t, t\rangle \right| &\le \widetilde{\mathbb{E}}^n_\theta\left[\left| \langle t, \widehat{V}^n_x - \widetilde{\mathbb{E}}^n_\theta[\widehat{V}^n_x]\rangle \right|^3 \right] \\ &\le \|t\|^3\, \widetilde{\mathbb{E}}^n_\theta\left[\left\| \widehat{V}^n_x - \widetilde{\mathbb{E}}^n_\theta[\widehat{V}^n_x] \right\|^3 \right].\end{aligned}$$

For $\varepsilon > 0$, by (4.138) of Lemma 4.8.3, 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 $\mathcal{H}^n_x(\theta^n)$ to $\mathcal{H}_x$ established in Lemma 4.8.2, for σ-a.e. θ, there exists $\varepsilon > 0$ such that $\mathcal{H}^n_x(\theta^n) - \varepsilon I$ is positive definite for all sufficiently large n (possibly depending on θ) and for $t \in U$,

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

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 (4.140), the uniformity of the bound in (4.141) follows from the definition in (4.83), (4.82) and the aforementioned uniform continuity of Φ_p in x. $\square$

4.8.4 Proof of the joint density estimate

We now combine the lemmas established in Sections 4.8.1–4.8.3 to prove the estimate for the density $\bar{h}^n_\theta$ of $\bar{S}^n$ obtained in Proposition 4.6.4.

Proof of Proposition 4.6.4. Fix $n \in \mathbb{N}$. Combining Lemma 4.8.1, (4.84) and (4.136) of Lemma 4.8.2, we see that for $x \in \mathbb{J}_p$ and σ-a.e. θ,

$$\bar{h}^n_\theta(x) = \frac{n}{2\pi} e^{-n\Psi^*_p(x)} e^{\sqrt{n}R^n_x(\theta^n)} \frac{n}{2\pi} \int_{\mathbb{R}^2} \hat{\mu}^n_{x,\theta}(t)dt. \tag{4.144}$$

When compared with (4.85) and (4.86), to prove the proposition, it suffices to show that

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

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

$$\int_{\mathbb{R}^2} \hat{\mu}^n_{x,\theta}(t)dt = \int_U \hat{\mu}^n_{x,\theta}(t)dt + \int_{U^c} \hat{\mu}^n_{x,\theta}(t)dt. \tag{4.145}$$

Now, by the estimate (4.140) in Lemma 4.8.4, we have for $C \in (0, 1)$ and $s > 1$,

$$\left|\int_{U^c} \hat{\mu}^n_{x,\theta}(t)dt\right| \le \int_{U^c} \left|\hat{\mu}^n_{x,\theta}(t)\right| dt \le C^{n-s} \int_{U^c} \left|\hat{\mu}^n_{x,\theta}(t)\right|^{s/n} dt. \tag{4.146}$$

From the definition of $\hat{\mu}^n_{x,\theta}$ in (4.136) and Lemma 4.8.1, we see that $|\hat{\mu}^n_{x,\theta}(t)|^{s/n}$ is integrable. Hence, the right hand side of (4.146) tends to zero exponentially fast as n tends to infinity. Moreover, the convergence is uniform in a neighborhood of x by (4.140) from Lemma 4.8.4.

Recall the definition of $\hat{\mu}^n_{x,\theta}$ in (4.136). By (4.133) and (4.134), the characteristic function of $\widehat{V}^n_x$ is given by $\hat{\mu}^n_{x,\theta}\left(\frac{t}{\sqrt{n}}\right)e^{-itc^n_x(\theta^n)}$. Since the sequence $(\widehat{V}^n_x)_{n\in\mathbb{N}}$ satisfies the Lyapunov-type condition stated in (4.137) of Lemma 4.8.3, 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

$$\hat{\mu}^n_{x,\theta}\left(\frac{t}{\sqrt{n}}\right)e^{-itc^n_x(\theta^n)} \to \exp\left(-\frac{1}{2}\langle \mathcal{H}_x t, t\rangle\right). \tag{4.147}$$

Now, by (4.141) of Lemma 4.8.4 and (4.147), we may apply the dominated convergence theorem, and use (4.147) to obtain for σ a.e. θ,

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

with $\mathcal{H}_x$ as in (4.22). Using standard properties of Gaussian integrals, this implies that

$$\int_U \hat{\mu}^n_{x,\theta}(t)dt = \frac{2\pi}{n}\det \mathcal{H}_x^{-1/2}\exp\left(\left\|\mathcal{H}_x^{-1/2}c^n_x(\theta^n)\right\|^2\right)(1+o(1)), \tag{4.148}$$

Combining (4.86), (4.144), (4.145), (4.148) and the estimate of the integral over U^c in (4.146), we conclude that the asymptotic expansion for the density $\bar{h}^n_\theta(x)$ given in (4.85) holds uniformly for x in any compact subset of $\mathbb{J}_p$.

Finally, by the definition of λ_x in (4.19) and the inverse function theorem, the mapping $x \mapsto \lambda_x$ is smooth. Therefore, combining (4.86), (4.22), (4.83) and (4.84), we conclude $\bar{g}^n_\theta$ is smooth. $\square$

4.9 Infimum of the rate function

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

Proof of Lemma 4.3.2. Recall from (4.16) and (4.20), that we have the following expression for the rate function: for $t \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{I}_p(t) &= \inf_{\tau_1 \in \mathbb{R}, \tau_2 > 0 : \tau_1 \tau_2^{-1/p} = t} \Psi_p^*(\tau_1, \tau_2) \\
&= \inf_{\tau_1 \in \mathbb{R}, \widetilde{\tau}_2 > 0 : \tau_1 \widetilde{\tau}_2^{-1} = t} \Psi_p^*(\tau_1, \widetilde{\tau}_2^p) \\
&= \inf_{\widetilde{\tau}_2 > 0} \Psi_p^*(\widetilde{\tau}_2 t, \widetilde{\tau}_2^p), \qquad (4.149)
\end{aligned}$$

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

By Lemmas 5.8 and 5.9 of [33], Ψ_p is essentially smooth, convex and lower semicontinuous; see Definition 2.3.5 of [21] 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.,

$$\Psi_p^*(\tau t, \tau^p) = s_1 \tau t + s_2 \tau^p - \Psi_p(s_1, s_2), \qquad (4.150)$$

where, by (4.15), $\Psi_p(s_1, s_2) = \int \Lambda_p(u s_1, s_2) \gamma_2(du)$, with γ_2 being the standard Gaussian measure and Λ_p defined as in (4.14). Note that s_1, s_2 satisfy the following first order conditions:

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

where ∂_i represents the partial derivative with respect to s_i, for $i = 1, 2$. From [33, Lemma 5.9], we can exchange the order of differentiation and integration to obtain

$$\begin{aligned} \partial_1 \Psi_p(s_1, s_2) &= \int_{\mathbb{R}} u \partial_1 \Lambda_p(us_1, s_2) \gamma_2(du), \\ \partial_2 \Psi_p(s_1, s_2) &= \int_{\mathbb{R}} \partial_2 \Lambda_p(us_1, s_2) \gamma_2(du). \end{aligned} \tag{4.151}$$

To calculate these integrals, we first recall the expression for Λ_p established in [33, Lemma 5.7],

$$\Lambda_p(s_1, s_2) = -\frac{1}{p} \log(1 - ps_2) + \log M_{\gamma_p}\left(\frac{s_1}{(1 - ps_2)^{1/p}}\right), \tag{4.152}$$

where M_{γ_p} denotes the moment generating function of the measure γ_p with density defined in (4.13). Differentiation yields

$$\begin{aligned} \partial_1 \Lambda_p(us_1, s_2) &= \frac{M'_{\gamma_p}\left(\frac{us_1}{(1-ps_2)^{1/p}}\right)}{M_{\gamma_p}\left(\frac{us_1}{(1-ps_2)^{1/p}}\right)} \frac{1}{(1 - ps_2)^{1/p}}, \\ \partial_2 \Lambda_p(us_1, s_2) &= \frac{1}{1 - ps_2} + \frac{M'_{\gamma_p}\left(\frac{us_1}{(1-ps_2)^{1/p}}\right)}{M_{\gamma_p}\left(\frac{us_1}{(1-ps_2)^{1/p}}\right)} \frac{us_1}{(1 - ps_2)^{(p+1)/p}}. \end{aligned} \tag{4.153}$$

Combining all the above relations, we obtain

$$\tau t = \int_{\mathbb{R}} \frac{M'_{\gamma_p}\left(\frac{us_1}{(1-ps_2)^{1/p}}\right)}{M_{\gamma_p}\left(\frac{us_1}{(1-ps_2)^{1/p}}\right)} \frac{u}{(1 - ps_2)^{1/p}} \gamma_2(du), \tag{4.154}$$

$$\begin{aligned} \tau^p &= \int_{\mathbb{R}} \left(\frac{1}{1 - ps_2} + \frac{M'_{\gamma_p}\left(\frac{us_1}{(1-ps_2)^{1/p}}\right)}{M_{\gamma_p}\left(\frac{us_1}{(1-ps_2)^{1/p}}\right)} \frac{us_1}{(1 - ps_2)^{(p+1)/p}} \right) \gamma_2(du) \\ &= \frac{1}{1 - ps_2} + \frac{\tau t s_1}{1 - ps_2}, \end{aligned} \tag{4.155}$$

and note that (4.155) implies

$$\tau^p p s_2 + \tau t s_1 = \tau^p - 1. \tag{4.156}$$

Now, in view of (4.149), to compute $\mathbb{I}_p(t)$ we have to first take the derivative of $\Psi_p^*(\tau t, \tau^p)$ with respect to τ and set it to 0. Note that in the following, s_1, s_2 are functions of τ and t satisfying (4.154) and (4.155). Using (4.15) and (4.149), we first rewrite $\Psi_p(s_1, s_2)$ as

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

From equations (4.150)-(4.156), we obtain

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

Setting the derivative computed above to 0, we conclude that the minimum over $\tau > 0$ in (4.149) 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 (4.16), proves Lemma 4.3.2. $\square$

4.10 Proof of the Central Limit Theorem for the empirical measure

Proof of Lemma 4.4.2. Let $(Z_j^n, j = 1, \ldots, n)_{n\in\mathbb{N}}$ be independent standard Gaussian random variables. Then note that (e.g. see Section 4.3.4 or [74, Lemma 1])

$$\Theta_j^n \overset{(d)}{=} \frac{Z_j^n}{\|Z^n\|}, \tag{4.157}$$

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

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

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

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

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{j=1}^n \left[F\left(\sqrt{n}\frac{Z_j^n}{\|Z^n\|}\right) - \mathbb{E}\left[F(Z)\right]\right] \\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^n \left[F(Z_j^n) - \mathbb{E}[F(Z)] + F'(Z_j^n)\left(\frac{\sqrt{n}Z_j^n}{\|Z^n\|} - Z_j^n\right) + \frac{F''(Z_j^n)}{2}\left(\frac{\sqrt{n}Z_j^n}{\|Z^n\|} - Z_j^n\right)^2\right. \\
&\qquad\qquad \left. + \frac{F'''(\widetilde{Z}_j^n)}{6}\left(\frac{\sqrt{n}Z_j^n}{\|Z^n\|} - Z_j^n\right)^3\right],
\end{aligned}$$

$$= \hat{r}_n(F) + \frac{1}{\sqrt{n}}\hat{s}_n(F) + \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{F'''(\widetilde{Z}_j^n)}{6}\left(\frac{\sqrt{n}Z_j^n}{\|Z^n\|} - Z_j^n\right)^3 \tag{4.158}$$

where $\hat{r}_n(\cdot)$ and $\hat{s}_n(\cdot)$ are defined in (4.49) and (4.48), respectively, and $\widetilde{Z}_i^n \in \mathbb{R}$ 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 (4.158) 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 $n \in \mathbb{N}$,

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

Since $\widetilde{Z}_j^n$ 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 $N = N(w)$ such that a.s. for all $n > N$,

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

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

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

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

From the Gaussian concentration inequality (see [78, 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

$$\begin{aligned} \mathbb{P}\left(\frac{1}{\sqrt{n}} \left| \|Z^n\| - \sqrt{n} \right|^3 > \epsilon \right) &= \mathbb{P}\left(\left| \|Z^n\| - \sqrt{n} \right| > n^{1/6} \epsilon^{1/3} \right) \\ &\leq 2e^{-c\epsilon^{2/3} n^{1/3}} \\ &\to 0, \quad \text{as} \quad n \to \infty. \end{aligned} \tag{4.159}$$

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

$$\frac{1}{n} \sum_{j=1}^{n} \left| Z_j^n \right|^3 (1 + \left| Z_j^n \right|^q) \to \mathbb{E}\left[|Z|^3 (1 + |Z|^q) \right]. \tag{4.160}$$

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

$$\frac{\|Z^n\|}{\sqrt{n}} \to 1, \quad \text{a.s.} \tag{4.161}$$

Together, (4.159), (4.160) and (4.161) show that

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

We may then rewrite (4.158) as follows:

$$\frac{1}{\sqrt{n}}\left(\sum_{j=1}^{n} F\left(\sqrt{n}\frac{Z_j^n}{\|Z^n\|}\right) - \mathbb{E}\left[F(Z)\right]\right) = \hat{r}_n(F) + \frac{1}{\sqrt{n}}\hat{s}_n(F) + o\left(\frac{1}{\sqrt{n}}\right). \tag{4.162}$$

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:

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

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

$$\left(\widetilde{\Sigma}\right)_{ij} := \operatorname{Cov}(M_i, M_j), \quad \text{for} \quad i, j = 1, \ldots, 6, \tag{4.163}$$

where

$$(M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_8) := \left(F(Z), F'(Z)Z, F''(Z)Z^2, Z^2, G_1(Z), G_1'(Z)Z, G_2(Z), G_2'(Z)Z\right).$$

By the Skorokhod representation theorem, we can find $(\widetilde{\mathfrak{A}}_n, \widetilde{\mathfrak{B}}_n, \widetilde{\mathfrak{C}}_n, \widetilde{\mathfrak{D}}_n, \widetilde{\mathfrak{E}}_n, \widetilde{\mathfrak{F}}_n, \widetilde{\mathfrak{G}}_n, \widetilde{\mathfrak{H}}_n)$

and $\widetilde{M} := (\widetilde{\mathfrak{A}}, \widetilde{\mathfrak{B}}, \widetilde{\mathfrak{C}}, \widetilde{\mathfrak{D}}, \widetilde{\mathfrak{E}}, \widetilde{\mathfrak{F}}, \widetilde{\mathfrak{G}}, \widetilde{\mathfrak{H}})$ all defined on some common probability space, such that

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

and

$$(\widetilde{\mathfrak{A}}_n, \widetilde{\mathfrak{B}}_n, \widetilde{\mathfrak{C}}_n, \widetilde{\mathfrak{D}}_n, \widetilde{\mathfrak{E}}_n, \widetilde{\mathfrak{F}}_n, \widetilde{\mathfrak{G}}_n, \widetilde{\mathfrak{H}}_n) \to \widetilde{M} \text{ a.s.} \tag{4.164}$$

Now, we substitute $(\widetilde{\mathfrak{A}}_n, \widetilde{\mathfrak{B}}_n, \widetilde{\mathfrak{C}}_n, \widetilde{\mathfrak{D}}_n, \widetilde{\mathfrak{E}}_n, \widetilde{\mathfrak{F}}_n, \widetilde{\mathfrak{G}}_n, \widetilde{\mathfrak{H}}_n)$ into (4.162), and we first take care of r_n

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

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

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

Since $\widetilde{\mathfrak{B}}_n/\sqrt{n}$ and $\widetilde{\mathfrak{D}}_n/\sqrt{n}$ converge to 0 almost surely by (4.164), we consider the Taylor expansion of H_1 at $(0, 0)$:

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

$$= -\frac{y}{2}\mathbb{E}[F'(Z)Z] + O(x^2 + y^2).$$

Combining the last three displays, we obtain

$$\begin{aligned}\hat{r}_n(F) &\overset{(d)}{=} \widetilde{\mathfrak{A}}_n + \sqrt{n}\left(-\frac{\widetilde{\mathfrak{D}}_n}{2\sqrt{n}}\mathbb{E}[F'(Z)Z] + O\left(\frac{\widetilde{\mathfrak{B}}_n^2}{n} + \frac{\widetilde{\mathfrak{D}}_n^2}{n}\right)\right)\\ &= \widetilde{\mathfrak{A}}_n - \frac{1}{2}\mathbb{E}[F'(Z)Z]\widetilde{\mathfrak{D}}_n + \mathbb{E}[F'(Z)Z]O\left(\frac{\widetilde{\mathfrak{B}}_n^2}{\sqrt{n}} + \frac{\widetilde{\mathfrak{D}}_n^2}{\sqrt{n}}\right).\end{aligned}$$

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

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

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

$$\hat{r}_n(F) \Rightarrow \widetilde{\mathfrak{A}} - \frac{1}{2}\mathbb{E}[F'(Z)Z]\widetilde{\mathfrak{D}}, \tag{4.165}$$

as $n \to \infty$.

Similarly, for s_n we have

$$\begin{aligned}\hat{s}_n(F) &= \frac{1}{2}\sum_{j=1}^{n} F''(Z_j^n)(Z_j^n)^2\left(\frac{\sqrt{n}}{\|Z^n\|} - 1\right)^2\\ &= \frac{1}{2}n\left(\mathbb{E}[F''(Z)Z^2] + \frac{\mathfrak{C}_n}{\sqrt{n}}\right)\left(\frac{1}{(1 + \mathfrak{D}_n/\sqrt{n})^{1/2}} - 1\right)^2\\ &\overset{(d)}{=} \frac{1}{2}nH_2\left(\frac{\widetilde{\mathfrak{C}}_n}{\sqrt{n}}, \frac{\widetilde{\mathfrak{D}}_n}{\sqrt{n}}\right),\end{aligned}$$

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

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

Note that $\widetilde{\mathfrak{C}}_n/\sqrt{n}$ and $\widetilde{\mathfrak{D}}_n/\sqrt{n}$ converge to 0 almost surely by (4.164). 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

$$\begin{aligned} \hat{s}_n(F) &\overset{(d)}{=} \frac{1}{8}\mathbb{E}[F''(Z)Z^2]\widetilde{\mathfrak{D}}_n^2 + O\left(\frac{\widetilde{\mathfrak{C}}_n^3}{\sqrt{n}} + \frac{\widetilde{\mathfrak{D}}_n^3}{\sqrt{n}}\right) \\ &\Rightarrow \frac{1}{8}\mathbb{E}[F''(Z)Z^2]\widetilde{\mathfrak{D}}^2, \end{aligned} \tag{4.166}$$

as n tends to infinity, which holds since $\widetilde{\mathfrak{D}}_n \to \widetilde{\mathfrak{D}}$ 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

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

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:

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

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

$$= \hat{r}_n(G_i) + o(1). \tag{4.167}$$

Thus, the expansion in Lemma 4.4.2 follows from (4.157), (4.165) (4.166) and (4.167). The second assertion of the lemma is a consequence of (4.165), (4.166), the analog of (4.165) with F replaced with G_i and the joint convergence of $(\widetilde{\mathfrak{A}}_n, \widetilde{\mathfrak{D}}_n, \widetilde{\mathfrak{E}}_n) \Rightarrow (\widetilde{\mathfrak{A}}, \widetilde{\mathfrak{D}}, \widetilde{\mathfrak{E}})$ and $(\widetilde{\mathfrak{A}}_n, \widetilde{\mathfrak{D}}_n, \widetilde{\mathfrak{G}}_n) \Rightarrow (\widetilde{\mathfrak{A}}, \widetilde{\mathfrak{D}}, \widetilde{\mathfrak{G}})$. □

4.11 Proof of Lemma 4.6.2

Proof. For $p = 2$, γ_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-th Gaussian random variable with density as in (4.13). The moments of Y are given in [64] by

$$\mathbb{E}\left[Y^m\right] = \begin{cases} 0, & \text{m odd}, \\ \frac{p^{m/p}\Gamma\left(\frac{m+1}{p}\right)}{\Gamma\left(\frac{1}{p}\right)}, & \text{m even}. \end{cases} \tag{4.168}$$

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

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

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,

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

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

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

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

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

where the second to last inequality is due to Wendel [82, 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 (4.168) we have for $t \in \mathbb{R}$,

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

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 $m \in \mathbb{N} \cup \{0\}$

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

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

$$(2m+1)\left(1-\frac{1}{p}\right) - 1 < a < (2m+1)\left(1-\frac{1}{p}\right);$$

$$(2m+2n)\left(1-\frac{1}{p}\right)-\frac{3}{p}<b<(2m+2n)\left(1-\frac{1}{p}\right)-\frac{3}{p}+1. \tag{4.169}$$

Then we have the inequality

$$2n-\frac{2n}{p}-1-\frac{2}{p}<b-a<2n-\frac{2n}{p}+1-\frac{2}{p}. \tag{4.170}$$

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

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

where the inequality follows from (4.169) and (4.170). We further see that

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

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

4.12 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 (4.28) that the leading order term that depends on θ is $R_a^n(\theta^n)$, which, in turn, depends on θ only through $\Psi_{p,\theta}^n(\lambda_a)$, as evident from its definition in (4.84). From the definitions in (4.14), (4.19)and (4.82), we have

$$\Psi_{p,\theta}^n(\lambda_a) = \frac{1}{n} \sum_{j=1}^{n} \Lambda_p \left(\sqrt{n} \theta_j^n \lambda_{a,1}, \lambda_{a,2} \right), \tag{4.171}$$

where we suppress the θ^n dependence in $\Psi_{p,\theta}^n$. We first state a lemma regarding the properties of Λ_p in [33].

Lemma 4.12.1. [33, 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 4.3.7. From the definition of $R_a^n(\theta^n)$ in (4.84), it suffices to understand the behavior of (4.171). Since by (4.19), $\lambda_{a,2} < 1/p$. We may apply Lemma 4.12.1 in the following proof.

First, for $p = 2$, from (4.14) and (4.77), 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 (4.171) and the last display, $\Psi_{p,\theta}^n(\lambda_a)$ does not depend on θ and thus is a constant.

Next, consider $p > 2$. By 4.12.1, $\Lambda_p\left(\sqrt{\cdot}, \lambda_{a,2}\right)$ is concave but not linear. By the definition of Λ_p in (4.14) and the symmetry of the p-th Gaussian distribution (4.13), $\Lambda_p(\cdot, \lambda_{a,2})$ is an even function. Therefore, for $\theta^n \in \mathbb{S}^{n-1}$,

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

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

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

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

On the other hand, to identify the minimizers of $\theta^n \mapsto \Psi^n_{p,\theta}(\lambda_a)$, note from Lemma 4.12.1 and the fact that $\Lambda_p(\cdot, \lambda_{a,2})$ is even, we can write $\Psi^n_{p,\theta}(\lambda_a) = \mathcal{F}(\theta_1^n, \ldots, \theta_n^n)$, where $\mathcal{F} = \mathcal{F}_a$ is defined to be

$$\mathcal{F}(t_1, t_2, \ldots, t_n) := \frac{1}{n}\sum_{j=1}^{n} \Lambda_p\left(\sqrt{n t_j (\lambda_{a,1})^2}, \lambda_{a,2}\right). \tag{4.172}$$

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

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

Since $\mathcal{F}$ is strictly concave by Lemma 4.12.1, the minimum of $\mathcal{F}$ is obtained at the extreme points of $\mathcal{A}$, namely, the vectors, $\pm e_j$, $j = 1, \ldots, n$. Thus, by (4.172), the minimum of $\Psi^n_{p,\theta}(\lambda_a)$ is also attained at

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

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 4.12.1. □

4.13 Proof of Proposition 4.6.6

Proof. For the smooth function f, for any multi-index $\alpha = (\alpha_1, \dots, \alpha_d)$, we define $f_\alpha := \partial^{|\alpha|}_{\alpha_1,\dots,\alpha_d} f(x^*)$. By setting $\tilde{f}(x) = f(x) - f(x^*)$, without loss of generality, we assume $f(x^*) = 0$. Note that $\tilde{D}$ contains a neighborhood of $x^* = (0, 0, \dots, 0)$, which equals the origin, and $D = \tilde{D} \cap (\mathbb{R}^m \times \mathbb{R}^d_+)$. Since f is smooth in D and its minimum over $\mathrm{cl}(D)$ 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 smooth real-valued functions on D. We will proceed by making several changes of variables. We start with the transformation $T_1 : \mathbb{R}^m \times \mathbb{R}^d_+ \to \mathbb{R}^m \times \mathbb{R}^d_+$ 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 of this transformation 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

$$F(u) := f(T_1^{-1}(u)) = \sum_{i=1}^{m} \frac{f_{i,i}}{2} u_i^2 + \sum_{i=m+1}^{m+d} f_i u_i. \tag{4.173}$$

Let m^* be the maximum of F in D' or equivalently, of f in D. Next, consider an additional change of variables $T_2 :: \mathbb{R}^m \times \mathbb{R}_+^d \to \mathbb{R}_+ \times [\pi/2, \pi, 2]^m \times [0, \pi/2]^{d-1}$ by letting $(\xi, \theta, \phi) = T_2(u)$ be such that

$$\begin{aligned}
u_i &= \left(\frac{2\xi}{f_{i,i}}\right)^{1/2} \cos\theta_1 \cdots \cos\theta_{i-1} \sin\theta_i, & i &= 1, \dots, m, \\
u_{m+i} &= \frac{\xi}{f_{m+i}} \cos^2\theta_1 \cdots \cos^2\theta_m \cos^2\phi_1 \cdots \cos^2\phi_{i-1} \sin^2\phi_i, & i &= 1, \dots, d-1, \\
u_{m+d} &= \frac{\xi}{f_{m+d}} \cos^2\theta_1 \cdots \cos^2\theta_m \cos^2\phi_1 \cdots \cos^2\phi_{d-2} \cos^2\phi_{d-1}, & & \qquad (4.174)
\end{aligned}$$

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 ϕ_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 (4.173) and (4.174), $F(u) = \xi$ for $u \in \mathbb{R}^{m+d}$ and we have the Jacobian

$$\begin{aligned}
J &:= \frac{\partial(u_1, \cdots, u_{m+d})}{\partial(\xi, \theta_1, \cdots, \theta_m, \phi_1, \cdots, \phi_{d-1})} & (4.175) \\
&= \frac{(2\xi)^{m/2+d-1}}{f_{m+1} \cdots f_{m+d} \sqrt{f_{1,1} \cdots f_{m,m}}} \prod_{i=1}^{m} \cos^{2d+m-1-i}\theta_i \prod_{i=1}^{d-1} \cos^{2d-1-2i}\phi_i \sin\phi_i. & (4.176)
\end{aligned}$$

Next, define $G^n(u) := g^n(T_1^{-1}(u))$. Using the change of variables T_1, we see that

$$I^n := \int_D h^n(x)dx = \int_{D'} G^n(u) e^{-nF(u)} du. \tag{4.177}$$

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 $t \in [0, m^*]$, i.e., $D' \subset \cup_{t\in[0,m^*]}\{u \in \mathbb{R}^{k+1} : F(u) = t\}$, Note also that D' is bounded and ∇F is nonzero in D'. Using the method of resolution of multiple integrals [83, Theorem 9, Chapter V], we then have

$$I^n = \int_0^{m^*} \mathrm{r}^n(t) e^{-nt} dt, \tag{4.178}$$

where

$$\mathrm{r}^n(t) := \int_{\{F(u)=t\}} \frac{G^n(u)}{\sqrt{\sum_{i=1}^{k+1} F_{u_i}^2}} dA, \quad t \in [0, m^*], \tag{4.179}$$

with dA denoting the surface element of the surface $F(u) = t$. Then, recalling the second change of variables in (4.174), we have

$$\frac{dA}{||\nabla F||} = J d\theta d\phi.$$

If $\hat{r}^n(\xi, \theta, \phi)$ denotes the transformation of r^n under $T_2 \circ T_1$, then,

$$\mathrm{r}^n(t) = \frac{(2t)^{m/2+d-1}}{f_{m+1} \cdots f_{m+d} \sqrt{f_{1,1} \cdots f_{m,m}}} \tag{4.180}$$

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

$$= \frac{(2t)^{m/2+d-1} e^{\tilde{r}^n(t)}}{f_{m+1} \cdots f_{m+d} \sqrt{f_{1,1} \cdots f_{m,m}}} \prod_{i=1}^{m} \int_{-\pi/2}^{\pi/2} \cos^{2d+m-1-i}\theta d\theta \prod_{i=1}^{d-1} \int_0^{2\pi} \cos^{2d-1-2i}\phi \sin\phi d\phi, \tag{4.181}$$

for some function $\tilde{r}^n(t)$ that satisfies for some $\varepsilon > 0$, $|\tilde{r}^n(t)| \le Cn^\alpha t$ for n large and $t \in (0, \varepsilon)$ because $|r^n(x)| \le Cn^\alpha ||x||$ for n large and x in a neighborhood of the origin. From (4.178) and (4.181), we have

$$I^n = \frac{2^{m/2+d-1}}{f_{m+1}\cdots f_{m+d}\sqrt{f_{1,1}\cdots f_{m,m}}}\prod_{i=1}^{m}\int_{-\pi/2}^{\pi/2}\cos^{2d+m-1-i}\theta d\theta\prod_{i=1}^{d-1}\int_0^{2\pi}\cos^{2d-1-2i}\phi\sin\phi d\phi$$
$$\times\int_0^{m^*}t^{m/2+d-1}e^{-nt+\tilde{r}^n(t)}dt.$$

Now, we apply Theorem 2.1 from [66, Chapter 9], with p, r, q, λ, μ, and ν being t, $\tilde{r}^n$, $t^{m/2+d-1}$, $m/2+d$, 1 and 1 and obtain

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

□

4.14 A uniform deviation estimate

We now establish Lemma 4.6.9. 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 4.6.9. Fix $\varepsilon > 0$. We first apply Taylor's theorem to the function $\mathcal{K}$. For $x \in \mathbb{R}$ and $(t_1, t_2) \in \mathbb{D}$, note that $(xt_1, t_2) \in \mathbb{D}$, and thus

$$\mathcal{K}(xt_1, t_2) = \mathcal{K}(0, t_2) + \partial_1\mathcal{K}(\rho(xt_1, t_2), t_2)xt_1, \tag{4.182}$$

where $\rho(\cdot,\cdot)$ is a function with $\left|\rho(xt_1,t_2)\right| \le |xt_1|$. By the polynomial growth assumption, there exist $q, \tilde{C} \in (0,\infty)$ such that

$$\sup_{\|t\|<\varepsilon} |\partial_1 \mathcal{K}(xt_1,t_2)| \le \tilde{C}(1+\varepsilon^q |x|^q), \tag{4.183}$$

$$\sup_{\|t\|<\varepsilon} \left|\partial_{1j} \mathcal{K}(xt_1,t_2)\right| \le \tilde{C}(1+\varepsilon^q |x|^q), \quad \text{for} \quad j=1,2. \tag{4.184}$$

By (4.182), for $\|t\|<\varepsilon$,

$$\sum_{j=1}^n \left(\mathcal{K}\left(\frac{\sqrt{n} Z_j}{\left\| Z^{(n)} \right\|} t_1, t_2 \right) - \mathbb{E}\left[\mathcal{K}(Zt_1,t_2) \right] \right) = t_1 \left(I_1^n(t) + I_2^n(t) \right), \tag{4.185}$$

where

$$I_1^n(t) := \sum_{j=1}^n \left(\partial_1 \mathcal{K}\left(\rho\left(\frac{\sqrt{n} Z_j}{\left\| Z^{(n)} \right\|} t_1, t_2 \right), t_2 \right) \frac{\sqrt{n} Z_j}{\left\| Z^{(n)} \right\|} - \partial_1 \mathcal{K}\left(\rho\left(Z_j t_1, t_2 \right), t_2 \right) Z_j \right), \tag{4.186}$$

$$I_2^n(t) := \sum_{j=1}^n \left(\partial_1 \mathcal{K}\left(\rho\left(Z_j t_1, t_2 \right), t_2 \right) Z_j - \mathbb{E}\left[\partial_1 \mathcal{K}(\rho(Zt_1,t_2),t_2) Z \right] \right). \tag{4.187}$$

The proof follows in several steps.

Step 1. We will use the Gaussian concentration inequality to establish a bound on $I_1^n(t)$. Specifically we claim that there exist $C_1 \in (0,\infty)$ and a random integer $N_1 \in \mathbb{N}$ such that $\mathbb{P}'$-almost surely, for $n \ge N_1$,

$$\sup_{\|t\|<\varepsilon} \left| n^{-\alpha} I_1^n(t) \right| \le C_1. \tag{4.188}$$

To prove (4.188), note from (4.182) that for $x \in \mathbb{R}$, $\|t\| < \varepsilon$,

$$\frac{d}{dx}\left(\partial_1 \mathcal{K}(\rho(xt_1, t_2), t_2)x\right) = \begin{cases} \dfrac{d}{dx}\left(\dfrac{\mathcal{K}(xt_1, t_2) - \mathcal{K}(0, t_2)}{t_1}\right) = \partial_1 \mathcal{K}(xt_1, t_2), & \text{if } t_1 \neq 0, \\ \partial_1 \mathcal{K}(0, t_2), & \text{if } t_1 = 0. \end{cases}$$

Together with (4.183), this implies that for $x \in \mathbb{R}$,

$$\sup_{\|t\|<\varepsilon} \left| \frac{d}{dx}\left(\partial_1 \mathcal{K}(\rho(xt_1, t_2), t_2)x\right) \right| = \sup_{\|t\|<\varepsilon} |\partial_1 \mathcal{K}(xt_1, t_2)| \leq \tilde{C}(1 + \varepsilon^q |x|^q).$$

Therefore, by the mean value theorem, there exists $\tilde{Z}_j \in \mathbb{R}$ lying between Z_j and $\sqrt{n} Z_j / \left\|Z^{(n)}\right\|$ such

$$\left| \partial_1 \mathcal{K}\left(\frac{\sqrt{n} Z_j}{\left\|Z^{(n)}\right\|} t_1, t_2\right) \frac{\sqrt{n} Z_j}{\left\|Z^{(n)}\right\|} - \partial_1 \mathcal{K}\left(Z_j t_1, t_2\right) Z_j \right| = \left| \partial_1 \mathcal{K}(\tilde{Z}_j^n t_1, t_2) \right| \left| \frac{\sqrt{n} Z_j}{\left\|Z^{(n)}\right\|} - Z_j \right|.$$

Since $\widetilde{Z}_j$ lies between Z_j and $\sqrt{n} Z_j / \left\|Z^{(n)}\right\|$, and $\sqrt{n} / \left\|Z^{(n)}\right\|$ converges to 1 almost surely. For each $0 < \bar{C} < \infty$, there exists a random integer $N_1' \in \mathbb{N}$ such that $\mathbb{P}'$-almost surely, for $n > N_1'$,

$$\left|\widetilde{Z}_j\right| < \left|Z_j\right| \bar{C} \quad \text{for } j = 1, \ldots, n.$$

Combining the last three displays and (4.186), we obtain $\mathbb{P}'$-almost surely for $n > N_1'$,

$$\begin{aligned} \sup_{\|t\|<\varepsilon} \left| n^{-\alpha} I_1^n(t) \right| &\leq \left| \frac{\sqrt{n}}{\left\|Z^{(n)}\right\|} - 1 \right| \tilde{C} \frac{1}{n^\alpha} \sum_{j=1}^n (1 + (\bar{C}\varepsilon)^q \left|Z_j\right|^q) \left|Z_j\right| \\ &= \left| \frac{(\sqrt{n} - \left\|Z^{(n)}\right\|) n^{1/2-\alpha}}{\left\|Z^{(n)}\right\| / \sqrt{n}} \right| \tilde{C} \frac{1}{n} \sum_{j=1}^n (1 + (\bar{C}\varepsilon)^q \left|Z_j\right|^q) \left|Z_j\right|. \end{aligned} \tag{4.189}$$

From the Gaussian concentration inequality [78, Theorem 3.1.1], given any $C_1' \in (0, \infty)$, there exists $c \in (0, \infty)$ such that

$$\mathbb{P}\left(n^{1/2-\alpha}\left|\left\|Z^{(n)}\right\| - \sqrt{n}\right| > C_1'\right) = \mathbb{P}\left(\left|\left\|Z^{(n)}\right\| - \sqrt{n}\right| > n^{\alpha-1/2}C_1'\right) \leq 2e^{-c(C_1')^2 n^{2\alpha-1}},$$

which is summable because $\alpha > 1/2$. Hence, by the Borel-Cantelli lemma, there exists a random integer $N_1'' \in \mathbb{N}$ such that $\mathbb{P}'$-almost surely for $n \geq N_1''$,

$$n^{1/2-\alpha}\left|\left\|Z^{(n)}\right\| - \sqrt{n}\right| \leq C_1'. \tag{4.190}$$

On the other hand, since $(Z_j)_{j\in\mathbb{N}}$ are independent, by the strong law of large numbers, as $n \to \infty$, almost surely

$$\frac{1}{n}\sum_{j=1}^{n}(1 + (\bar{C}\varepsilon)^q \left|Z_j\right|^q)\left|Z_j\right| \to \mathbb{E}\left[(1 + (\bar{C}\varepsilon)^q |Z|^q)|Z|\right] \quad \text{and} \quad \frac{\left\|Z^{(n)}\right\|}{\sqrt{n}} \to 1. \tag{4.191}$$

Together, (4.189), (4.190), and (4.191) imply (4.188).

Step 2. We now establish a bound on $I_2^n(t)$ defined in (4.187). We claim that there exist $C_2 \in (0, \infty)$ and a random integer $N_2 \in \mathbb{N}$ such that $\mathbb{P}'$-almost surely, for $n \geq N_2$,

$$\sup_{\|t\|<\varepsilon}\left|n^{-\alpha}I_2^n(t)\right| \leq C_2. \tag{4.192}$$

With a view to proving (4.192), recall the definition of q in (4.183). Since $\alpha > 1/2$, we may choose $\beta > 0$ so that

$$2\alpha - 1 - 2\beta q > 0 \quad \text{and define} \quad r_n := n^\beta. \tag{4.193}$$

Next, define the truncation of $(x, (t_1, t_2)) \mapsto \partial_1\mathcal{K}(\rho(xt_1, t_2), t_2)x$ as follows: for $x \in \mathbb{R}$ and

$t = (t_1, t_2) \in B_\varepsilon(0)$,

$$\mathbb{T}_n(x,t) := \begin{cases} \partial_1 \mathcal{K}\left(\rho\left(xt_1, t_2\right), t_2\right) x, & \text{if } |x| < r_n, \\ \partial_1 \mathcal{K}\left(\rho\left(\operatorname{sgn}(x) r_n t_1, t_2\right), t_2\right) \operatorname{sgn}(x) r_n, & \text{if } |x| \geq r_n, \end{cases} \tag{4.194}$$

where sgn is the sign function defined on $\mathbb{R}$ by $\operatorname{sgn}(x) = 1$ if $x \geq 0$ and $\operatorname{sgn}(x) = -1$ if $x < 0$. We bound (4.192) above by the sum of three terms,

$$\sup_{\|t\|<\varepsilon} \left| n^{-\alpha} I_2^n(t) \right| \leq I_{21}^n + I_{22}^n + I_{23}^n, \tag{4.195}$$

where

$$I_{21}^n := \sup_{\|t\|<\varepsilon} \left| \frac{1}{n^\alpha} \sum_{j=1}^n \left(\mathbb{T}_n(Z_j, t) - \mathbb{E}[\mathbb{T}_n(Z, t)] \right) \right|, \tag{4.196}$$

$$I_{22}^n := \sup_{\|t\|<\varepsilon} \left| \frac{1}{n^\alpha} \sum_{j=1}^n \left(\partial_1 \mathcal{K}(\rho(Z_j t_1, t_2), t_2) Z_j - \mathbb{T}_n(Z_j, t) \right) \right|, \tag{4.197}$$

$$I_{23}^n := \sup_{\|t\|<\varepsilon} \left| \mathbb{E}\left[\frac{1}{n^\alpha} \sum_{j=1}^n \left(\partial_1 \mathcal{K}(\rho(Z_j t_1, t_2), t_2) Z_j - \mathbb{T}_n(Z_j, t) \right) \right] \right|. \tag{4.198}$$

For the first term I_{21}^n in (4.196), we start by proving the following.

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

Proof of the claim. First, note that by (4.182) and (4.194), for $|x| < r_n$, $\|t\| < \varepsilon$ 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 ξ is a constant such that $|\xi| \leq |xt_2|$ from Taylor's theorem, and for $|x| < r_n$, $\|t\| < \varepsilon$

and $t_1 = 0$,

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

By (4.184) and the last two displays, we have $|x| < r_n$, $\|t\| < \varepsilon$,

$$\left|\partial_{t_1}\mathbb{T}_n(x,t)\right| \le \tilde{C}(1+\varepsilon^q r_n^q)r_n^2. \tag{4.199}$$

Similarly, for $|x| < r_n$, $\|t\| < \varepsilon$ and $t_1 \neq 0$,

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

where by Taylor's theorem ξ is a constant such that $|\xi| \le |xt_2|$. Also for $|x| < r_n$, $\|t\| < \varepsilon$ and $t_1 = 0$, by (4.194),

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

The last two displays and (4.184) imply that

$$\left|\partial_{t_2}\mathbb{T}_n(x,t)\right| \le \tilde{C}(1+\varepsilon^q r_n^q)r_n, \quad \text{for } x\in\mathbb{R} \text{ and } \|t\| \le \varepsilon. \tag{4.200}$$

Thus, the claim follows from (4.199) and (4.200). □

For $n \in \mathbb{N}$, define

$$\delta_n := \frac{\varepsilon}{\tilde{C}(1+\varepsilon^q r_n^q)r_n^2 n^{1-\alpha}} \quad \text{and} \quad k_n := \left\lceil\left(\tilde{C}(1+\varepsilon^q r_n^q)r_n^2 n^{1-\alpha}\right)^2\right\rceil. \tag{4.201}$$

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

$$|\mathbb{T}_n(x,u) - \mathbb{T}_n(x,v)| \leq \|u-v\|\, \tilde{C}(1+\varepsilon^q r_n^q) r_n^2 \leq \frac{2\varepsilon}{n^{1-\alpha}}. \tag{4.202}$$

Therefore, we have

$$\begin{aligned}
I_{21}^n &\leq \sup_{k=1,\dots,k_n} \sup_{t\in B_{\delta_n}(l_k)} \left| \frac{1}{n^\alpha} \sum_{j=1}^n \Big(\mathbb{T}_n(Z_j,t) - \mathbb{E}[\mathbb{T}_n(Z,t)] \Big) \right| \\
&\leq \sup_{k=1,\dots,k_n} \left| \frac{1}{n^\alpha} \sum_{j=1}^n \left(\sup_{t\in B_{\delta_n}(l_k)} \mathbb{T}_n(Z_j,t) - \mathbb{E}\left[\sup_{t\in B_{\delta_n}(l_k)} \mathbb{T}_n(Z_j,t) \right] \right) \right| \\
&\quad + n^{1-\alpha} \left| \mathbb{E}\left[\sup_{t\in B_{\delta_n}(l_k)} \mathbb{T}_n(Z,t) - \inf_{t\in B_{\delta_n}(l_k)} \mathbb{T}_n(Z,t) \right] \right| \\
&\leq \sup_{k=1,\dots,k_n} \left| \frac{1}{n^\alpha} \sum_{j=1}^n \left(\sup_{t\in B_{\delta_n}(l_k)} \mathbb{T}_n(Z_j,t) - \mathbb{E}\left[\sup_{t\in B_{\delta_n}(l_k)} \mathbb{T}_n(Z_j,t) \right] \right) \right| + 2\varepsilon,
\end{aligned} \tag{4.203}$$

where the last inequality follows from (4.202). Together, (4.183) and (4.194) 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 [78, Theorem 2.2.6], for any $n \in \mathbb{N}$ and $C_3 \in (0,\infty)$ we have

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

which is summable in n by (4.201) and (4.193). By the Borel-Cantelli lemma and (4.203), it follows that there exists a random integer $N_3 \in \mathbb{N}$ such that $\mathbb{P}'$-almost surely, for $n \geq N_3$, almost surely,

$$I_{21}^n \leq C_3 + 2\varepsilon. \tag{4.204}$$

Next, we deal with the quantity I_{22}^n in (4.197). Note that by (4.197), (4.194) and (4.183)

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

Hence for $C_4 \in (0, \infty)$, Markov's inequality implies

$$\begin{aligned} \mathbb{P}(I_{22}^n > C_4) &\leq \mathbb{P}\left(\frac{1}{n^\alpha} \sum_{i=1}^{n} \left(\left|Z_j\right| \tilde{C}(1 + \varepsilon^q \left|Z_j\right|^q) + r_n \tilde{C}(1 + \varepsilon^q r_n^q) \right) 1_{\{\left|Z_j\right| > r_n\}} > C_4 \right) \\ &\leq \frac{1}{C_4} n^{1-\alpha} \mathbb{E}\left[\left(\left|Z_j\right| \tilde{C}(1 + \varepsilon^q \left|Z_j\right|^q) + r_n \tilde{C}(1 + \varepsilon^q r_n^q) \right) 1_{\{\left|Z_j\right| > r_n\}} \right]. \end{aligned}$$

Now, for any $k \in \mathbb{N} \cup \{0\}$, the Laplace approximation (see e.g. [83, Chapter 2]) implies

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

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 (4.193), by the Borel-Cantelli lemma, there exists a random integer $N_5 \in \mathbb{N}$ such that $\mathbb{P}'$-almost surely, for $n \geq N_5$,

$$I_{22}^n \leq C_4. \tag{4.205}$$

Finally from (4.197), (4.198) and (4.205), we see that $\mathbb{P}'$-almost surely $I_{23}^n \leq \mathbb{E}[I_{22}^n] \leq C_4$ for $n \geq N_5$. When combined with (4.195), (4.204) and (4.205), this yields (4.192). Now, by (4.185), (4.188) and (4.192), we have almost surely, for $\|t\| \leq \varepsilon$ and $n \geq \max\{N_1, N_2\}$,

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

which implies that (4.97) holds $\mathbb{P}'$-almost surely and concludes the proof of the lemma.

□

Chapter Five

Open questions and ongoing work

In this thesis, we study large deviations estimates for high-dimensional measures and convex bodies. Here, we list some ongoing and future works:

- In Section 2.6.2, we verified the asymptotic thin shell condition for random vectors uniformly distributed on an Orlicz ball, only when the associated Orlicz function is superquadratic. It is natural to ask what would happen in the case of subquadratic (but superlinear) Orlicz functions. A special case of this would include ℓ_p^n balls for $p \in [1, 2)$, which is fully understood. In this particular case, one needs to look at (sharp) large deviation estimates for heavy-tailed random variables (because Cramé's Theorem does not apply anymore).
- In the study of intersections of Orlicz balls, we have thus far only understood the asymptotic behavior of the volumes of the intersections for a certain range of values of the radius. The full range is not understood, and one expects the exponential rate of decay in these regions to be different from the usual one. Therefore, new ideas and techniques would likely be needed to obtain the sharp large deviation estimates.
- In the case of quenched random projections, in Chapter 4 we studied only 1-dimensional projection. We are currently working on the case when the dimension is fixed to be a constant $k \in \mathbb{N}$. It would be interesting to ask what the LDP or sharp large deviation estimate would be when the projection dimension k_n of a random vector in $\mathbb{R}^n$ grows with n. This question is nontrivial because it is not even clear what common underlying space the projection should be embedded into as n is increased.
- In the annealed settings, we consider more general objects than ℓ_p^n balls and include superquadratic Orlicz balls and a class of Gibbs measures. Other objects of

interests in convex geometry including zonotopes and Lorentz balls could also be considered. On the other hand, for the quenched setting, it would also be interesting to consider generalization of our results to understand more unstructured probability measures than the normalized volume measures on ℓ_p^n balls.

- Many of our results focus on the theoretical aspects of projections of high-dimensional measures. One potential application of these results is to design hypothesis testing algorithms for distinguishing data sampled from different underlying high-dimensional measures.

- Another possible application of our results is on the refinement of the Johnson-Lindenstrauss Lemma. The Johnson-Lindenstrauss (J-L) Lemma shows that without any distributional assumption on data, high-dimensional data can be mapped to a lower-dimensional space such that the Euclidean geometry (distances between data points) of the data is mostly preserved. More precisely, the J-L Lemma suggests that for a set of n points, it is possible to embed these points to a subspace of dimension $\log(n)/\varepsilon^2$ while the distances between points are "perturbed by at most ε." The result in the J-L Lemma is universal and no distribution on the data is assumed. A natural question is to ask whether given some information of the underlying data, one can obtain a more refined result and further reduce the dimension of the lower-dimensional space.

Bibliography

[1] D. Alonso-Gutiérrez and J. Bastero, *Approaching the Kannan-Lovász-Simonovits and variance conjectures*, Vol. 2131, Springer, 2015.

[2] D. Alonso-Gutiérrez and J. Prochno, *Thin-shell concentration for random vectors in Orlicz balls via moderate deviations and Gibbs measures*, Journal of Functional Analysis (2021), 109291.

[3] D. Alonso-Gutiérrez, J. Prochno, and C. Thäle, *Large deviations for high-dimensional random projections of ℓ_n^p-balls*, Advances in Applied Mahematics **99** (2018), 1–35.

[4] D. Alonso-Gutiérrez, J. Prochno, and C. Thäle, *Large deviations, moderate deviations, and the KLS conjecture*, Journal of Functional Analysis **280** (2021), no. 1, 108779.

[5] B. Anderson, J. Jackson, and M. Sitharam, *Descartes' rule of signs revisited*, American Mathematical Monthly **105** (1998), no. 5, 447–451.

[6] C. Andriani and P. Baldi, *Sharp estimates of deviations of the sample mean in many dimensions*, Annales de l'Institut Henri Poincare (B) Probability and Statistics **33** (1997), no. 3, 371 –385.

[7] J. Angst and G. Poly, *A weak Cramér condition and application to Edgeworth expansions*, Electronic Journal of Probability **22** (2017), no. 0.

[8] M. Anttila, K. Ball, and I. Perissinaki, *The central limit problem for convex bodies*, Transactions of the American Mathematical Society **355** (2003), no. 12, 4723–4735.

[9] R. Bahadur and R. R. Rao, *On deviations of the sample mean*, Annals of Mathematical Statistics **31** (1960), no. 4, 1015–1027.

[10] R. Bahadur and R.R. Rao, *On deviations of the sample mean*, Annals of Mathematical Statistics **31** (1960), no. 4, 1015–1027.

[11] 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.

[12] F. Barthe, F. Gamboa, L.-V. Lozada-Chang, and A. Rouault, *Generalized Dirichlet distributions on the ball and moments*, ALEA Latin American Journal Probability and Mathematical Statistics **7** (2010), 319–340.

[13] F. Barthe, O. Guédon, S. Mendelson, and A. Naor, *A probabilistic approach to the geometry of the ℓ_p^n-ball*, Annals of Probability **33** (2005), no. 2, 480–513.

[14] G. Ben Arous, A. Dembo, and A. Guionnet, *Aging of spherical spin glasses*, Probability Theory and Related Fields **120** (2001), no. 1, 1–67.

[15] N. Bleistein and R. Handelsman, *Asymptotic expansions of integrals*, Courier Corporation, 1986.

[16] K. Borovkov, *On the convergence of projections of uniform distributions on balls*, Theory of Probability and Its Applications **35** (1991), no. 3, 546–550.

[17] 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.

[18] U. Brehm and J. Voigt, *Asymptotics of cross sections for convex bodies*, Beiträge Algebra Geom **41** (2000), no. 2, 437–454.

[19] N. R. Chaganty and J. Sethuraman, *Strong large deviation and local limit theorems*, Annals of Probability **21** (1993), no. 3, 1671–1690.

[20] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed., Wiley, 2001.

[21] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Bartlett Publishers, Boston, MA, 1993.

[22] P. Diaconis and D. Freedman, *Asymptotics of graphical projection pursuit*, Annals of Statistics **12** (1984), no. 3, 793–815.

[23] I. H. Dinwoodie and S. L. Zabell, *Large deviations for exchangeable random vectors*, Annals of Probability **20** (1992), no. 3, 1147–1166.

[24] P. Dupuis, V. Laschos, and K. Ramanan, *Large deviations for configurations generated by Gibbs distributions with energy functionals consisting of singular interaction and weakly confining potentials*, Electronic Journal of Probability **25** (2020).

[25] R. Durrett, *Probability: theory and examples*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2010.

[26] C. Esseen, *Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law*, Acta Mathematica **77** (1945), 1 –125.

[27] K. Fernando and P. Hebbar, *Higher order asymptotics for large deviations – Part I*, Asymptotic Analysis **121** (2021), no. 3-4, 219–257.

[28] B. Fleury, *Between Paouris concentration inequality and variance conjecture*, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 2010, pp. 299–312.

[29] Bruno Fleury, *Concentration in a thin Euclidean shell for log-concave measures*, Journal of Functional Analysis **259** (2010), no. 4, 832–841.

[30] G. Folland, *Real analysis : modern techniques and their applications*, New York: Wiley, 1999.

[31] J. Freedman and J.W.T. Tukey, *A projection pursuit algorithm for exploratory data analysis*, IEEE Transactions on Computers **9** (1974), 881–890.

[32] D. J. Fresen, *A simplified proof of CLT for convex bodies*, Proceedings of the American Mathematical Society **149** (2021), no. 1, 345–354.

[33] 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, 2016, pp. 253–270.

[34] ———, *Large deviations for random projections of ℓ^p balls*, Annals of Probability **45** (2017), 4419–4476.

[35] O. Guédon and E. Milman, *Interpolating thin-shell and sharp large-deviation estimates for lsotropic log-concave measures*, Geometric and Functional Analysis **21** (2011), no. 5, 1043.

[36] J. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms II*, Vol. 305, Springer science & business media, 2013.

[37] P.J. Huber, *Projection pursuit*, Annals of Statistics **13** (1985), 435–475.

[38] P. Indyk and R. Motwani, *Approximate nearest neighbors: towards removing the curse of dimensionality*, STOC '98 (Dallas, TX), 1999, pp. 604–613.

[39] W. B. Johnson and J. Lindenstrauss, *Extensions of Lipschitz mappings into a Hilbert space*, Conference in modern analysis and probability (New Haven, Conn., 1982), 1984, pp. 189–206.

[40] C. Joutard, *Strong large deviations for arbitrary sequences of random variables*, Annals of the Institute of Statistical Mathematics **65** (2012), no. 1, 49–67.

[41] ———, *Multidimensional strong large deviation results*, Metrika **80** (2017), no. 6-8, 663–683.

[42] Z. Kabluchko and J. Prochno, *The maximum entropy principle and volumetric properties of Orlicz balls*, Journal of Mathematical Analysis and Applications **495** (2021), no. 1, 124687.

[43] Z. Kabluchko, J. Prochno, and C. Thäle, *High-dimensional limit theorems for random vectors in ℓ_p^n-balls*, Communications in Contemporary Mathematics **21** (2019), no. 1, 1750092.

[44] Z. Kabluchko, J. Prochno, and C. Thäle, *High-dimensional limit theorems for random vectors in ℓ_p^n-balls. II*, Communications in Contemporary Mathematics **23** (2021), no. 3, 1950073.

[45] D. Kannan and V. Lakshmikanthan, *Handbook of stochastic analysis and applications*, First, Statistics: A Series of Textbooks and Monographs, CRC Press, 2001.

[46] R. Kannan, L. Lovász, and M. Simonovits, *Isoperimetric problems for convex bodies and a localization lemma*, Discrete and Computational Geometry **13** (1995), no. 3-4, 541–559.

[47] T. Kaufmann, *Sharp Asymptotics for q-Norms of Random Vectors in High-Dimensional ℓ_p^n-Balls*, arXiv e-prints (2021).

[48] S. S. Kim, *Problems at the interface of probability and convex geometry: Random projections and constrained processes*, Ph.D. Thesis, 2017.

[49] 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.

[50] S.S. Kim and K. Ramanan, *A conditional limit theorem for high-dimensional ℓ_p spheres*, Journal of Applied Probability **55** (2018), 1060–1077.

[51] B. Klartag, *A central limit theorem for convex sets*, Inventiones Mathematicae **168** (2007), 91–131.

[52] ———, *A central limit theorem for convex sets*, Inventiones mathematicae **168** (2007), 91–131.

[53] A. Klenke, *Probability Theory: A Comprehensive Course*, Springer Science & Business Media, 2013.

[54] A. V. Kolesnikov, E. Milman, et al., *The KLS isoperimetric conjecture for generalized Orlicz balls*, Annals of Probability **46** (2018), no. 6, 3578–3615.

[55] Y.-T. Liao and K. Ramanan, *Quenched large deviations for random projections of high-dimensional measures*. In preparation.

[56] ———, *Sharp large deviation estimate and its application to asymptotic convex geometry*. Preprint.

[57] Y.-T. Liao and K. Ramanan, *Estimating probabilities in asymptotic convex geometry: a case study*, 2020. In preparation.

[58] Y.-T. Liao and K. Ramanan, *Geometric sharp large deviations for random projections of ℓ_p^n spheres and balls*, arXiv e-prints (2020), arXiv:2001.04053, available at 2001.04053.

[59] W. Liu and L. Wu, *Large deviations for empirical measures of mean-field Gibbs measures*, Stochastic Processes and their Applications **130** (2020), no. 2, 503–520.

[60] O. Maillard and R. Munos, *Compressed least-squares regression*, Advances in Neural Information Processing Systems 22, 2009, pp. 1213–1221.

[61] E. Meckes, *Approximation of projections of random vectors*, Journal of Theoretical Probability **25** (2012), no. 2, 333–352.

[62] M. E. Muller, *A note on a method for generating points uniformly on n-dimensional spheres*, Communications of the ACM **2** (1959), no. 4, 19–20.

[63] J. Musielak, *Orlicz Spaces and Modular Spaces*, Lecture Notes in Mathematics, vol. 1034, Springer, 1983.

[64] S. Nadarajah, *A generalized normal distribution*, Journal of Applied Statistics **32** (2005), no. 7, 685–694.

[65] A V Nagaev, *Integral limit theorems taking large deviations into account when Cramér's condition does not hold. I*, Theory of Probability and its Applications **14** (1969), no. 1, 51–64.

[66] F. Olver, *Asymptotics and special functions*, AK Peters/CRC Press, 1997.

[67] G. Paouris, *Small ball probability estimates for log-concave measures*, Transactions of the American Mathematical Society **364** (2012), no. 1, 287–308.

[68] Grigoris Paouris, *Concentration of mass on convex bodies*, Geometric & Functional Analysis GAFA **16** (2006), no. 5, 1021–1049.

[69] C. H. Papadimitriou, P. Raghavan, H. Tamaki, and S. Vempala, *Latent semantic indexing: a probabilistic analysis*, J. Comput. System Sci. **61** (2000), no. 2, 217–235. Special issue on the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (Seattle, WA, 1998). MR1802556

[70] M. Pilipczuk and J. O. Wojtaszczyk, *The negative association property for the absolute values of random variables equidistributed on a generalized Orlicz ball*, Positivity **12** (2008), no. 3, 421–474.

[71] S. T. Rachev and L. Rüschendorf, *Approximate independence of distributions on spheres and their stability properties*, Annals of Probability **19** (1991), no. 3, 1311–1337.

[72] R. Rockafellar, *Convex analysis*, Princeton University Press, 1970.

[73] G Schechtman and M Schmuckenschläger, *Another remark on the volume of the intersection of two l_p^n balls*, Geometric aspects of functional analysis, 1991, pp. 174–178.

[74] G. Schechtman and J. Zinn, *On the volume of the intersection of two l_p^n balls*, Proceedings of the American Mathematical Society **110** (1990), no. 1, 217–217.

[75] ———, *On the volume of the intersection of two l_p^n balls*, Proceedings of the American Mathematical Society **110** (1990), no. 1, 217–224.

[76] M. Schmuckenschläger, *Clt and the volume of intersections of ℓ_p^n-balls*, Geometriae Dedicata **85** (2001), 189–195.

[77] 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.

[78] R. Vershynin, *High-dimensional probability: an introduction with applications in data science*, Vol. 47, Cambridge University Press, 2018.

[79] C. Villani, *Optimal Transport: Old and New*, Vol. 338, Springer Science & Business Media, 2008.

[80] 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

[81] R. Wang, X. Wang, and L. Wu, *Sanov's theorem in the Wasserstein distance: a necessary and sufficient condition*, Statistics & Probability Letters **80** (2010), no. 5, 505–512.

[82] J. G. Wendel, *Note on the Gamma function*, American Mathematical Monthly **55** (1948), no. 9, 563–564.

[83] R. Wong, *Asymptotic approximations of integrals*, Society for Industrial and Applied Mathematics, 2001.

[84] R. Zia, E. Redish, and S. McKay, *Making sense of the Legendre transform*, American Journal of Physics **77** (2009), no. 7, 614–622.