# Concentration Phenomena on High-Dimensional $\ell_p^n$-Balls

by


Tom Kaufmann

Matriculation number 108013207000

Born 15.03.1994


A thesis handed in to the graduate faculty of the

Ruhr-University Bochum

in partial fulfillment of the requirements for the degree of

Doctor Rerum Naturalium

in the department of Mathematics

July, 2022

# Acknowledgment

First and foremost I want to sincerely thank my PhD supervisor *Christoph Thäle*. During my master's degree he introduced me to the beautiful and interesting field of stochastic geometry and was both a great teacher who was passionate about his subject and a very supportive supervisor. As my PhD supervisor he was a source of constant insight, enthusiasm, and instructive criticism, and really helped me grow as a mathematician. I thank him greatly for the opportunity of doing my PhD with him and always being highly approachable and taking the time for my colleagues and me. He was the best supervisor I could have asked for.

I am also grateful to my second supervisor *Asja Fischer*, through whom I got to know the field of machine learning, which has become a major interest of mine beside my main area of research. I thank her for several semesters of fun teaching and very enjoyable collaboration.

Moreover, I want to express my gratitude to my current and former colleagues *Nils Heerten*, *Holger Sambale*, *Carina Betken*, *Kathrin Meier*, *Tristan Schiller*, *Daniel Rosen*, *Gilles Bonnet*, *Anna Gusakova*, *Anastas Baci*, and *Nicola Turchi*. You not only were of great help academically numerous times, but you also made working at RUB for the last 3+ years a lot of fun. Furthermore, my sincere thanks go to *Tanja Schiffmann* for continually going above and beyond to help my colleagues and me with whatever problem we might have had and being very kind in doing so.

Further, I would like to thank *Nina Gantert* for agreeing to review my PhD thesis and *Joscha Prochno* for inviting me to the university of Passau in March of 2022 and offering a lot of helpful and instructive insights into stochastic geometry.

I thank the *Ruhr-University Bochum* and the *Faculty of Mathematics*, namely *Eva Glasmachers* and *Peter Eichelsbacher*, the latter of whom I specifically would like to thank for giving me the opportunity of studying a year abroad at Durham University in the United Kingdom in 2015 via the ERASMUS scholarship.

In my very first maths lecture the dean at the time, Herold Dehling, told us that "Mathematics is not something you study on your own". This was true during my time as an undergraduate and it continued to be true for my time as a PhD student. Hence I would like to thank my friends and colleagues from Uni for the both difficult and exciting shared experience of the study of mathematics, namely *Johannes Pieperbeck*, *Oliver Brammen*, *Maxim Kukol* and *Hannah Wollny*. Special thanks go to my great friend and colleague *Janina Hübner*, with whom every challenge academia might throw at us seemed conquerable. "Wir machen das schon." :)

Doing a PhD is hard. Doing it during one of the biggest global pandemics in history certainly did not make it any easier. I want to thank everyone who helped me keep up something approximating a normal social life, despite the trying but necessary restrictions. First, I would like to thank *Maria Rechenberger* for keeping me grounded and cheering me up whenever I was stressed. Moreover, I want to thank *Stefan Rieche*, my Boulder-circle *Hendrik Enge*, *Katharina Dreyer*, *Lennart Schürmann*, *Hanne Köster*, *Chiara Steller*, and *Elias Kerperin*, and both the *Weltenrand Community* and the *KSG Essen e.V. Abteilung Tabletop*, specifically *Sandra Steier*, *Torben Arndt*, and *Andreas Proplesch*.

The person I have most to thank for is my father *Ralf Kaufmann*. Not just for his continuous support, but also for sparking an interest and joy in learning and doing mathematics in me at an early age and keeping the very same interest himself. He continues to show me that learning does not need to stop at some point in life and there are always new challenges to set yourself and overcome. I would also like to thank my family, my mother *Gabriele Vossen*, my sister *Lea Kaufmann*, my grandparents *Erika* and *Helmut Burggraf* and *Katharina Schneider*.

I would also like to express my gratitude to *Hans Zander*, my school maths teacher, for not just teaching his subject, but being passionate about it and being able to pass this on to his students.

# Contents

# Chapter 1

# Introduction

This chapter consists of two sections. The first section gives an overview of the field of asymptotic geometric analysis, its origin, objects of study, and methodological toolbox. Specifically, the ideas of large deviations theory and sharp large deviations theory are briefly outlined, as they motivate several of the problems tackled within this thesis. Lastly, the primary object of study of the present work, the $\ell_p^n$-ball, shall be defined and discussed with respect to its overall relevance to the field of research, and some problems with respect to it are presented. The second section provides a guideline of this thesis, outlining the contents of each chapter and putting them into their respective research contexts, that is, pointing out the underlying publications they are based on and the relevant research preceding these results.

## 1.1 General introduction

In $n$-dimensional Euclidean space there is a one-to-one correspondence between norms and symmetric convex bodies. Any given norm $\|\cdot\|$ on $\mathbb{R}^n$ defines a symmetric convex body in the form of its unit ball

$$\mathbb{B}_{\|\cdot\|} := \{x \in \mathbb{R}^n : \|x\| \leq 1\}$$

and, vice versa, a symmetric convex body $K \subset \mathbb{R}^n$ induces a norm $\|\cdot\|_K$ on $\mathbb{R}^n$ via the Minkowski functional

$$\|x\|_K := \inf\{r \in [0,\infty) : x \in rK\}, \qquad x \in \mathbb{R}^n, \tag{1.1}$$

with respect to which $K$ itself is the unit ball $\mathbb{B}_{\|\cdot\|_K}$. We can thus see how the study of norms (or normed spaces) and symmetric convex bodies are closely related.

The study of convex bodies in high dimensions, known today as asymptotic geometric analysis, has arisen from the local theory of Banach spaces, which aimed at analyzing infinite-dimensional normed spaces via their local substructures, such as their unit balls. Although there is some debate on what exact problems and perspectives are at the core of the local theory of Banach spaces (see [96]), the characterization best befitting the connection to asymptotic geometric analysis seems to be the one of Lindenstrauss and Milman [86, p. 1151]:

*The local theory of Banach spaces deals with convex bodies in $\mathbb{R}^n$ where $n$ is finite but large. The main theme of the theory is a quantitative study of the structure of such sets and asymptotic estimates of various parameters associated with them as $n \to \infty$. [...]*

*The name "local theory" is applied to two somewhat different topics:*

1. *The quantitative study of $n$-dimensional normed spaces as $n \to \infty$.*
2. *The relation of the structure of an infinite-dimensional space and its finite-dimensional subspaces.*

Given an infinite-dimensional Banach space, local structures like its unit ball are naturally of infinite dimension as well, and since working in infinite dimensions is inherently more difficult than working in the finite-dimensional setting, it is a fruitful approach to instead study the finite-dimensional counterparts of such structures asymptotically in the limit of the dimension. This was the motivating impulse giving rise to the field of asymptotic geometric analysis and has yielded some highly relevant results, such as solutions to Banachs' hyperplane problem by Gowers in [40] or the unconditional basic sequence problem by Gowers and Maurey in [41] (see [90] for a broader context on these results).

Despite having its origin in the realm of functional analysis, the field has since established itself in its own right, also considering problems beyond the study of centrally symmetric convex bodies that occur naturally as the unit balls of Banach spaces. High-dimensional convexity furthermore has a large number of applications, e.g., in signal processing, such as compressed sensing (see [20, 34]) and sparse signal recovery (see [114, Chapter 10]), or random information and approximation theory (see, e.g., [49, 50, 51, 83]). Since a discussion of the many applications of high-dimensional convexity is beyond the scope of this thesis, we refer to the excellent book of Vershynin [114] for a more comprehensive look into where its concepts are used in a wide variety of data-driven fields of work and study.

In high dimensions convex bodies exhibit certain regularities, such as volume concentration phenomena (see, e.g., [18, 45, 46]), which make it highly useful to approach them from a probabilistic perspective. As pointed out in [9], it might seem counterintuitive to analyze something exhibiting regularities from a probabilistic perspective, as probability concerns itself with studying the nature of irregularity, i.e., randomness, of given quantities. But as with well-known limit theorems from probability, such as the law of large numbers and the central limit theorem (CLT), with large sample sizes (and analogously – with high dimensionality) random objects exhibit interesting patterns well characterized in the language of probability and vice versa. Thus, one can view asymptotic geometric analysis as being located somewhat at the intersection between geometry, functional analysis, and probability theory.

Many results analogous to those from classic probability have been found for high-dimensional convex sets, one of the most notable being the central limit theorem for convex bodies shown by Klartag [77, 78] based on the work of Anttila, Ball and Perissinaki [8]. He showed that for fixed $k \in \mathbb{N}$ the $k$-dimensional marginal distributions of isotropic convex bodies in high dimensions are approximately Gaussian.

This central limit theorem was actually shown in a much more general setting for isotropic log-concave measures on $\mathbb{R}^n$, and uniform distributions on convex bodies are merely a special case for such measures. However, while the expansion of results from convex bodies to log-concave measures is an interesting area of research, sometimes referred to as "*Geometrization of Probability*" (see [87, 89]), this thesis will only focus on convex bodies in high dimensions.

The aforementioned concentration phenomena run against our understanding of geometric objects from three-dimensional space and are, on the contrary, often beautifully counterintuitive. To illustrate this, let us give a very classic example of high-dimensional concentration of mass by considering the volume of the cube $[-1, 1]^n$ and standard unit ball $\mathbb{B}_2^n$ with respect to the Euclidean norm as the dimension tends to infinity. One can see directly that $\mathbb{B}_2^n$ is the inball of $[-1, 1]^n$, intersecting the cube at the midpoints of all its facets. Calculating the volume of both $[-1, 1]^n$ and $\mathbb{B}_2^n$ yields that

$$\mathrm{vol}_n([-1,1]^n) = 2^n \qquad \text{and} \qquad \mathrm{vol}_n(\mathbb{B}_2^n) = \frac{\pi^{n/2}}{\Gamma\big(1+\frac{n}{2}\big)},$$

with $\Gamma(\cdot)$ denoting the Gamma function. Thus, we can see that $\mathrm{vol}_n([-1, 1]^n)$ tends to infinity in $n$, and applying Stirling's formula for the Gamma function (see (2.4)) yields that $\mathrm{vol}_n(\mathbb{B}_2^n)$ behaves like $n^{-n/2}$, that is, tends to zero in $n$. So despite $\mathbb{B}_2^n$ intersecting all facets of $[-1,1]^n$ at their midpoints and obviously being convex, the volume it encloses tends to zero with increasing dimension $n$, while that of the cube tends to infinity. The conclusion thus has to be that almost all of the volume of the cube is concentrated "in the corners", that is, outside of the ball. Also, in relationship to the cube, the ball tends to "disappear" in a volumetric sense. The fact that the volume, i.e., the Lebesgue measure, which should yield a homogeneous distribution of mass, exhibits such concentration phenomena in high dimensions is highly counterintuitive.

Figure 1.1 aims to illustrate how this volumetric relationship of $[-1,1]^n$ and $\mathbb{B}_2^n$ in high dimensions could be thought of visually, that is, the volume of $\mathbb{B}_2^n$ tending to zero, specifically in relation to $[-1,1]^n$, contrasted with the classic way we think of the geometric relationship of the cube and its inball.

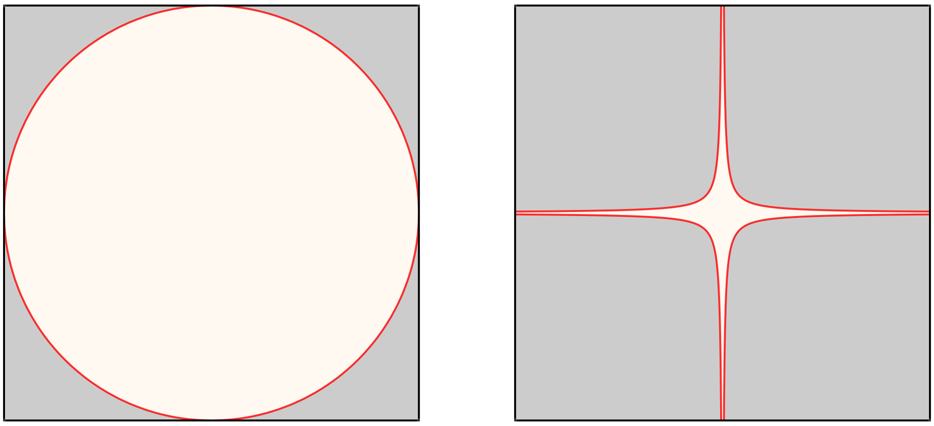

FIGURE 1.1: Two-dimensional representation of the geometric (left) and the volumetric (right) relationship between $[-1,1]^n$ and $\mathbb{B}_2^n$ in high dimensions.

To add even more counterintuitive behaviour, Klartag showed the central limit theorem for high-dimensional isotropic convex bodies in [77, 78] by proving a so-called thin-shell concentration, i.e., that the Euclidean norm of uniform random vectors in isotropic convex bodies in high dimensions heavily concentrates in a thin shell of radius $\sqrt{n}$ (see Figure 1.2). (Note that isotropy – or "being in isotropic position" – of a convex body means that it has unit volume, its barycenter lies at the origin, and its inertia matrix is a multiple of the identity matrix). So merely thinking that volume generally concentrates around the boundary or in the "corners" of a convex body, as the previous example might suggest, would not be correct.

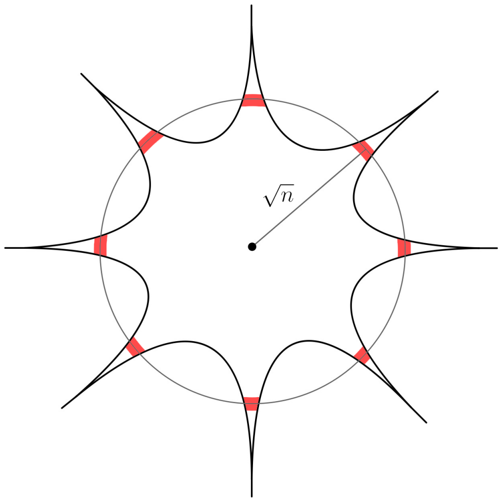


FIGURE 1.2: Volumetric representation for the concentration of mass (red) within a thin shell of radius $\sqrt{n}$ in an isotropic convex body in high dimensions.

Throughout this thesis we will primarily be analyzing the unit balls of the finite-dimensional counterparts of the sequence space $\ell_p$, $p \in (0,\infty]$, where $\ell_p$ is defined as the space of absolutely $p$-summable sequences in $\mathbb{R}$, i.e.,

$$\ell_p := \big\{(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} : \|(x_n)_{n\in\mathbb{N}}\|_{\ell_p} < \infty\big\}.$$

with the $\ell_p$-norm $\|(x_n)_{n\in\mathbb{N}}\|_{\ell_p}$ defined as

$$\|(x_n)_{n\in\mathbb{N}}\|_{\ell_p} := \begin{cases} \left(\sum\limits_{n=1}^{\infty} |x_n|^p\right)^{1/p} & : p < \infty \\ \sup\limits_{n\in\mathbb{N}} \{|x_n|\} & : p = \infty, \end{cases}$$

(which for $p \in (0,1)$ is only a quasi-norm, as the triangle inequality does not hold). The corresponding unit ball with respect to the $\ell_p$-norm is thus

$$\mathbb{B}_{\ell_p} := \big\{(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} : \|(x_n)_{n\in\mathbb{N}}\|_{\ell_p} \leq 1\big\},$$

which is infinite-dimensional, just as $\ell_p$ itself. Equipped with this norm, for $p \in [1,\infty]$, $\ell_p$ is a Banach space and is of great interest from the perspective of functional analysis, both for its own sake and its close connection to the omnipresent $L_p$-space, that is, the space of absolutely $p$-integrable real-valued functions, defined as

$$L_p := \big\{f : \mathbb{R} \to \mathbb{R} : \|f\|_{L_p} < \infty\big\}, \tag{1.2}$$

with the $L_p$-norm $\|f\|_{L_p}$ defined as

$$\|f\|_{L_p} := \begin{cases} \left( \int\limits_{\mathbb{R}} |f(x)|^p \,\mathrm{d}x \right)^{1/p} & : p < \infty \\ \operatorname*{ess\,sup}\limits_{x \in \mathbb{R}} |f(x)| & : p = \infty, \end{cases} \tag{1.3}$$

with $\operatorname{ess\,sup}_{x \in \mathbb{R}} |f(x)|$ denoting the essential supremum. (Note, that this again only denotes a norm for $p \in [1, \infty]$ and only on the class of functions that differ on sets of positive measure.) This connection is due to the fact that $\ell_p$ essentially forms a discretization of $L_p$ by considering the counting measure instead of the Lebesgue measure. Hence, as previously explained, analyzing the finite-dimensional analogues of $\ell_p$ as the dimension tends to infinity is a relevant and promising line of inquiry. For $p \in (0, \infty]$, $n \in \mathbb{N}$, and $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ let us define the $\ell_p^n$-norm on $\mathbb{R}^n$ as

$$\|x\|_p := \begin{cases} \left( \sum\limits_{i=1}^n |x_i|^p \right)^{1/p} & : p < \infty \\ \max\{|x_1|, \dots, |x_n|\} & : p = \infty, \end{cases}$$

(again, only defining a quasi-norm for $p \in (0, 1)$). Thus, $\mathbb{R}^n$ equipped with the $\ell_p^n$-(quasi)-norm is the finite-dimensional counterpart to $\ell_p$. We then define

$$\mathbb{B}_p^n := \{x \in \mathbb{R}^n : \|x\|_p \leq 1\} \qquad \text{and} \qquad \mathbb{S}_p^{n-1} := \{x \in \mathbb{R}^n : \|x\|_p = 1\}$$

to be the *unit* $\ell_p^n$*-ball* and *unit* $\ell_p^n$*-sphere*, respectively (sometimes simply referred to as the *p-ball* and the *p-sphere*).

Another reason why we will be considering $\mathbb{B}_p^n$ and $\mathbb{S}_p^{n-1}$ specifically, besides their connection to the sequence space $\ell_p$, is that their Minkowski functional $\|\cdot\|_{\mathbb{B}_p^n} = \|\cdot\|_p$ has a convenient form (see Remark 2.4.8), which allows us to construct random vectors from i.i.d. random variables that are equal in distribution to random vectors from $\mathbb{B}_p^n$ and $\mathbb{S}_p^{n-1}$ with a multitude of distributions. This, in turn, makes many functionals of random vectors in $\mathbb{B}_p^n$ and $\mathbb{S}_p^{n-1}$ accessible for calculations. This form of reconstruction of random vectors from $\mathbb{B}_p^n$ and $\mathbb{S}_p^{n-1}$ via i.i.d. random vectors is what we will refer to as *probabilistic representation* and it will be one of the main tools within this thesis and will be explained in greater detail in Section 2.4.1. So overall, we consider $\ell_p^n$-balls in high dimensions because they are both relevant and accessible.

A detailed overview of the results for $\ell_p^n$-balls contained within this thesis will be given in Section 1.2, however we shall give a rough outline of some of the quantities generally of interest when considering $\ell_p^n$-balls, focusing specifically on those quantities at the center of this work with the goal of introducing and motivating them.

The first canonical objects of study are volume distributions. As we have seen in the previous example regarding the volumes of the ball and the cube, which happen to be $\ell_p^n$-balls for $p = 2$ and $p = \infty$, respectively, the volumetric behaviour of $\mathbb{B}_p^n$ is by no means equal for all values of $p$. The volume of $\ell_p^n$-balls was shown by Dirichlet in [32] to be

$$\mathrm{vol}_n(\mathbb{B}_p^n) = \frac{\left(2\,\Gamma\left(1+\frac{1}{p}\right)\right)^n}{\Gamma\left(1+\frac{n}{p}\right)}.$$

However, we also saw in the above example how not only considering the volumes of $\mathbb{B}_2^n$ and $\mathbb{B}_\infty^n$ and their limits, but also their geometric relationship, i.e., $\mathbb{B}_2^n$ being the inball of $\mathbb{B}_\infty^n$, gave us some understanding where the mass of $\mathbb{B}_\infty^n$ is concentrated. Thus, a relevant question to ask about $\ell_p^n$-balls is regarding their volumes relative to each other, or rather, their intersection volume, as the dimension tends to infinity. (Note that insight on the volume distribution within $\ell_p^n$-balls can be gained via different approaches as well, may it be by considering the marginals of uniform distributions, showing thin-shell concentrations etc.). The intersection volume of different $\ell_p^n$-balls will be considered in Chapter 5.

Another natural quantity of interest is the projection behaviour of $\ell_p^n$-balls. That is, given a sequence of random vectors $X^{(n)}$ with certain distributions on $\mathbb{B}_p^n$ or $\mathbb{S}_p^{n-1}$ and a sequence of random subspaces $E^{(n)}$, one could ask for the properties of the projection point $P_E X$ of $X^{(n)}$ onto $E^{(n)}$, i.e., its distribution or its norm $\|P_E X\|$, as in Figure 1.3. Random projections of distributions on $\ell_p^n$-balls will be the topic of Chapter 4.

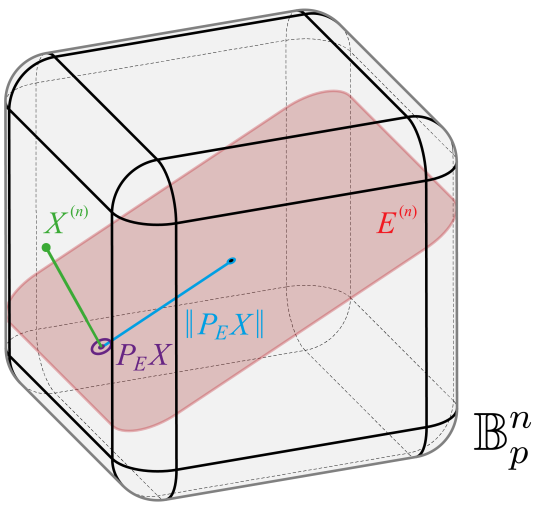


FIGURE 1.3: Projection $P_E X$ of a random vector $X^{(n)} \in \mathbb{B}_p^n$ onto a random $(n-1)$-dimensional subspace $E^{(n)}$ and its norm $\|P_E X\|$ for $n = 3$ and $p > 2$.

The last quantity of interest we want to mention is the empirical measure with respect to a random vector within $\mathbb{B}_p^n$ or $\mathbb{S}_p^{n-1}$. For such a random vector $X^{(n)} = \big(X_1^{(n)}, \dots, X_n^{(n)}\big)$ the empirical measure is a random measure on $\mathbb{R}$, defined as

$$\nu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i^{(n)}},$$

where $\delta_x$ denotes the Dirac measure for some $x \in \mathbb{R}$. Understanding the behaviour of the empirical measure gives insight into the mean behaviour of the coordinates of $X^{(n)}$ and is more accessible for calculations when considering functionals of the form $\frac{1}{n}\sum_{i=1}^{n} f\big(X_i^{(n)}\big)$ for suitable $f$. The behaviour of the empirical measures with respect to the coordinates of random vectors from $\mathbb{B}_p^n$ will be examined in Chapter 3.

These are some of the quantities whose asymptotics are of interest and for which one would like to obtain concentration results such as large deviation principles (LDPs), which will be one of the main focus points of this thesis.

The field of large deviations has only been introduced into asymptotic geometric analysis fairly recently by Gantert, Kim, and Ramanan [36] in 2017, who derived an LDP for projections of $\ell_p^n$-balls onto one-dimensional subspaces. This has since spawned a wave of large deviation results in asymptotic geometric analysis (see, e.g., [4, 5, 35, 61, 62, 65, 69, 70, 71, 72, 73, 74, 75, 76, 85]), among which were the results contained in this thesis. While their concrete research context will be given in Section 1.2 and a more detailed look into large deviations theory provided in Section 2.3, let us give a rough sketch of the basic ideas in order to motivate the results in this work and conceptually contrast them with the *sharp* large deviation results also contained herein, whose details and background will be given in Section 5.1.

Generally speaking, the field of large deviations theory concerns itself with the study of rare events, that is, deviations of sequences of random variables $(X_n)_{n\in\mathbb{N}}$ from their expectation beyond the Gaussian scale (of course, the theory has quite some more depth to it, as it not only considers sequences of random variables, but also general families of probability distributions (see Definition 2.3.2), but for now we shall stick to a more simplified perspective). To be more precise, by *large deviations* we mean deviations of order at least $n$, whereas deviations of order between $\sqrt{n}$ and $n$ are referred to as *moderate deviations* Take the simple case of a sequence of real-valued random variables $(X_n)_{n\in\mathbb{N}}$ that concentrate around their expectation in $n \in \mathbb{N}$ (e.g., the empirical average of i.i.d. random variables). The goal of large deviations theory is to characterize how the probability of deviations of order $n$ decays, that is, to give two functions $s : \mathbb{N} \to (0, \infty)$ and $\mathcal{I}_X : \mathbb{R} \to [0, \infty)$, called the *speed* and the *rate function*, respectively, such that for $\epsilon > 0$

$$\lim_{n\to\infty} \frac{1}{s(n)} \log \mathbb{P}\Big(X^{(n)} - \mathbb{E}\left[X^{(n)}\right] \geq n\epsilon\Big) = -\mathcal{I}_X(\epsilon). \tag{1.4}$$

This is essentially a special case of the definition of a large deviation principle (see Definition 2.3.2). In other words, we want to give two functions, such that

$$\mathbb{P}\Big(X^{(n)} - \mathbb{E}\left[X^{(n)}\right] \geq n\epsilon\Big) = e^{-s(n)[\mathcal{I}_X(\epsilon)+o(1)]}, \tag{1.5}$$

where $o(1)$ denotes a sequence that tends to zero as $n \to \infty$. This is the type of result we would like to derive for functionals of random vectors in convex bodies, specifically $\ell_p^n$-balls, as the dimension tends to infinity, since in the setting of high-dimensional convex geometry the sequence parameter of $(X_n)_{n\in\mathbb{N}}$ coincides with the dimension of the ambient space.

While (1.5) nicely illustrates how deviation probabilities are characterized, (1.4) underlines that at the core of large deviations theory are asymptotic results on a *logarithmic* scale. While those are surely useful, having similar results on a non-logarithmic scale would be preferable, as one can gain much more accurate probability estimates for deviation events from them. If we wanted to get a non-logarithmically scaled probability estimate via a large deviation result such as (1.5), we would have to contend with an unknown prefactor $e^{-s(n)o(1)}$. Hence, one would be interested additionally in characterizing that prefactor, i.e., finding a $c_X : \mathbb{N} \times \mathbb{R} \to [0, \infty)$ such that

$$\mathbb{P}\Big(X^{(n)} > z\Big) = c_X(n, z)\, e^{-s(n)\mathcal{I}_X(z)}(1 + o(1)).$$

for $z > \mathbb{E}[X^{(n)}]$. Note that the error term in the above is no longer contained in the exponent, hence the result is not logarithmically scaled. Characterizing the decay of large deviation probabilities on a non-logarithmic scale is what we refer to as *sharp* large deviations theory. It was introduced into asymptotic geometric analysis very recently by Liao and Ramanan in [85] in 2020 and promises to be a fruitful line of inquiry for future work. Besides interests from the purely mathematical perspective, the non-logarithmic nature of the results, i.e., their asymptotic sharpness, is very useful for applications like importance sampling algorithms, that is, generating samples of very rare events, where classic Monte-Carlo methods become inefficient (see [85, Section 3]).

## 1.2 Guideline

This section will outline the specific results contained in this thesis and point out their respective research context. The present thesis can be thematically split into two parts with different overarching approaches: The first deals with problems from the theory of large deviations and is comprised of Chapter 3 and Chapter 4, while the second part addresses problems from sharp large deviations theory and is contained in Chapter 5.

**Chapter 2:** In this introductory chapter, we will define the concepts and notation we will need throughout this work. This will encompass presenting some basic notation and definitions from probability theory, among other things introducing the $p$-generalized Gaussian distributions in Section 2.2, which make up the core building block of the probabilistic representations integral to all of the results in this thesis. Section 2.3 will then explain the basic notions of large deviations and present some essential methods and tools of large deviations theory. The chapter will then conclude with Section 2.4, defining the relevant probability distributions on $\ell_p^n$-balls, such as the uniform distribution $\mathbf{U}_{n,p}$ on $\mathbb{B}_p^n$ and the so-called cone probability measure $\mathbf{C}_{n,p}$ on $\mathbb{S}_p^{n-1}$, which assigns to a set on $\mathbb{S}_p^{n-1}$ the probability given by the volume of the cone it encloses with the origin relative to the volume of $\mathbb{B}_p^n$ (see (2.12)). Also, some well-established results for $\ell_p^n$-balls we will need throughout this thesis are presented.

Let us provide some context by mentioning some results for $\ell_p^n$-balls that have been shown. For the Euclidean sphere $\mathbb{S}_2^{n-1}$ the Poincaré-Maxwell-Borel lemma states that the joint distribution of any fixed number of $k$ coordinates of a random vector with distribution $\mathbf{C}_{n,2}$ is approximately standard Gaussian (see [31]). This was extended to $\mathbb{S}_p^{n-1}$ for any $p \in [1, \infty]$ by Rachev and Rüschendorf [99] and Naor and Romik [93]. Rachev and Rüschendorf [99] and Schechtman and Zinn [106] then also provided a probabilistic representation for random vectors with distributions $\mathbf{U}_{n,p}$ and $\mathbf{C}_{n,p}$. For $p \in (0, \infty]$ this was generalized by Barthe, Guédon, Mendelson and Naor [13], who gave a probabilistic representation for a class of "mixtures" of $\mathbf{C}_{n,p}$ and $\mathbf{U}_{n,p}$. For a Borel probability measure $\mathbf{W}$ on $[0, \infty)$ they defined the class of distributions

$$\mathbf{P}_{n,p,\mathbf{W}} := \mathbf{W}(\{0\})\mathbf{C}_{n,p} + \Psi\mathbf{U}_{n,p}$$

on $\mathbb{B}_p^n$, where $\Psi$ is an appropriate $p$-radial density that depends on $\mathbf{W}$, and provided a convenient representation of $\mathbf{P}_{n,p,\mathbf{W}}$ via a random vector of $p$-generalized Gaussians. The choice of $\mathbf{W}$ determines how exactly the cone probability measure and the uniform distribution are "mixed". This class of measures and its corresponding representations have gained considerable interest in asymptotic geometric analysis and were used in a variety of applications (see [4, 5, 12, 36, 92, 94, 109], to name just a few). These probabilistic representation results will be presented in Section 2.4, followed by a brief discussion on polar integration in the spirit of [98, Section 3.2]. Therein we will both establish useful polar integration formulae and reflect on why the results within this thesis and the surrounding research context focus specifically on $\ell_p^n$-balls.

**Chapter 3:** The chapter begins by introducing a class of distributions on $\ell_p^n$-balls of the following form: For some suitable homogeneous function $f : \mathbb{R}^n \to [0, \infty)$ we construct weighted versions $\mathbf{U}_{n,p,f}$ and $\mathbf{C}_{n,p,f}$ of the uniform distribution $\mathbf{U}_{n,p}$ and the cone probability measure $\mathbf{C}_{n,p}$ in the sense that their densities are weighted by $f$. We accordingly construct a weighted analogue to $\mathbf{P}_{n,p,\mathbf{W}}$ in the form of

$$\mathbf{P}_{n,p,\mathbf{W},f} := \mathbf{W}(\{0\})\mathbf{C}_{n,p,f} + \Psi \mathbf{U}_{n,p,f}.$$

For all of these distributions probabilistic representations results via $p$-generalized Gaussian random variables in the spirit of [13] are derived, which also need to be weighted accordingly via the function $f$. These weighted $p$-radial distributions will be of great use when considering $p$-balls in other spaces than standard Euclidean space.

In this chapter, we study concentration phenomena on $p$-balls in both Euclidean space and within finite-dimensional Schatten trace classes $\mathcal{S}_p^n$ in matrix space. Generally, for a given $p \in (0, \infty]$, the Schatten trace class $\mathcal{S}_p$ is the Banach space of compact linear operators between two Hilbert spaces whose singular values form a sequence within the sequence space $\ell_p$. We will, however, focus on the finite-dimensional Schatten trace classes $\mathcal{S}_p^n$, i.e., the spaces of $(n \times n)$ matrices (with real, complex or quaternionic entries) whose singular values form a vector in $\ell_p^n$ (that is, $\mathbb{R}^n$ with the $\ell_p^n$-norm). Additionally, we will also consider their self-adjoint subclasses, that is, the spaces of self-adjoint $(n \times n)$ matrices whose eigenvalues also form a vector in $\ell_p^n$. The unit balls in these Schatten trace classes $\mathcal{S}_p^n$ are what we will refer to as *matrix $p$-balls*.

There has been a rising interest in the study of these Schatten trace classes and their unit balls in recent years. For example, Guédon and Paouris [47] provided concentration inequalities for points uniformly distributed within the matrix $p$-ball. Moreover, König, Meyer, and Pajor [80] showed that the isotropic constants of matrix $p$-balls (for $p \in [1, \infty]$) are bounded. Barthe and Cordero-Erausquin [11] derived variance estimates, Radke and Vritsiou [100] proved the thin-shell conjecture, and Vritsiou [115] showed the variance conjecture for the operator norm in $\mathcal{S}_p^n$. Hinrichs, Prochno and Vybíral [52, 53] derived optimal bounds for the entropy numbers and sharp estimates for the Gelfand numbers of natural embeddings of $\mathcal{S}_p^n$ and Prochno and Strzelecki [97] also considered the approximation numbers of such embeddings and studied their relationship to the Gelfand and Kolmogorov numbers. Kabluchko, Prochno and Thäle [63, 64] gave the exact asymptotic volumes and volume ratios of matrix $p$-balls and studied their intersection volumes. Also, Kabluchko, Prochno and Thäle [62, 63] studied the eigenvalue distribution as well as singular value distribution of random matrices distributed according to the cone probability measure and the uniform distribution in matrix $p$-balls. It is a well-known fact from random matrix theory that eigenvalues of self-adjoint random matrices behave, figuratively speaking, like particles of a gas, insofar as they "repell each other", which is why eigenvalue distributions of random matrices are often used in physics to model particle interactions of gasses. This means mathematically that the distributions of eigenvalues contain a factor that vanishes if two eigenvalues are close to each other, thereby allocating less and less probability to this event. The same holds for singular values as well. Hence, eigen-/singular values exhibit a tendency to spread themselves out evenly and do not have accumulation points. Following a line of argument in the spirit of [104] in combination with an approach from log-potential theory, Kabluchko, Prochno, and Thäle showed that for such random matrices the vector of the eigen-/singular values has respective distribution $\mathbf{C}_{n,p,f}$ and $\mathbf{U}_{n,p,f}$ on the Euclidean $\ell_p^n$-ball $\mathbb{B}_p^n$, with $f$ being the suitable repulsion factor between the eigen-/singular values of the random matrices (see, e.g., [6]).

The aim of this chapter is to put this last result into a wider context by investigating the eigenvalue and singular value distribution of random matrices that have the analogue distribution to $\mathbf{P}_{n,p,\mathbf{W}}$ on matrix $p$-balls. Using similar arguments as [62, 63], we will show that the vector of eigenvalues of such a random matrix also is $p$-radially distributed according to $\mathbf{P}_{n,p,\mathbf{W},f}$ on $\mathbb{B}_p^n$, with $f$ being the appropriate repulsion factor again, and the same holds for the vector of singular values on the non-negative segment of $\mathbb{B}_p^n$, denoted as $\mathbb{B}_{p,+}^n$. This connection paves the way to approach concentration phenomena on matrix $p$-balls via those on $\mathbb{B}_p^n$ with appropriately weighted distributions.

As an application of the connection just described, we study the large deviation behaviours of random elements in Euclidean and matrix $p$-balls. In case of Euclidean $\ell_p^n$-balls, the results of Kim and Ramanan [74] are of particular interest to us. For a random vector with distribution $\mathbf{C}_{n,p}$ they gave a large deviation principle for the empirical measure of its coordinates. Their findings are in the spirit of the theorem of Sanov [29, Theorem 2.1.10], as the corresponding rate function is given by the relative entropy (see (3.19)) perturbed by a $p$-th moment penalty. We want to expand on their results and give a large deviation principle for the empirical measure of a random vector with distribution $\mathbf{P}_{n,p,\mathbf{W}}$. We will show that, even though the distribution $\mathbf{P}_{n,p,\mathbf{W}}$ is highly dependent on the choice of $\mathbf{W}$, for certain classes of $\mathbf{W}$ the corresponding rate function will be universal to all $\mathbf{P}_{n,p,\mathbf{W}}$. The results of Kim and Ramanan have been further extended by Frühwirth and Prochno [35], who derived a Sanov-type large deviation principle for the empirical measure of random vectors uniformly distributed in Orlicz balls, which are a generalization of $\ell_p^n$-balls (see Remark 2.4.8).

In case of the matrix $p$-ball, an analogue result to that of Kim and Ramanan [74] has been given by Kabluchko, Prochno and Thäle [62]. They derived a large deviation principle for the empirical spectral measure, i.e., the empirical measure with respect to the eigenvalues or singular values of random matrices that are distributed according to the uniform distribution or the cone probability measure on the matrix $p$-ball. We will derive similar results for the analogue of $\mathbf{P}_{n,p,\mathbf{W}}$ on matrix $p$-balls and show a similar universality of the rate function. To do so, we will utilize the probabilistic representations for the eigenvalue and singular value distributions we derived beforehand.

Summarizing, the overall goals of this chapter are threefold. First, we want to expand the results from [13] to weighted $p$-radial distributions $\mathbf{P}_{n,p,\mathbf{W},f}$. This will be done in Section 3.2. Second, we want to show that for self-adjoint and non-self-adjoint random matrices, which are distributed according to the analogue of $\mathbf{P}_{n,p,\mathbf{W}}$ on matrix $p$-balls, the corresponding eigenvalue and singular value distributions are given by $\mathbf{P}_{n,p,\mathbf{W},f}$ on $\mathbb{B}_p^n$ (and its non-negative analogue on $\mathbb{B}_{p,+}^n$), with $f$ being the appropriate repulsion factor. This will be done in Section 3.3. And third, Sections 3.4 and 3.5 will then use the previous results to derive several large deviation principles for Euclidean and matrix $p$-balls, respectively. We will prove a large deviation principle for the empirical measure of the coordinates of a random vector with distribution $\mathbf{P}_{n,p,\mathbf{W}}$ on $\mathbb{B}_p^n$. Then we will show large deviation principles for the empirical spectral measures (for eigenvalues and singular values) of random matrices distributed according to the analogue of $\mathbf{P}_{n,p,\mathbf{W}}$ on matrix $p$-balls by using the representations of the eigenvalue and singular value distributions as $\mathbf{P}_{n,p,\mathbf{W},f}$ from Section 3.3 for suitable choices of $f$. The chapter will, however, begin in Section 3.1 by establishing the necessary notation and basic concepts.

Chapter 3 is partly based on the paper

- Kaufmann, T., and Thäle, C. [72]: Weighted $p$-radial distributions on Euclidean and matrix $p$-balls with applications to large deviations. *Journal of Mathematical Analysis and Applications*, (2022).

**Chapter 4:** In this chapter, we consider products of uniform random variables from the Stiefel manifold of orthonormal $k$-frames in $\mathbb{R}^n$ with $k \leq n$ and random vectors from the $\ell_p^n$-ball $\mathbb{B}_p^n$ with $p$-radial distributions $\mathbf{P}_{n,p,\mathbf{W}}$. The distribution of this product geometrically corresponds to the projection of the distributions $\mathbf{P}_{n,p,\mathbf{W}}$ on $\mathbb{B}_p^n$ onto a random $k$-dimensional subspace. We derive large deviation principles on the space of probability measures on $\mathbb{R}^k$ for sequences of such projections.

The setting of this chapter is a generalization of the one initiated by Kabluchko and Prochno [59] and can be described as follows. For $k \leq n$, the Stiefel manifold $\mathbb{V}_{n,k}$ is the set of all orthonormal $k$-frames in $\mathbb{R}^n$, i.e., the set of all $k$-tuples of orthonormal vectors $v_1, \dots, v_k$ in $\mathbb{R}^n$. Arranging these vectors into a $(k \times n)$ matrix $V$ with rows $v_1^T, \dots, v_k^T$, we have the identification

$$\mathbb{V}_{n,k} = \left\{ V \in \mathbb{R}^{k \times n} \colon VV^T = I_k \right\},$$

where $I_k$ denotes the $(k \times k)$ identity matrix. We denote by $\mathbf{U}_{n,k,\mathbb{V}}$ the uniform distribution on $\mathbb{V}_{n,k}$ and by $V_{n,k}$ the corresponding random variable. For random vectors $X^{(n)}$ taking values in $\mathbb{R}^n$ we may regard $V \in \mathbb{V}_{n,k}$ as a linear map $V \colon \mathbb{R}^n \to \mathbb{R}^k$ and study the distribution of the vectors $VX^{(n)} \in \mathbb{R}^k$, which we denote by

$$\mu_{VX^{(n)}}(A) := \mathbb{P}\left(VX^{(n)} \in A\right)$$

for any Borel set $A \subseteq \mathbb{R}^k$. In addition, we may also choose $V_{n,k} \in \mathbb{V}_{n,k}$ at random according to $\mathbf{U}_{n,k,\mathbb{V}}$. In this case, the distribution of $V_{n,k}X^{(n)}$, which we denote by

$$\mu_{V_{n,k}X^{(n)}}(A) := \mathbb{P}\left(V_{n,k}X^{(n)} \in A\right),$$

is a random probability measure on $\mathbb{R}^k$, that is, a random variable taking values in the space $\mathcal{M}_1(\mathbb{R}^k)$ of probability measures on $\mathbb{R}^k$, that is equipped with the topology of weak convergence. This can geometrically be interpreted as the projection of the distribution of the random vector $X^{(n)}$ onto a uniform random $k$-dimensional subspace.

We are interested in large deviation principles for the random probability measures $\mu_{V_{n,k}X^{(n)}}$, where $X^{(n)} \in \mathbb{B}_p^n$ with distribution $\mathbf{P}_{n,p,\mathbf{W}}$ for some Borel probability measure $\mathbf{W}$ on $[0, \infty)$. Kabluchko and Prochno [59] gave very general LDPs for random matrices in the orthogonal group and the Stiefel manifold, and showed an LDP for $k$-dimensional projections of the special case of the uniform distribution $\mathbf{U}_{n,p}$ on $\mathbb{B}_p^n$ as an application. Based on [59], the results of this chapter largely extend the set of projected distributions for which such an LDP is shown from the uniform distribution $\mathbf{U}_{n,p}$ to the aforementioned class of $p$-radial distributions $\mathbf{P}_{n,p,\mathbf{W}}$.

In particular, we will see that the large deviation behaviour observed by Kabluchko and Prochno [59] is universal for a large class of probability measures on $\mathbb{B}_p^n$. Moreover, we shall describe geometrically motivated distributions on $\mathbb{B}_p^n$ for which the LDP needs a suitable modification, which we also provide.

We should also delineate this chapter's content from the results shown by Kim and Ramanan in [76, Theorems 2.4 & 2.6], who have shown, among other results, LDPs for projections of uniform random vectors in $\mathbb{B}_p^n$ onto uniform random $k$-dimensional subspaces. By the same arguments as put forth in [59], we note that while the settings are quite similar, the key difference is in the object of study. In [76] it is the *projection point itself*, hence yielding an LDP on $\mathbb{R}^k$, whereas in both [59] and this chapter the object of study is the *projected distribution* on $\mathbb{R}^k$, thus the main result yields an LDP on the space $\mathcal{M}_1(\mathbb{R}^k)$ of probability measures on $\mathbb{R}^k$.

The chapter will start off in Section 4.1 by briefly listing the notation and background material we will need to formulate the central theorems, which in turn are presented in Section 4.2. Section 4.3 will then contain their respective proofs, which can be delineated into three steps.

Firstly, we use the probabilistic representation for random vectors in $\mathbb{B}_p^n$ with distribution $\mathbf{P}_{n,p,\mathbf{W}}$ to reformulate the target random measure $\mu_{V_{n,k}X^{(n)}}$ and show that this measure behaves asymptotically like a different, more simple measure $\tilde{\mu}_{V_{n,k}X^{(n)}}$. This is done by proving that their distance in the Lévy-Prokhorov metric (see (4.4)) tends to zero for all $V \in \mathbb{V}_{n,k}$. Secondly, it is shown that this asymptotic vanishing of the Lévy-Prokhorov metric between $\mu_{V_{n,k}X^{(n)}}$ and the simpler measure $\tilde{\mu}_{V_{n,k}X^{(n)}}$ is sufficient to infer a weak LDP for $\mu_{V_{n,k}X^{(n)}}$, if one is given for $\tilde{\mu}_{V_{n,k}X^{(n)}}$. Lastly, one then shows that every $\mu_{V_{n,k}X^{(n)}}$ lies in a closed and compact subset of $\mathcal{M}_1(\mathbb{R}^k)$, where the notion of a weak LDP coincides with that of a full LDP. Thus, one can infer a full LDP for $\mu_{V_{n,k}X^{(n)}}$ by carrying over the LDP for $\tilde{\mu}_{V_{n,k}X^{(n)}}$ established in [59] in the manner just described.

Chapter 4 is partly based on the paper

- Kaufmann, T., Sambale, H., and Thäle, C. [70]: Large deviations for uniform projections of $p$-radial distributions on $\ell_p^n$-balls. arXiv:2203.00476 (2022).

**Chapter 5:** This final chapter turns the focus from results in the realm of large deviations theory to results from *sharp* large deviations (SLD) theory. It has the distinct advantage over classical large deviations theory that it gives tail asymptotics on a non-logarithmic scale and can provide concrete and asymptotically exact tail estimates for specific $n \in \mathbb{N}$, which makes them significantly more useful for practical applications. Moreover, a lot of idiosyncrasies of the underlying distributions, that are drowned out on the LDP scale, are still visible on the SLD scale, thus giving a deeper understanding of the geometric interpretation of the quantities involved.

The chapter will begin with Section 5.1 by expanding on the brief outline of SLD theory in the introduction and presenting the results of Bahadur and Ranga Rao [10], who gave the original impetus for the theory. Further, the differences to large deviations theory are underlined, both regarding their goals and methodology. Specifically, the saddle point method is layed out, which will be used for the essential local density estimates in the subsequent proofs.

As Section 5.2 and Section 5.3 are based partly on individual research papers, their content and outline will be given here individually.

In Section 5.2 sharp large deviation results of Bahadur-Ranga Rao-type are provided for the $q$-norm of random vectors distributed on the $\ell_p^n$-ball $\mathbb{B}_p^n$ according to the cone probability measure $\mathbf{C}_{n,p}$ or the uniform distribution $\mathbf{U}_{n,p}$ for $1 \leq q < p < \infty$.

The behaviour of the $q$-norm $\|Z\|_q$ of a random vector $Z$ in $\mathbb{B}_p^n$ was first studied by Schechtman and Zinn [106], who derived concentration inequalities for $\|Z\|_q$ with $Z \sim \mathbf{C}_{n,p}$ and $Z \sim \mathbf{U}_{n,p}$ for $q > p$. This is closely related to the intersection volume of $t$-multiples of volume-normalized $\ell_p^n$-balls $\mathbb{D}_p^n := \mathrm{vol}_n(\mathbb{B}_p^n)^{-1/n}\mathbb{B}_p^n$, i.e., $\mathrm{vol}_n(\mathbb{D}_p^n \cap t\mathbb{D}_q^n)$ with $t \in [0, \infty)$, for which Schechtman and Schmuckenschläger [105] gave asymptotics for different values of $t$. Schechtman and Zinn [107] expanded their previous results in [106], by not only considering the $q$-norm, but also images of random vectors under Lipschitz functions in general. Thus, they gave concentration inequalities for $f(Z)$, with $Z \sim \mathbf{C}_{n,p}$ and $Z \sim \mathbf{U}_{n,p}$, $p \in [1, 2)$, and $f$ a Lipschitz function with respect to the Euclidean norm. Schmuckenschläger [108] provided a CLT for $\|Z\|_q$ with $Z \sim \mathbf{C}_{n,p}$ and $Z \sim \mathbf{U}_{n,p}$ and used it to refine the previous intersection results in [105] for all $t \in (0, \infty)$. Naor [92] gave concentration inequalities for $\|Z\|_q^q$ with $Z \sim \mathbf{C}_{n,p}$, showed that the total variation distance between $\mathbf{C}_{n,p}$ and the normalized surface measure $\sigma_{n,p}$ on $\mathbb{S}_p^{n-1}$ tends to zero proportional to $n^{-1/2}$, and used the previously mentioned results to show a concentration inequality for $\|Z\|_q^q$ with $Z \sim \sigma_{n,p}$. He also discussed how concentration results similar to Schechtman and Zinn [107] for $\|Z\|_q$ could already be derived from previous results of Gromov and Milman [44] for the concentration of Lipschitz functions on convex bodies. Kabluchko, Prochno and Thäle [61] gave a multivariate CLT for $(\|Z\|_{q_1}, \ldots, \|Z\|_{q_d})$ with $Z \sim \mathbf{U}_{n,p}$ in the spirit of [108] and also considered the asymptotics for the intersection volume of multiple $\ell_p^n$-balls, i.e., $\mathrm{vol}_n(\mathbb{D}_p^n \cap t_1\mathbb{D}_{q_1}^n \cap \cdots \cap t_d\mathbb{D}_{q_d}^n)$ with $t_i \in [0, \infty)$. This CLT was furthermore applied by the same authors to infer a central limit theorem for the length of $\mathbb{B}_p^n$ projected onto a line with uniform random direction. Moreover, they provided an LDP for $\|Z\|_q$ with $Z \sim \mathbf{C}_{n,p}$ and $Z \sim \mathbf{U}_{n,p}$. In a follow-up paper [65], the same authors showed a CLT for $\|Z\|_q$ with the distribution of $Z$ taken from the class $\mathbf{P}_{n,p,\mathbf{W}}$ of $p$-radial distributions established in [13]. Finally, they gave a moderate and a large deviation principle for $\|Z\|_q$ with $Z \sim \mathbf{P}_{n,p,\mathbf{W}}$.

Recently, a new tool from large deviations theory was introduced to asymptotic geometric analysis by Liao and Ramanan [85]. They gave sharp large deviation results in the spirit of Bahadur and Ranga Rao [10] and Petrov [95] for the projections of random points in $\ell_p^n$-balls with distributions $\mathbf{C}_{n,p}$ and $\mathbf{U}_{n,p}$ onto a fixed one-dimensional subspace.[1] Other works in asymptotic geometric analysis have also employed methods from sharp large deviations theory as well, such as Kabluchko and Prochno [60], who derived asymptotic volumes for generalizations of $\ell_p^n$-balls, known as Orlicz balls (see Remark 2.4.8), and showed a Schechtman-Schmuckenschläger-type result by considering intersection volumes of Orlicz balls. Their results on Orlicz balls were then expanded upon by Alonso-Guiterréz and Prochno in [3], who gave the exact asymptotic volume of Orlicz balls and provided thin-shell concentrations for them, augmenting their results into sharp asymptotics under certain conditions.

Section 5.2 will follow closely in the footsteps of Liao and Ramanan [85] and establish SLD results for the $q$-norms of random vectors with distribution $\mathbf{C}_{n,p}$ and $\mathbf{U}_{n,p}$. Furthermore, we will use these results to expand on works of Schechtman and Schmuckenschläger [105], Schmuckenschläger [108], and Kabluchko, Prochno and Thäle [61] for intersection volumes of $\ell_p^n$-balls by giving sharp asymptotics for $\mathrm{vol}_n(\mathbb{D}_p^n \cap t\mathbb{D}_q^n)$ at a considerably improved rate for $1 \leq q < p < \infty$ and $t > C(p, q)$ bigger than some con-

[1]Note that the case for $\mathbf{U}_{n,p}$ was originally included in the second preprint version of the work of Liao and Ramanan [85] (arXiv:2001.04053v2), but has subsequently been removed in the current third preprint version (arXiv:2001.04053v3) to be contained in another, yet unpublished paper. Thus, while we generally refer to the third preprint version of [85] (as listed in the references), those references regarding $\mathbf{U}_{n,p}$ necessarily concern the second version. Where we refer to specific results therein, we will hence write [85, v2].

stant dependent on $p$ and $q$. Additionally, we will also apply our results for $\ell_p^n$-spheres to retain sharp asymptotics for the length of the projection of an $\ell_p^n$-ball onto the line spanned by a uniform random direction.

Let us give a brief outline of Section 5.2: In Section 5.2.1 some necessary notation and definitions will be provided and we will recapitulate some relevant preexisting results, followed by a short discussion on the Weingarten map in Section 5.2.2. This will be needed for the geometric Laplace integration results used later in the chapter. In Section 5.2.3 we will present the main results regarding the $q$-norms of random vectors on $\ell_p^n$-spheres and $\ell_p^n$-balls and outline the idea of the two central proofs. Section 5.2.4 and Section 5.2.5 will contain the applications of the main results to intersection volumes and random projections of $\ell_p^n$-balls mentioned above. In Section 5.2.6 we will reformulate the target probabilities from the main results in terms of useful probabilistic representations, using well-established representations of random vectors in $\ell_p^n$-balls of Schechtman and Zinn [106] and Rachev and Rüschendorf [99]. In Section 5.2.7 local density approximations of these probabilistic representations will be provided. In Sections 5.2.8 and 5.2.9 we will prove the SLD results for $\ell_p^n$-spheres and $\ell_p^n$-balls, respectively, by integrating over the density estimates. For that, we will utilize some geometric results for asymptotic expansions of Laplace integrals from Andriani and Baldi [7] and Breitung and Hohenbichler [19], thereby finishing Section 5.2.

In the second major part of this chapter, Section 5.3, we consider the $p$-generalized arithmetic-geometric mean ($p$-AGM) inequality for vectors chosen randomly from the $\ell_p^n$-ball. For $n \in \mathbb{N}$ and $x_1, \dots, x_n \in \mathbb{R}^n$ the classic AGM inequality states that

$$\left(\prod_{i=1}^n |x_i|\right)^{1/n} \leq \frac{1}{n}\sum_{i=1}^n |x_i|.$$

Additionally, for $p > 0$ the $p$-AGM inequality expands the above for the $p$-generalized mean, i.e., for $(x_1, \dots, x_n) \in \mathbb{R}^n, n \in \mathbb{N}$, we have

$$\left(\prod_{i=1}^n |x_i|\right)^{1/n} \leq \left(\frac{1}{n}\sum_{i=1}^n |x_i|^p\right)^{1/p}.$$

It was shown by Gluskin and Milman [38] that for a random vector $X^{(n)} \in \mathbb{R}^n$ uniformly distributed on the standard $(n-1)$-dimensional unit sphere $\mathbb{S}_2^{n-1}$ in $\mathbb{R}^n$, one can reverse the $p$-AGM inequality for $p = 2$ up to a scalar constant with high probability, which was then extended to $p = 1$ by Aldaz [1, 2]. Kabluchko, Prochno, and Vysotsky [66] provided a CLT and an LDP for the ratio of the two sides of the $p$-AGM inequality for any $p \in [1, \infty)$ and $X^{(n)}$ with distribution $\mathbf{U}_{n,p}$ or $\mathbf{C}_{n,p}$ on $\mathbb{B}_p^n$. Finally, Thäle [111] expanded the results of [66] to a CLT and a moderate deviation principle (MDP) for the ratio of the two sides of the $p$-AGM inequality with the corresponding random vector $X^{(n)} \in \mathbb{B}_p^n$ having distribution $\mathbf{P}_{n,p,\mathbf{W}}$ in the spirit of [13], which includes $\mathbf{U}_{n,p}$ and $\mathbf{C}_{n,p}$ as special cases. However, the arguments of Thäle show that the properties of interest of a random vector $X^{(n)} \in \mathbb{B}_p^n$ are independent of the $p$-radial component of its distribution, as long as the directional distribution is given by $\mathbf{C}_{n,p}$ and its $p$-radial distribution has no atom at zero (see (5.69)).

The purpose of Section 5.3 is to develop further the LDP of [66] into SLD results in the spirit of Bahadur and Ranga Rao [10]. For a random vector $X^{(n)} \in \mathbb{B}_p^n$ with directional distribution $\mathbf{C}_{n,p}$ in the sense of (5.69) we now want to give sharp asymptotics for the probability of the ratio of the two sides of the $p$-AGM inequality being bigger than a constant $\theta \in [0, 1]$. We thereby provide concrete and asymptotically exact estimates on a non-logarithmic scale for the probability of the inequality being improvable or reversible up to a constant, respectively.

The main results regarding the sharpening of the $p$-AGM inequality are stated right at the beginning of Section 5.3. Since their proof orients itself heavily on the proof in Section 5.2, it will also contain three parts: the first part provides a probabilistic representation for the ratios of the two sides of the $p$-AGM inequality in Section 5.3.1, the second part gives an asymptotic density estimate for this probabilistic representation in Section 5.3.2, and the third part consists of the final proof of the section's main result by integrating over said density estimate in Section 5.3.3.

Chapter 5 is partly based on the papers

- Kaufmann, T. [69]: Sharp asymptotics for $q$-norms of random vectors in high-dimensional $\ell_p^n$-balls. *Modern Stochastics: Theory and Applications* (2021),
- Kaufmann, T., and Thäle, C. [71]: Sharpening the probabilistic arithmetic-geometric mean inequality. arXiv:2112.04340 (2021).

# Chapter 2

# Preliminaries and notation

In this chapter basic notation and foundational concepts of probability and geometry are established. Specifically, the basics of large deviations theory are introduced, as they will play an integral role throughout the chapters.

## 2.1 Notation and basic definitions

Let $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{C}$ be the natural, real, and complex numbers, respectively, and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. For $d \in \mathbb{N}$ denote by $\mathbb{R}^d$ the $d$-dimensional Euclidean space with the standard scalar product $\langle \cdot\,, \cdot \rangle$, which induces the Euclidean norm $\| \cdot \|_2$, analogue for $\mathbb{C}^d$. Further, let $\mathbb{R}_+ := [0, \infty)$ be the non-negative real numbers and $\mathbb{R}_+^d := [0, \infty)^d$ the non-negative $d$-dimensional Euclidean space. For a complex number $z \in \mathbb{C}$, denote by $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$ its real and imaginary component, respectively. For ease of notation we write $(x_1, \dots, x_d) \in \mathbb{R}^d$ for a column vector and for $x, y \in \mathbb{R}^d$, we write their product $x^T y$ as $xy$, skipping the explicit transpose. The same holds for matrices and matrix multiplication. For a general set $A$ define the indicator function

$$\mathbf{1}_A(x) = \begin{cases} 1 & : x \in A \\ 0 & : x \notin A, \end{cases}$$

and denote by $\partial A, \overline{A}, A^\circ$, and $A^c$ respectively its boundary, closure, interior, and complement with respect to the underlying topology. Set $\mathcal{B}(\mathbb{R}^d)$ to be $\sigma$-field of Borel sets in $\mathbb{R}^d$ and $\mathrm{vol}_d$ to be the $d$-dimensional Lebesgue measure on $\mathbb{R}^d$.

While we have given the definition of $L_p$ in (1.2), we set a more general version as

$$L_p\left(\mathbb{R}^d\right) := \left\{ f : \mathbb{R}^d \to \mathbb{R} \, : \int\limits_{\mathbb{R}^d} |f(x)|^p \,\mathrm{d}x < \infty \right\}$$

for $d \in \mathbb{N}$ and $p < \infty$ (for $p = \infty$ consider the essential supremum of $f$, analogue to (1.3)). For a function $f : \mathbb{R}^d \to \mathbb{R}$ we define the effective domain of $f$ as

$$\mathrm{Dom}(f) := \left\{ x \in \mathbb{R}^d : f(x) < \infty \right\}.$$

Further, for $c \in \mathbb{R}$ we call $\{x \in \mathbb{R}^d : f(x) = c\}$ and $\{x \in \mathbb{R}^d : f(x) \leq c\}$ the $c$-level set and $c$-sublevel set, respectively.

We say a set $A$ is convex if it contains all convex combinations of its elements, that is, $\lambda x + (1-\lambda)y \in A$ for all $x, y \in A, \lambda \in [0,1]$. A compact and convex subset of $\mathbb{R}^d$ with non-empty interior is called a convex body. A real-valued function $f : \mathbb{R}^d \to \mathbb{R}$ is called convex if the set above its graph is convex, i.e., if $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$ for all $x, y \in \mathbb{R}^d$ and $\lambda \in [0,1]$. If this inequality holds strictly for all arguments $x \neq y$ and $\lambda \in (0,1)$, the function $f$ is called strictly convex. A real-valued function $f$ is called (strictly) concave, on the other hand, if $(-f)$ is (strictly) convex.

Next we shall set down some notation for different derivatives. For $g : \mathbb{R}^d \to \mathbb{R}^k$, we denote by $J_x g(x^*)$ the Jacobian of $g$ with respect to $x$ evaluated at $x^* \in \mathbb{R}^d$, and for $f : \mathbb{R}^d \to \mathbb{R}$ by $\nabla_x f(x^*)$ and $\mathcal{H}_x f(x^*)$ the gradient and Hessian of $f$ with respect to $x$ evaluated at $x^* \in \mathbb{R}^d$, respectively, and use the shorthand derivative notation

$$f_{[i_1,\ldots,i_d]}(x^*) = \frac{\partial^{i_1}}{\partial x_1^{i_1}} \cdots \frac{\partial^{i_d}}{\partial x_d^{i_d}} f(x)\big|_{x=x^*}. \tag{2.1}$$

Given a convex real-valued function $f : \mathbb{R}^d \to \mathbb{R}$, we denote by

$$f^*(x) = \sup_{\tau \in \mathbb{R}^d} \left[ \langle x, \tau \rangle - f(\tau) \right] \tag{2.2}$$

its Legendre-Fenchel transform. We shall briefly list two of its useful properties. For a proof of the first we refer the reader to [103, Theorem 26.5] and for a proof of the second to [37, Chapter 4, Section 18].

**Lemma 2.1.1** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex and lower semi-continuous function with* $\mathrm{Dom}(f)^\circ \neq \emptyset$ *and* $\inf(f) > -\infty$*. Additionally, let $f$ be differentiable on* $\mathrm{Dom}(f)^\circ$ *and* $\lim_{n\to\infty} \|\nabla_x f(x_n)\|_2 = +\infty$ *for any sequence $(x_n)_{n\in\mathbb{N}}$ converging to a point on* $\partial\mathrm{Dom}(f)$*. Then the following holds:*

(1) *For any $x \in \mathrm{Dom}(f^*)$ the supremum in the Legendre-Fenchel transform $f^*(x)$ is uniquely attained in*

$$\tau(x) = (\nabla_\tau f)^{-1}(x) \in \mathrm{Dom}(f),$$

*with $(\nabla_\tau f)^{-1}$ denoting the inverse of the derivative of $f$. Hence*

$$f^*(x) = \langle x, \tau(x)\rangle - f(\tau(x)) = \big\langle x, (\nabla_\tau f)^{-1}(x)\big\rangle - f\big((\nabla_\tau f)^{-1}(x)\big).$$

(2) *The Legendre-Fenchel transform is an involution on $\mathrm{Dom}(f)^\circ$, that is, for all $x \in \mathrm{Dom}(f)^\circ$ we have $f(t) = (f^*)^*(t)$.*

The other transform we want to introduce is the Fourier transform. For a function $f : \mathbb{R}^d \to \mathbb{R}$ with $f \in L_1(\mathbb{R}^d)$ we define its Fourier transform at $t \in \mathbb{R}^d$ to be

$$\mathcal{F}(f)(t) := \int\limits_{\mathbb{R}^d} e^{\langle it,y\rangle}\, f(y)\,\mathrm{d}y. \tag{2.3}$$

We choose this normalization of the Fourier transform such that in its applications to densities of probability distributions it coincides with the corresponding characteristic function. This way we can easily use results from both Fourier analysis and probability theory without intermediate renormalization. The inverse Fourier transform, denoted by $\mathcal{F}^{-1}(\cdot)$, of a function $f : \mathbb{R}^d \to \mathbb{R}$ at some $y \in \mathbb{R}^d$ is then defined as

$$\mathcal{F}^{-1}(\mathcal{F}(f))(y) := \frac{1}{(2\pi)^d} \int\limits_{\mathbb{R}^d} e^{-\langle it,y\rangle}\, \mathcal{F}(f)(t)\,\mathrm{d}t.$$

If the inverse Fourier transform $\mathcal{F}^{-1}(\mathcal{F}(f))$ of $\mathcal{F}(f))$ is itself absolutely integrable, then the Fourier inversion theorem states $\mathcal{F}^{-1}(\mathcal{F}(f))(y) = f(y)$ (cf. [110, Theorem 1.9]).

Closing this section, we define Euler's gamma and beta function and provide a well-known approximation result for the former. For $x, y \in \mathbb{R}_+$ define

$$\Gamma(x) := \int\limits_0^\infty t^{x-1}e^{-t}\,\mathrm{d}t \qquad \text{and} \qquad \mathrm{B}(x,y) := \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

to be Euler's gamma function and beta function, respectively. For $x > 0$, Stirling's approximation formula for the gamma function states that

$$\Gamma(x) = \sqrt{\frac{2\pi}{x}}\Big(\frac{x}{e}\Big)^x\Big(1 + \mathcal{O}\Big(\frac{1}{x}\Big)\Big), \tag{2.4}$$

where $\mathcal{O}\big(\frac{1}{x}\big)$ denotes an approximation error with asymptotic upper bound $\frac{1}{x}$.

## 2.2 Probability

Given a probability space $(\Omega, \Sigma, \mathbb{P})$, a measurable space $(E, \mathcal{E})$, and a random variable $X : \Omega \to E$, we call $\mu := \mathbb{P} \circ X^{-1}$ the distribution of $X$ and write $X \sim \mu$ to denote that for any $A \in \mathcal{E}$ it holds that

$$\mathbb{P}(X \in A) = \int_E \mathbf{1}_A(x)\, \mu(\mathrm{d}x).$$

The distribution $\mu = \mathbb{P} \circ X^{-1}$ of a random variable $X$ is also often denoted as $\mathcal{D}(X)$. For a real-valued random variable $X : \Omega \to \mathbb{R}$ with distribution $\mu$ we write

$$\mathbb{E}[X] := \int_{\mathbb{R}} x\, \mu(\mathrm{d}x)$$

for the expectation of $X$, often writing $\mathbb{E}X$ if it is clear from context what expectation is considered, and $\mathrm{Var}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2]$ for the variance of $X$ if $\mathbb{E}[X] < \infty$. For some $r > 0$ we denote by

$$\mathbb{E}[X^r] := \int_{\mathbb{R}} x^r\, \mu(\mathrm{d}x) \qquad \text{and} \qquad \mathbb{E}[|X|^r] := \int_{\mathbb{R}} |x|^r\, \mu(\mathrm{d}x) \tag{2.5}$$

the $r$-th moment and $r$-th absolute moment of $X$, respectively. More generally, we denote by $m_r(\mu)$ the $r$-th absolute moment of the probability measure $\mu$, defined as above. The expectation of a random vector $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$ is the vector of its coordinate expectations $\mathbb{E}[X] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d])$ and analogue holds for its variance. For a probability distribution $\mu$ on $\mathbb{R}$ let $\mu^{\otimes d}$ denote its $d$-fold product measure on $\mathbb{R}^d$. Let $X$ be a random vector in $\mathbb{R}^d$, then for $\tau \in \mathbb{R}^d$

$$\varphi_X(\tau) := \mathbb{E}\big[e^{\langle \tau, X\rangle}\big] \qquad \text{and} \qquad \Lambda_X(\tau) := \log \mathbb{E}\big[e^{\langle \tau, X\rangle}\big]$$

are the moment generating function and cumulant generating function of $X$, respectively, where we often omit the index when it is clear from context.

Let us present some useful properties of the cumulant generating function in the following lemma. The statements therein and their proof can be found in [73, Lemma 1.1.4] and its subsequent proof. Alternatively, they follow from the standard properties of the moment generating function (see, e.g., [27, Theorem 5.4]) and cumulant generating function (see, e.g., [29, Lemma 2.2.31]), together with the properties of the Legendre-Fenchel transform laid out in Lemma 2.1.1.

**Lemma 2.2.1** *Let $X$ be a random vector in $\mathbb{R}^d$ that is not almost surely constant and has cumulant generating function $\Lambda_X$. Then it holds that*

*(1)* $\Lambda_X$ *is convex and lower semi-continuous,*

*(2)* $\Lambda_X$ *is infinitely differentiable and strictly convex on* $\mathrm{Dom}(\Lambda_X)^\circ$*,*

*(3)* $\Lambda_X^*$ *is infinitely differentiable and strictly convex on* $\mathrm{Dom}(\Lambda_X^*)^\circ$*,*

*(4) for* $x \in \mathrm{Dom}(\Lambda_X^*)^\circ$ *there exists a unique* $\tau(x) \in \mathrm{Dom}(\Lambda_X)^\circ$*, such that*

$$\Lambda_X^*(x) = \langle x, \tau(x)\rangle - \Lambda_X(\tau(x))$$

*with* $\tau(x) = (\nabla_\tau \Lambda_X)^{-1}(x)$*,*

*(5) the interior of the effective domain of* $\Lambda_X^*$ *is given by*

$$\mathrm{Dom}(\Lambda_X^*)^\circ = \nabla_\tau \Lambda_X(\mathrm{Dom}(\Lambda_X)^\circ).$$

**Remark 2.2.2** For a random vector $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$ one defines the characteristic function to be the function mapping some $t \in \mathbb{R}^d$ to $\mathbb{E}\big[e^{\langle it, X\rangle}\big]$. Note here that, on the one hand, this is just the moment generating function of $X$ at a complex argument $it \in \mathbb{C}^d$, i.e., $\varphi_X(it) = \mathbb{E}\big[e^{\langle it, X\rangle}\big]$. On the other hand, if $X$ possesses a density $f$, due to our chosen normalization in (2.3) the characteristic function is simply the Fourier transform of $f$, that is, $\mathcal{F}(f)(t) = \mathbb{E}\big[e^{\langle it, X\rangle}\big]$. Thus, we have

$$\varphi_X(it) = \mathbb{E}\Big[e^{\langle it, X\rangle}\Big] = \mathcal{F}(f)(t),$$

meaning that these concepts can be used interchangeably, which will be beneficial in the proofs within the following chapters.

For two random variables $X, Y$ with finite expectations we denote by $\mathrm{Cov}[X, Y] := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$ their covariance and for an $\mathbb{R}^d$-valued random vector $X$ with finite coordinate-wise expectations we write $\mathrm{Cov}[X]$ for its covariance matrix, i.e., the $(d \times d)$ matrix of covariances of its components. For two random variables $X, Y$ with the same distribution we write $X \overset{\mathcal{D}}{=} Y$.

We now consider a few specific distributions of random variables. We say a real-valued random variable $X$ is gamma distributed with shape $a > 0$ and rate $b > 0$ if its distribution has density

$$f_{\mathbf{G}}(x) := \frac{b^a}{\Gamma(a)}\, x^{a-1}\, e^{-bx}\, \mathbf{1}_{(0,\infty)}(x), \qquad x \in \mathbb{R},$$

with respect to the Lebesgue measure on $\mathbb{R}$. We denote this by $X \sim \mathbf{G}(a,b)$. For $a=1$ we call this an exponential distribution and write $X \sim \mathbf{E}(b)$. Similarly, we say a real-valued random variable $X$ is beta distributed with parameters $a,b>0$ if its distribution has Lebesgue density

$$f_{\mathbf{B}}(x) := \frac{1}{B(a,b)}\, x^{a-1}\,(1-x)^{b-1}\,\mathbf{1}_{(0,1)}(x), \qquad x \in \mathbb{R},$$

and write $X \sim \mathbf{B}(a,b)$. Note the well-known relation between the gamma and beta distribution, stating that for independent $X \sim \mathbf{G}(a,c)$ and $Y \sim \mathbf{G}(b,c)$ it holds that

$$\frac{X}{X+Y} \sim \mathbf{B}(a,b). \tag{2.6}$$

Finally, we say a real-valued random variable $X$ has a generalized Gaussian distribution if its distribution has Lebesgue density

$$f_{\mathrm{gen}}(x) := \frac{b}{2\,a\,\Gamma\left(\frac{1}{b}\right)}\, e^{-(|x-\mu|/a)^b}, \qquad x \in \mathbb{R},$$

where $\mu \in \mathbb{R}$ and $a,b>0$, and denote this by $X \sim \mathbf{N}_{\mathrm{gen}}(\mu,a,b)$.

As mentioned in the introduction and as will be shown in further detail in Section 2.4.1, the generalized Gaussian distributions are intimately connected to the geometry of $\ell_p^n$-balls and serve as the essential building block when constructing useful probabilistically equivalent representations for random vectors in with a wide variety of distributions. For these constructions we will be using the specific generalized Gaussian distributions $\mathbf{N}_p := \mathbf{N}_{\mathrm{gen}}(0, p^{1/p}, p)$ (for $p \in (0,\infty)$) with density

$$f_{\mathbf{N}_p}(x) := \frac{1}{2\,p^{1/p}\,\Gamma\left(1+\frac{1}{p}\right)}\, e^{-|x|^p/p}, \qquad x \in \mathbb{R},$$

and $\tilde{\mathbf{N}}_p := \mathbf{N}_{\mathrm{gen}}(0,1,p)$, which has density

$$f_{\tilde{\mathbf{N}}_p}(x) := \frac{1}{2\,\Gamma\left(1+\frac{1}{p}\right)}\, e^{-|x|^p}, \qquad x \in \mathbb{R}.$$

We will refer to random variables $X$ with distribution $\mathbf{N}_p$ or $\tilde{\mathbf{N}}_p$ as having *$p$-generalized Gaussian distribution* (it will always be clear from context which concrete distribution we will be referring to). Note that for $p=2$ it holds that $\mathbf{N}_2$ is simply the standard normal distribution.

**Remark 2.2.3** In the literature both $\mathbf{N}_p$ and $\tilde{\mathbf{N}}_p$ are used. For example, the papers [4, 5, 36, 69, 70, 71, 74] consider $\mathbf{N}_p$, while [13, 62, 63, 72, 106] work with $\tilde{\mathbf{N}}_p$. This results merely in different normalization factors when constructing probabilistic representations. The results within this thesis are also reflective of this, as in Chapter 3 we use $\tilde{\mathbf{N}}_p$, whereas in Chapter 4 and Chapter 5 we employ $\mathbf{N}_p$, as this is most appropriate to the respective research contexts of the related publications [72] and [69, 70, 71].

For $X \sim \mathbf{N}_p$ and $r > 0$ the $r$-th absolute moment of $X$ is given by

$$\mathbb{E}\left[|X|^r\right] = \frac{p^{r/p}}{r+1} \frac{\Gamma\left(1 + \frac{r+1}{p}\right)}{\Gamma\left(1 + \frac{1}{p}\right)}, \tag{2.7}$$

which can be seen from [61, Lemma 4.1].

## 2.3 Large deviations theory

In this chapter large deviations theory is briefly introduced. We begin with the basic ideas of the theory, including the general definition of a large deviation principle, and then present the methodological toolbox of large deviations theory, i.e., a collection of useful results we will often rely on throughout this thesis. For further background on large deviations and proofs of the related results presented in this section we refer the reader to the monographs [29, 30, 67].

### 2.3.1 Idea of the theory

At the end of the general introduction in Section 1.1 a rough draft of the goals of large deviations theory was given, i.e., that it intends to - roughly speaking - characterize the probabilistic decay of sequences of rare events. But let us start with a well-known concrete example and then expand out to more general settings.

For a sequence $(X_n)_{n\in\mathbb{N}}$ of i.i.d. real-valued random variables with $\mathbb{E}[X_1] = \mu < \infty$ denote by $(S_n)_{n\in\mathbb{N}}$ the sequence of their partial sums

$$S_n := \sum_{i=1}^{n} X_i.$$

We know from the law of large numbers that the sequence $\left(\frac{1}{n}S_n\right)_{n\in\mathbb{N}}$ of the empirical averages converges almost surely to $\mu$ as $n$ tends to infinity. Hence, we know that the probability of a deviation of $S_n$ from its expectation $n\mu$ of order $n$ converges to 0 in $n \in \mathbb{N}$, that is, for any $x > 0$ it holds that

$$\lim_{n\to\infty} \mathbb{P}\left(|S_n - n\mu| \geq nx\right) = 0.$$

However, we do not know how exactly this convergence to zero – that is, this probabilistic decay – of the deviation probabilities takes place, both in terms of the sequence parameter $n \in \mathbb{N}$ and the deviation size $x$. The CLT tells us, given $\sigma^2 := \mathrm{Var}[X_1] \in (0, \infty)$, that $\frac{1}{\sqrt{n}}(S_n - n\mu)$ converges in distribution to a normal distribution $\mathcal{N}(0, \sigma^2)$ as $n$ tends to infinity, and therefore

$$\lim_{n\to\infty} \mathbb{P}\left(|S_n - n\mu| \geq \sqrt{n}x\right) = 1 - \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-x}^{x} e^{-y^2/2\sigma^2} \mathrm{d}y.$$

This means the probability for a deviation of $S_n$ from $n\mu$ of order $\sqrt{n}$ is approximately Gaussian. The Berry-Esseen Theorem (see [114, Theorem 2.1.3]) additionally yields that the error of the Gaussian approximation given by the CLT is of order $n^{-1/2}$, meaning we also have a rate of convergence for the above. Deviations of order up to $\sqrt{n}$ are often refered to as "Gaussian fluctuations" or "normal deviations", as they are characterized via the CLT. Beyond order $\sqrt{n}$ one talks about *moderate deviations* (of order between $\sqrt{n}$ and $n$) and *large deviations* (of order $n$ and higher), the latter of which is the main area of focus of this thesis.

Note that the CLT yields a limiting normal distribution - and hence Gaussian behaviour of fluctuations of order up to $\sqrt{n}$ - regardless of the underlying distribution of the involved random variables $X_i$, given the necessary moment-conditions. The CLT is therefore very universal in this regard.

Turning our focus to *large* deviations, in the previous context of the empirical average one can rewrite deviation probabilities for some $x > \mathbb{E}[X_1]$ using the moment generating function $\varphi_X$ of the $X_i$. Assume that $\varphi_X(\tau) < \infty$ for all $\tau \in \mathbb{R}$, then it follows from Markov's inequality that for $\tau > 0$,

$$\mathbb{P}\left(S_n \geq nx\right) = \mathbb{P}\left(e^{\tau S_n} \geq e^{\tau nx}\right) \leq e^{-\tau nx}\mathbb{E}\left[e^{\tau S_n}\right] = e^{-\tau nx}\varphi_X(\tau)^n = e^{-n[\tau x - \log\varphi_X(\tau)]}$$

(cf. [81, Equation (1.1.2)]). Since one intends to give the best possible estimate of the above, optimizing for $\tau > 0$ gives

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(S_n \geq nx\right) \leq -\sup_{\tau>0}\left[\tau x - \log\varphi_X(\tau)\right] = \Lambda_X^*(x), \tag{2.8}$$

where $\Lambda_X^*$ is the Legendre-Fenchel transform of the cumulant generating function $\Lambda_X(\tau) = \log \varphi_X(\tau)$ of the $X_i$ (the supremum can be considered over all $\tau \in \mathbb{R}$, since for $x > 0$ we have $\tau(x) > 0$, as shown in [81, Lemma 1.4.1], so the above reformulation is admissible). For a given $x > 0$ the corresponding $\tau(x) \in \mathbb{R}$ at which the supremum is attained is the solution to the equation

$$x = \frac{\partial}{\partial \tau} \Lambda_X(\tau(x)) = \frac{\mathbb{E}\left[X e^{\tau(x) X}\right]}{\mathbb{E}\left[e^{\tau(x) X}\right]}. \tag{2.9}$$

We have seen in Lemma 2.2.1 that for $X$ not being almost surely constant and $x \in \mathrm{Dom}(\Lambda_X^*)^\circ$ such a $\tau(x) \in \mathrm{Dom}(\Lambda_X)^\circ$ exists and is unique. Thereby we already have a way to describe the probabilistic decay of deviations of order $n \in \mathbb{N}$ via

$$\mathbb{P}(S_n \geq nx) \leq e^{-n[\Lambda_X^*(x) + o(1)]},$$

again denoting by $o(1)$ a sequence that tends to zero as $n \to \infty$.

We note that, as remarked in [81, p.2], an additional take-away of (2.9) is that transforming $X$ with the density $e^{\tau(x)X}/\varphi_X(\tau(x))$ yields a new random variable with expectation $x$. This fact, or rather its multi-dimensional counterpart, will be an essential tool in Chapter 5, therein referred to as an exponential measure tilt.

In addition to (2.8), one can show that $\Lambda_X^*$ is also a lower bound for logarithmic deviation probabilities of the empirical average, that is, that in the limit there holds an equality. This, in fact, is the well-known theorem of Cramér (cf. [30, Theorem I.4]).

**Proposition 2.3.1** *Let $(X_i)_{i \in \mathbb{N}}$ be a family of i.i.d. random variables such that for all $t \in \mathbb{R}$ it holds that $\varphi_X(t) < \infty$. Then for every $x > \mathbb{E}[X_1]$*

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} S_n \geq x\right) = -\Lambda_X^*(x).$$

As shown in [29, Corollary 6.1.6] the condition that $\varphi_X(t) < \infty$ for all $t \in \mathbb{R}$ in the theorem of Cramér can be relaxed to $\varphi_X$ being merely finite in a neighbourhood around zero, while still retaining the same results. Since we will only work with the multi-dimensional version of Cramér's theorem, which will be presented in Proposition 2.3.3, we will only include said relaxation of the condition therein, as the one-dimensional version serves just to illustrate the core concepts of the theorem and large deviations theory overall.

Hence, we know that for sequences of real-valued i.i.d. random variables with sufficiently finite exponential moments we can characterize the large deviation behaviour of their empirical averages via the Legendre-Fenchel transform of their cumulant generating function, i.e.,

$$\mathbb{P}\left(\frac{1}{n}S_n \geq x\right) = e^{-n\left[\Lambda^*_{\mathbb{X}}(x)+o(1)\right]}. \tag{2.10}$$

Of course there are other sequences of random variables in other spaces for which one would also like to characterize the large deviation behaviour in a manner such as (2.10), i.e., describe the probabilistic decay of large deviation events via two functions, one depending on the sequence parameter and one on the size of the deviation. In general, one would like such a description for general sequences of probability distributions on Polish spaces as well, that is, separable and completely metrizable topological spaces. Finding the functions that allow this is one of the main goals of large deviations theory, and the capability of describing logarithmic probabilities in such a fashion is referred to as a *large deviation principle* (LDP). While the theorem of Cramér was established in the 1930's, the following formal and substantially more general definition of an LDP is due to work of Varhadan (see [112, 113]), who incorporated previous results like those of Cramér into an overarching field of study. If $\mathbb{X}$ is a topological space, we write $\mathcal{B}(\mathbb{X})$ for the $\sigma$-field of Borel sets in $\mathbb{X}$. If $\mathbb{X}$ is separable and completely metrizable, we call $\mathbb{X}$ a Polish space.

**Definition 2.3.2** Let $\mathbb{X}$ be a Polish space equipped with the Borel $\sigma$-field $\mathcal{B}(\mathbb{X})$ and $(\mathbf{P}_n)_{n\in\mathbb{N}}$ a sequence of probability measures on $\mathbb{X}$. We say that $(\mathbf{P}_n)_{n\in\mathbb{N}}$ satisfies a large deviation principle on $\mathbb{X}$ if there are two functions $s:\mathbb{N}\to(0,\infty)$, such that $s(n)\to\infty$ as $n\to\infty$, and $\mathcal{I}:\mathbb{X}\to[0,\infty]$, such that $\mathcal{I}$ is lower semi-continuous and

$$a)\quad \liminf_{n\to\infty}\frac{1}{s(n)}\log\mathbf{P}_n(O) \quad\geq\quad -\mathcal{I}(O) \qquad\qquad \text{for all } O\in\mathcal{B}(\mathbb{X}) \text{ open,}$$

$$b)\quad \limsup_{n\to\infty}\frac{1}{s(n)}\log\mathbf{P}_n(C) \quad\leq\quad -\mathcal{I}(C) \qquad\qquad \text{for all } C\in\mathcal{B}(\mathbb{X}) \text{ closed,}$$

where for $B\in\mathcal{B}(\mathbb{X})$ we define $\mathcal{I}(B) := \inf_{x\in B}\mathcal{I}(x)$. We call $s(\cdot)$ the speed and $\mathcal{I}(\cdot)$ the rate function. We say that $\mathcal{I}(\cdot)$ is a good rate function, if it has compact sublevel sets. Further, we say that $(\mathbf{P}_n)_{n\in\mathbb{N}}$ satisfies a weak LDP on $\mathbb{X}$ if a) holds but b) only needs to hold for compact sets.

For sequences $(X_n)_{n\in\mathbb{N}}$ of random variables one applies the above definition to the sequence of their distributions. Hence, we see that the theorem of Cramér in Proposition 2.3.1 states that the sequence of empirical averages of real-valued i.i.d. random variables with finite exponential moments satisfies an LDP with speed $n$ and good rate function given by the Legendre-Fenchel transform of the cumulant generating function.

We have previously outlined the universality of the CLT with respect to the underlying distributions of the involved random variables. This is generally not the case for LDPs, which we can already see exemplified in the theorem of Cramér for the empirical average: as the rate function is given by the Legendre-Fenchel transform of the cumulant generating function, it is hence dependent of the underlying distribution. We thus get an impression of the higher sensitivity of large deviation results towards the distributions of the involved random variables.

Since we also apply the definition of LDPs to random measures, let us briefly address this setting as well. For a Polish space $\mathbb{X}$ we denote by $\mathcal{M}_1(\mathbb{X})$ the space of probability measures on $\mathbb{X}$ endowed with the topology of weak convergence and recall that $\mathcal{M}_1(\mathbb{X})$ is itself again Polish (see [68, Theorem 4.2, Lemma 4.3, Lemma 4.5]), e.g., when equipped with the Lévy-Prokhorov metric (see (4.4)). Then a sequence of random measures on $\mathbb{X}$ is just a sequence of random variables on $\mathcal{M}_1(\mathbb{X})$ and the definition of an LDP can be applied as previously.

### 2.3.2 Large deviations toolbox

In this section we present some results from large deviations theory which we will frequently use to prove the results within this thesis. We start off by formulating a multi-dimensional version of the theorem of Cramér (cf. [29, Theorem 2.2.30]) with the relaxed condition on the exponential moments established in [29, Corollary 6.1.6].

**Proposition 2.3.3** *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random vectors in $\mathbb{R}^n$ with cumulant generating function $\Lambda_X$. If the origin is an interior point of* $\mathrm{Dom}(\Lambda_X)$*, then the sequence of empirical averages $\left(\frac{1}{n}S_n\right)_{n\in\mathbb{N}}$, $S_n := \sum_{i=1}^n X_i$, satisfies an LDP with speed $n$ and good rate function $\mathcal{I}(\cdot) = \Lambda_X^*(\cdot)$.*

Note that in the above proposition the sequence index and the dimension of the ambient space coincide, which generally does not need to be the case for the multi-dimensional Theorem of Cramér. But since we consider concentration phenomena for convex bodies as their dimension tends to infinity, it will be the case in our setting.

The upcoming proposition concerns the large deviation behaviour of two sequences of random variables with the same speed in a product space. For the result and its proof see [4, Proposition 2.4, Appendix A].

**Proposition 2.3.4** *Let $\mathbb{X}, \mathbb{Y}$ be Polish spaces. Let $(X_n)_{n\in\mathbb{N}}$, $(Y_n)_{n\in\mathbb{N}}$ be sequences of random variables in $\mathbb{X}$ and $\mathbb{Y}$, respectively. Assume that $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ are independent. Further assume that both $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ satisfy LDPs with the same speed $s(n)$ and respective good rate functions $\mathcal{I}_X : \mathbb{X} \to [0, \infty]$ and $\mathcal{I}_Y : \mathbb{Y} \to [0, \infty]$. Consider the sequence of random variables $(Z_n)_{n\in\mathbb{N}}$ on $\mathbb{X} \times \mathbb{Y}$ with $Z_n = (X_n, Y_n)$. Then $(Z_n)_{n\in\mathbb{N}}$ satisfies an LDP with speed $s(n)$ and good rate function $\mathcal{I}_Z$ with $\mathcal{I}_Z(z) = \mathcal{I}_X(x) + \mathcal{I}_Y(y)$ for all $z = (x, y) \in \mathbb{X} \times \mathbb{Y}$.*

The next result is the so-called contraction principle and it gives a way to transport an LDP from one sequence of random variables to another by virtue of a continuous map. The result and its proof can be found in [29, Theorem 4.2.1].

**Proposition 2.3.5** *Let $\mathbb{X}, \mathbb{Y}$ be Polish spaces and $f : \mathbb{X} \to \mathbb{Y}$ be a continuous function. Also let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables in $\mathbb{X}$ that satisfies an LDP with speed $s(n)$ and good rate function $\mathcal{I}_X$. Then the sequence of random variables $(Y_n)_{n\in\mathbb{N}} := (f(X_n))_{n\in\mathbb{N}}$ satisfies an LDP with speed $s(n)$ and good rate function $\mathcal{I}_Y(y) = \inf\{\mathcal{I}_X(x) \,|\, x \in \mathbb{X}, f(x) = y\}$.*

**Remark 2.3.6** In some of the upcoming LDP results we want to use the contraction principle in the following situation. Let $\mathbb{X}, \mathbb{Y}$ be Polish spaces, $f : \mathbb{X} \to \mathbb{Y}$ a continuous map, and $(\mu_n)_{n\in\mathbb{N}}$ a sequence of random measures on $\mathbb{X}$. Let $(\mu_n)_{n\in\mathbb{N}}$ satisfy an LDP on $\mathcal{M}_1(\mathbb{X})$ with speed $s : \mathbb{N} \to (0, \infty)$ and rate function $\mathcal{I}_\mu : \mathcal{M}_1(\mathbb{X}) \to [0, \infty]$. We then consider the sequence of random measures $(\nu_n)_{n\in\mathbb{N}}$ on $\mathbb{Y}$ with $\nu_n = \mu_n \circ f^{-1}$ and want to use the contraction principle to infer an LDP for $(\nu_n)_{n\in\mathbb{N}}$. In this case the function that is actually "transporting" the LDP is not $f : \mathbb{X} \to \mathbb{Y}$ itself, but $F : \mathcal{M}_1(\mathbb{X}) \to \mathcal{M}_1(\mathbb{Y})$ with $F(\mu) = \mu \circ f^{-1}$. So in general the continuity of $F$ has to be given rather than that of $f$. But the latter follows directly from the continuity of $f$ by the definition of weak convergence: as discussed, equipped with, e.g., the Lévy-Prokhorov metric (see (4.4)) $\mathcal{M}_1(\mathbb{X})$ and $\mathcal{M}_1(\mathbb{Y})$ are metric spaces, hence we can show the continuity of $F$ by proving sequential continuity of $F$. This can be done with respect to the weak convergence of measures, since it is equivalent to the Lévy-Prokhorov metric (see [68, Lemma 4.3]). Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}_1(\mathbb{X})$ that converges weakly to a probability measure

$\mu \in \mathcal{M}_1(\mathbb{X})$ as $n \to \infty$. That means that for every bounded continuous function $h : \mathbb{X} \to \mathbb{R}$ it holds that

$$\lim_{n\to\infty} \int h(x)\, \mu_n(\mathrm{d}x) = \int h(x)\, \mu(\mathrm{d}x).$$

To show the continuity of $F$ we need to show that $F(\mu_n)$ converges weakly to $F(\mu)$ as $n \to \infty$. Let $h : \mathbb{Y} \to \mathbb{R}$ be a bounded continuous function. Since $f$ is continuous, we know that $(h \circ f)$ is bounded and continuous. Thus, by the weak convergence of $(\mu_n)_{n\in\mathbb{N}}$ we have that

$$\begin{aligned}
\lim_{n\to\infty} \int h(x)\, F(\mu_n)(\mathrm{d}x) &= \lim_{n\to\infty} \int h(x)\, \left(\mu_n \circ f^{-1}\right)(\mathrm{d}x) \\
&= \lim_{n\to\infty} \int \left(h \circ f\right)(x)\, \mu_n(\mathrm{d}x) \\
&= \int \left(h \circ f\right)(x)\, \mu(\mathrm{d}x) \\
&= \int h(x)\, \left(\mu \circ f^{-1}\right)(\mathrm{d}x) \\
&= \int h(x)\, F(\mu)(\mathrm{d}x).
\end{aligned}$$

Hence, $F(\mu_n)$ converges weakly to $F(\mu)$ as $n \to \infty$ and thereby $F$ is continuous.

So far we have only covered LDP results for sequences of i.i.d. random variables. For sequences of non-identically distributed random variables, that do however exhibit a certain level of distributional convergence, the theorem of Gärtner-Ellis (see, e.g., [29, Theorem 2.3.6]) provides a useful way to gain an LDP.

**Proposition 2.3.7** *Let $(X^{(n)})_{n\in\mathbb{N}}$ be a sequence of random variables with cumulant generating functions $\Lambda_n$ and $k \in [1, \infty)$. We assume that for all $t \in \mathbb{R}$ the limit $\Lambda(t) := \lim_{n\to\infty} \frac{1}{n^k}\Lambda_n(n^k t)$ exists in $[-\infty, +\infty]$ and that the origin is an interior point of the effective domain* $\mathrm{Dom}(\Lambda)$*. We furthermore assume that $\Lambda$ is lower semi-continuous and differentiable on the interior of* $\mathrm{Dom}(\Lambda)$*. Then the sequence $(X^{(n)})_{n\in\mathbb{N}}$ satisfies an LDP with speed $n^k$ and rate function $\Lambda^*$.*

Lastly, we present a result that yields a weak LDP for a sequence of probability measures, if the limits of the logarithmic probabilities from Definition 2.3.2 have the same supremum over the base of the underlying topology (cf. [29, Theorem 4.1.11]).

**Proposition 2.3.8** *Let $\mathbb{X}$ be a Polish space and $(\mathbf{P}_n)_{n\in\mathbb{N}}$ a sequence of probability measures on $\mathbb{X}$. Further, let $\mathcal{A}$ be a base of the topology of $\mathbb{X}$. If for every $x \in \mathbb{X}$ it holds that*

$$\sup_{A\in\mathcal{A}:\, x\in A} \liminf_{n\to\infty} \frac{1}{n} \log \mathbf{P}_n(A) = \sup_{A\in\mathcal{A}:\, x\in A} \limsup_{n\to\infty} \frac{1}{n} \log \mathbf{P}_n(A),$$

*then $(\mathbf{P}_n)_{n\in\mathbb{N}}$ satisfies a weak LDP with speed $n$ and rate function*

$$I(x) := \sup_{A\in\mathcal{A}:\, x\in A} \liminf_{n\to\infty} \frac{1}{n} \log \mathbf{P}_n(A).$$

## 2.4 $\ell_p^n$-balls

In this section we establish the distributions on $\ell_p^n$-balls that are at the core of the results within this thesis. Furthermore, we present some probabilistic representations that are essential to their proofs. Furthermore, we list some classic polar integration tools, which are also frequently of use.

### 2.4.1 Probabilistic representation results on $\ell_p^n$-balls

As mentioned in the introduction, for $p \in (0,\infty]$, $n \in \mathbb{N}$, and $x = (x_1,\dots,x_n) \in \mathbb{R}^n$ we denote by

$$\|x\|_p := \begin{cases} \left(\sum\limits_{i=1}^n |x_i|^p\right)^{1/p} & : p < \infty \\ \max\{|x_1|,\dots,|x_n|\} & : p = \infty \end{cases}$$

the $\ell_p^n$-norm of $x$ (which for $p \in (0,1)$ is only a quasi-norm) and set

$$\mathbb{B}_p^n := \{x \in \mathbb{R}^n : \|x\|_p \le 1\} \qquad \text{and} \qquad \mathbb{S}_p^{n-1} := \{x \in \mathbb{R}^n : \|x\|_p = 1\} \tag{2.11}$$

to be the unit ball and unit sphere with respect to this (quasi-)norm, respectively. We define the uniform distribution on $\mathbb{B}_p^n$ and the cone probability measure on $\mathbb{S}_p^{n-1}$ as

$$\mathbf{U}_{n,p}(\,\cdot\,) := \frac{\mathrm{vol}_n(\,\cdot\,)}{\mathrm{vol}_n(\mathbb{B}_p^n)} \qquad \text{and} \qquad \mathbf{C}_{n,p}(\,\cdot\,) := \frac{\mathrm{vol}_n\left(\{rx : r \in [0,1], x \in \,\cdot\,\}\right)}{\mathrm{vol}_n(\mathbb{B}_p^n)}, \tag{2.12}$$

where the cone probability measure (often just called the cone measure) assigns to a set $A \subset \mathbb{S}_p^{n-1}$ the volume of the cone it encloses with the origin relative to the volume of $\mathbb{B}_p^n$ (see Figure 2.1).

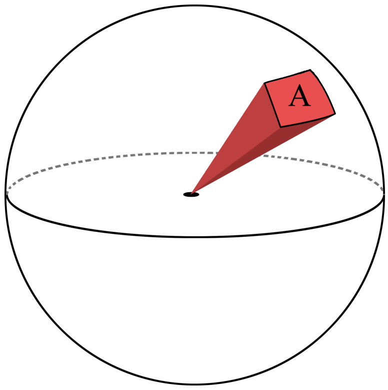


FIGURE 2.1: Cone which a set $A \subset \mathbb{S}_p^{n-1}$ encloses with the origin for $n = 2, p = 2$.

**Remark 2.4.1** On $\mathbb{S}_p^{n-1}$ one can also consider the surface measure $\sigma_p^n$ , which is defined to be the sufficiently normalized $(n-1)$-dimensional Hausdorff measure. However, as discussed in [98, Section 3.1], the cone measure $\mathbf{C}_{n,p}$ is more canonical to consider on $\ell_p^n$-balls, as it is unique in its ability on the $\ell_p^n$-sphere to be probabilistically represented as in Proposition 2.4.2. This is due to the fact that this representation result relies heavily on polar integration arguments, and the cone measure has been shown in [93, Proposition 1] to be the unique measure on $\mathbb{S}_p^{n-1}$ such that the polar integration formula (2.14) holds for any integrable $f : \mathbb{R}^n \to \mathbb{R}$. While $\mathbf{C}_{n,p}$ and $\sigma_p^n$ have been shown in [99] to coincide (exclusively) for $p \in \{1, 2, \infty\}$, their difference in form of their total variation distance

$$d_{TV}\left(\mathbf{C}_{n,p}, \sigma_p^n\right) := \sup\left\{|\mathbf{C}_{n,p}(A) - \sigma_p^n(A)| : A \in \mathcal{B}(\mathbb{S}_p^{n-1})\right\}$$

is bounded and was shown to decrease as $n^{-1/2}$ in [92].

For a random vector $X^{(n)} \in \mathbb{R}^n$, which is uniformly distributed on the (sufficiently scaled) standard Euclidean sphere, i.e., $n^{1/2}\,\mathbb{S}_2^{n-1} \subset \mathbb{R}^n$, and $k \in \mathbb{N}$, any $k$ coordinates $X_{i_1}^{(n)}, \ldots, X_{i_k}^{(n)}$ of $X^{(n)}$ seen as a random vector in $\mathbb{R}^k$ converge weakly in distribution to a Gaussian random vector. In short, for fixed $k \in \mathbb{N}$, the $k$-dimensional marginals of the uniform distribution on the high-dimensional sphere are approximately Gaussian. This result is the well-known Poincaré-Maxwell-Borel Lemma, which was shown around the start of the 20th century (see [31]). In 1991 this was generalized by Mogulski [91] and Rachev and Rüschendorf [99], who showed that for random vectors distributed according to the cone measure on the (sufficiently scaled) $\ell_p^n$-sphere, $p \in [1, \infty)$, the $k$-marginals converge to $p$-generalized Gaussian random vectors. This was incorporated by Rachev and Rüschendorf [99] and Schechtman and Zinn [106] into the following probabilistic representation for random vectors with distributions $\mathbf{C}_{n,p}$ and $\mathbf{U}_{n,p}$.

**Proposition 2.4.2** *Let* $p \in [1,\infty)$, $Y = (Y_1,\dots,Y_n)$ *be a random vector in* $\mathbb{R}^n$ *with* $Y_i \sim \mathbf{N}_p$ *i.i.d., and* $U$ *be an independent random variable uniformly distributed on* $[0,1]$. *Then*

*i) the random vector* $\frac{Y}{\|Y\|_p}$ *has distribution* $\mathbf{C}_{n,p}$ *and is independent of* $\|Y\|_p$,

*ii) the random vector* $U^{1/n}\frac{Y}{\|Y\|_p}$ *has distribution* $\mathbf{U}_{n,p}$.

For $p \in (0,\infty]$ this was generalized by Barthe, Guédon, Mendelson and Naor [13], who gave a probabilistic representation for a class of mixtures of $\mathbf{C}_{n,p}$ and $\mathbf{U}_{n,p}$. For a Borel probability measure $\mathbf{W}$ on $[0,\infty)$ they defined the class of distributions

$$\mathbf{P}_{n,p,\mathbf{W}} := \mathbf{W}(\{0\})\mathbf{C}_{n,p} + \Psi\mathbf{U}_{n,p} \tag{2.13}$$

on $\mathbb{B}_p^n$, where $\Psi(x) = \psi(\|x\|_p)$, $x \in \mathbb{B}_p^n$, is a $p$-radial density given by

$$\psi(s) = \frac{1}{p^{n/p}\,\Gamma\left(\frac{n}{p}+1\right)}\frac{1}{(1-s^p)^{\frac{n}{p}+1}}\left[\int\limits_{(0,\infty)} w^{\frac{n}{p}}\, e^{-\frac{1}{p}\left(\frac{s^p}{1-s^p}\right)w}\,\mathbf{W}(\mathrm{d}w)\right], \qquad 0 \le s \le 1,$$

and provided a convenient representation of $\mathbf{P}_{n,p,\mathbf{W}}$ via a random vector of $p$-generalized Gaussians (see Proposition 2.4.4).

The choice of $\mathbf{W}$ determines how exactly the cone measure and the uniform distribution get mixed. Heuristically, one can think of $\mathbf{W}$ as indicating how probability mass is distributed $p$-radially within $\mathbb{B}_p^n$ (see Figure 2.2).

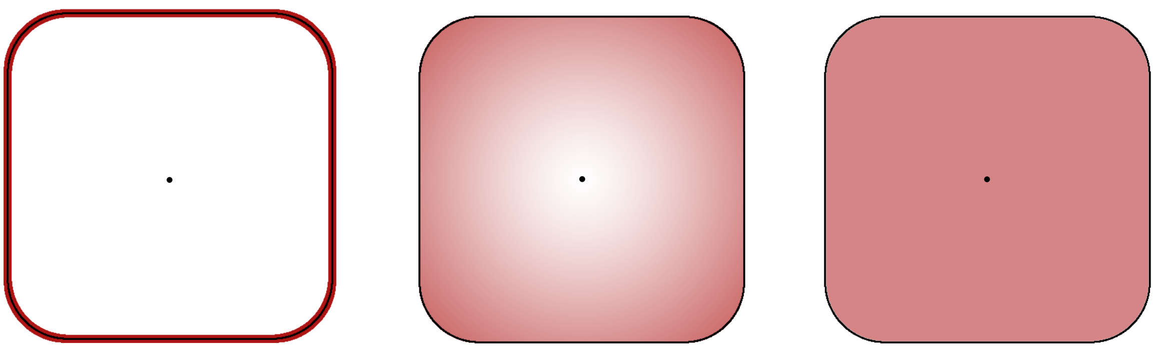

FIGURE 2.2: Distributions $\mathbf{P}_{n,p,\mathbf{W}}$ of probability mass (red) within $\mathbb{B}_p^n$ for $n=2$, $p>2$ and different choices of $\mathbf{W}$ for $\mathbf{W}=\delta_0$, $\mathbf{W}\in\mathcal{M}_1([0,\infty))$ and $\mathbf{W}=\mathbf{E}(1)$ (from left to right).

**Remark 2.4.3** It was shown in [13] that choosing $\mathbf{W} = \delta_0$ to be the Dirac measure at 0 yields that $\mathbf{P}_{n,p,\mathbf{W}} = \mathbf{C}_{n,p}$, and for $\mathbf{W} = \mathbf{E}(1)$, we have that $\mathbf{P}_{n,p,\mathbf{W}} = \mathbf{U}_{n,p}$. For $m \in \mathbb{N}$ choosing $\mathbf{W} = \mathbf{G}\big(\frac{m}{p}, \frac{1}{p}\big)$, i.e., a gamma distribution with shape $\frac{m}{p}$ and rate $\frac{1}{p}$, it can be shown that $\mathbf{P}_{n,p,\mathbf{W}}$ then corresponds to the projection of $\mathbf{C}_{n+m,p}$ onto its first $n$ coordinates. An analogue correspondence is given for $\mathbf{W} = \mathbf{G}\big(1 + \frac{m}{p}, \frac{1}{p}\big)$ and the projection of $\mathbf{U}_{n+m,p}$ onto its first $n$ coordinates (see [13]).

Thus, the motivation behind considering this class of distributions is twofold. First, as outlined in Remark 2.4.3, they encompass many relevant distributions on $\mathbb{B}_p^n$. The second reason we consider $\mathbf{P}_{n,p,\mathbf{W}}$ specifically is the following probabilistic representation result via $p$-generalized Gaussians shown for it in [13, Theorem 3].

**Proposition 2.4.4** *Let $n \in \mathbb{N}$ and $p \in (0, \infty)$. Let $\mathbf{W}$ be a Borel probability measure on $[0, \infty)$ and $W$ be a random variable with $W \sim \mathbf{W}$. Further, let $X_1, \dots, X_n$ be i.i.d. random variables with $X_i \sim \mathbf{N}_p$, which are independent of $W$. Then the random vector*

$$\frac{X}{(\|X\|_p^p + W)^{1/p}}$$

*has distribution $\mathbf{P}_{n,p,\mathbf{W}}$ on $\mathbb{B}_p^n$ as in* (2.13).

Both the class of probability measures $\mathbf{P}_{n,p,\mathbf{W}}$ and their above representation result can be formulated just as well for sequences $(\mathbf{W}_n)_{n\in\mathbb{N}}$ of Borel probability measures $\mathbf{W}_n$ on $[0, \infty)$ instead of a single fixed distribution $\mathbf{W}$. In such cases, we write

$$\mathbf{P}_{n,p,\mathbf{W}_n} := \mathbf{W}_n\left(\{0\}\right)\mathbf{C}_{n,p} + \Psi_n \mathbf{U}_{n,p},$$

with $\Psi_n$ defined as previously for $\mathbf{W}_n$.

As mentioned in Remark 2.2.3, different research papers use different versions of $p$-generalized Gaussian distributions $\mathbf{N}_p$ and $\tilde{\mathbf{N}}_p$ for their probabilistic representations. Hence, for the sake of completeness, we shall include a version of Proposition 2.4.4 using $\tilde{\mathbf{N}}_p$ instead of $\mathbf{N}_p$, as it will be used within Chapter 3. However, for notational brevity we will keep the same naming conventions as previously for $\mathbf{P}_{n,p,\mathbf{W}}$ etc., since the differences are comparably small and the areas of applications are clearly separated, with Chapter 3 using $\tilde{\mathbf{N}}_p$ and Chapter 4 and Chapter 5 using $\mathbf{N}_p$.

**Proposition 2.4.5** *Let $n \in \mathbb{N}$ and $p \in (0,\infty)$. Let $\mathbf{W}$ be a Borel probability measure on $[0,\infty)$ and $W$ be a random variable with $W \sim \mathbf{W}$. Further, let $X_1, \dots, X_n$ be i.i.d. random variables with $X_i \sim \tilde{\mathbf{N}}_p$, which are independent of $W$. Then the random vector*

$$\frac{X}{(\|X\|_p^p + W)^{1/p}}$$

*has distribution*

$$\mathbf{P}_{n,p,\mathbf{W}} := \mathbf{W}(\{0\})\,\mathbf{C}_{n,p} + \Psi \mathbf{U}_{n,p}$$

*on $\mathbb{B}_p^n$, where $\Psi(x) = \psi(\|x\|_p)$, $x \in \mathbb{B}_p^n$, is a $p$-radial density with*

$$\psi(s) = \frac{1}{\Gamma\left(\frac{n}{p}+1\right)} \frac{1}{(1-s^p)^{\frac{n}{p}+1}} \left[ \int\limits_{(0,\infty)} w^{\frac{n}{p}}\, e^{-\frac{s^p}{1-s^p}w}\, \mathbf{W}(\mathrm{d}w) \right], \qquad 0 \le s \le 1.$$

### 2.4.2 Polar integration

At the very start of this thesis in (1.1) we discussed the Minkowski functional of a symmetric convex body $K \subset \mathbb{R}^n$, set to be

$$\|x\|_K := \inf\{r \in [0,\infty) : x \in rK\}, \qquad x \in \mathbb{R}^n,$$

which defines a norm $\|\cdot\|_K$. Generally, we say a function $f : \mathbb{R}^n \to \mathbb{R}$ is $K$-radially symmetric - or just $K$-radial, for short - if $f(x)$ is only dependent on $\|x\|_K$. If a probability measure's distribution function is $K$-radial, we call it a $K$-radial distribution. For the special case $K = \mathbb{B}_p^n$ we speak of $p$-radial distributions.

Since distributions given by radially symmetric densities (such as $\mathbf{P}_{n,p,\mathbf{W}_n}$) play a central role in this thesis, we need a tool to work with them efficiently. This tool is provided by the polar integration formula. Let $K \subset \mathbb{R}^n$, $n \in \mathbb{N}$, be a set that is star-shaped with respect to the origin and has finite non-zero volume. We define the uniform distribution on $K$ and the cone probability measure on the boundary $\partial K$ as

$$\mathbf{U}_K(\,\cdot\,) := \frac{\mathrm{vol}_n(\,\cdot\,)}{\mathrm{vol}_n(K)} \qquad \text{and} \qquad \mathbf{C}_K(\,\cdot\,) := \frac{\mathrm{vol}_n(\{rx : r \in [0,1], x \in \,\cdot\,\})}{\mathrm{vol}_n(K)},$$

respectively. We can now formulate the general version of the polar integration formula.

**Lemma 2.4.6** *For any set $K \subset \mathbb{R}^n$, $n \in \mathbb{N}$, that is star-shaped with respect to the origin, contains the origin in its interior, and has finite non-zero volume, and for any measurable function $h : \mathbb{R}^n \to \mathbb{R}$ it holds that*

$$\int_{\mathbb{R}^n} h(x)\,\mathrm{d}x = n\,\mathrm{vol}_n(K)\int_0^\infty r^{n-1}\int_{\partial K} h(ry)\,\mathbf{C}_K(\mathrm{d}y)\,\mathrm{d}r.$$

The proof of this is the same as that of Proposition 3.3 in [98], which deals with the case where $K$ is a symmetric convex body (see also [93, Proposition 1]). Note that setting $K = \mathbb{B}_p^n$ yields $\mathbf{U}_{\mathbb{B}_p^n} = \mathbf{U}_{n,p}$ and $\mathbf{C}_{\mathbb{B}_p^n} = \mathbf{C}_{n,p}$ and the polar integration formula for $\ell_p^n$-balls, i.e., for any measurable function $h : \mathbb{R}^n \to \mathbb{R}$ it holds that

$$\int_{\mathbb{R}^n} h(x)\,\mathrm{d}x = n\,\mathrm{vol}_n(\mathbb{B}_p^n)\int_0^\infty r^{n-1}\int_{\mathbb{S}_p^{n-1}} h(ry)\,\mathbf{C}_{n,p}(\mathrm{d}y)\,\mathrm{d}r. \tag{2.14}$$

Polar integration actually plays a key role in the proofs of the probabilistic representation results in the previous section. Since we will need a polar integration formula in the non-negative orthant $\mathbb{R}_+^n$ of $\mathbb{R}^n$ later in Chapter 3, we will present one for that specific case.

**Corollary 2.4.7** *For any set $K \subset \mathbb{R}_+^n$, $n \in \mathbb{N}$, that is star-shaped with respect to the origin, contains the origin in its interior with respect to $\mathbb{R}_+^n$, and has finite non-zero volume, and for any measurable function $h : \mathbb{R}_+^n \to \mathbb{R}$ it holds that*

$$\int_{\mathbb{R}_+^n} h(x)\,\mathrm{d}x = n\,\mathrm{vol}_n(K)\int_0^\infty r^{n-1}\int_{\partial K} h(ry)\,\mathbf{C}_K(\mathrm{d}y)\,\mathrm{d}r.$$

The proof of this again follows along the same lines as that of Proposition 3.3 in [98], and hence will be omitted here.

**Remark 2.4.8** Let us briefly address why we are considering $\ell_p^n$-balls specifically instead of other symmetric convex bodies, with the discussion in this remark being largely based on [98, Proposition 3.3]. As pointed out therein, probabilistic representations of Schechtman-Zinn-type as in Proposition 2.4.2 can be found for other symmetric convex bodies $K$ as well (this holds even for merely star-shaped bodies, although for these $\|\cdot\|_K$ does not define a norm as it generally is not absolutely homogeneous). For any such body $K$, a random vector $Z$ with $K$-radial distribution $\psi_Z$, and an independent $U \sim \mathrm{Unif}[0,1]$ it holds that

i) $\frac{Z}{\|Z\|_K}$ has distribution $\mathbf{C}_K$ and is independent of $\|Z\|_K$,

ii) the random vector $U^{1/n}\,\frac{Z}{\|Z\|_K}$ has distribution $\mathbf{U}_K$.

The important difference to $\mathbb{B}_p^n$ is, however, that by the specific shape of the Minkowski functional of $\ell_p^n$-balls we can construct a $p$-radial distribution for $Z$ such that its coordinates are i.i.d. and exhibit a convenient distribution in form of $\mathbf{N}_p$. For $\mathbb{B}_p^n$ the related Minkowski functional has the form

$$\|x\|_{\mathbb{B}_p^n} = \|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p},$$

which is absolutely homogeneous and for which one can construct the $p$-radial distribution

$$\psi_Z(\|x\|_p) = (2\,p^{1/p}\,\Gamma(1+1/p))^{-n} e^{-\frac{1}{p}\|x\|_p^p} = \prod_{i=1}^n f_{\mathbf{N}_p}(x_i),$$

resulting in a probabilistic representation via i.i.d. $p$-generalized Gaussians. To understand why $\ell_p^n$-balls specifically are so accessible for calculations, let us consider a more general case. Assume that for some $K \subset \mathbb{R}^n$ the Minkowski functional has the form

$$\|x\|_K = F\left(\sum_{i=1}^n f_i(x_i)\right), \tag{2.15}$$

where the $f_1, \ldots, f_n$ are so-called Orlicz functions, that is, real-valued, even, and convex functions with $f_i(0) = 0$ and $f_i(x_i) > 0$ for $x_i \neq 0$ for all $i \in \{1, \ldots, n\}$, and $F$ is non-negative and invertible. Then, up to normalizing constants, one can construct a $K$-radial density

$$\psi_Z(\|x\|_K) = e^{-F^{-1}(\|x\|_K)} = e^{-\sum_{i=1}^n f_i(x_i)} = \prod_{i=1}^n e^{-f_i(x_i)}. \tag{2.16}$$

This yields that a random vector $Z$ with $K$-radial density $\psi_Z$ has independent coordinates. Choosing $f_i = f$ for all $i \in \{1, \ldots, n\}$ then yields that the coordinates of $Z$ are identically distributed as well. Using such Orlicz functions $f$ and $f_1, \ldots, f_n$, one can define the objects

$$\mathbb{B}_f^n := \left\{x \in \mathbb{R}^n : \sum_{i=1}^n f(x_i) \leq n\right\} \qquad \text{and} \qquad \mathbb{B}_{f_1,\ldots,f_n}^n := \left\{x \in \mathbb{R}^n : \sum_{i=1}^n f_i(x_i) \leq n\right\},$$

respectively called Orlicz balls and Musielak–Orlicz balls (cf. [75, Section 3.3]). These are generalizations of $\ell_p^n$-balls and have been subject to a lot of promising research (see [3, 14, 35, 56, 60, 75]). However, it is actually the case that $\ell_p^n$-balls are in fact the *only* convex bodies, for which the Minkowski functional has the form as in (2.15).

This is due to the fact that for any Orlicz function except $f(x) = |x|^p$, $x \in \mathbb{R}$, the expression on the right-hand side in (2.15) does not define a norm, as the condition of absolute homogeneity is not satisfied. If we postulate an Orlicz function $f$ to be absolutely homogeneous of some degree $\alpha > 0$, it directly follows that $f(x) = |x|^\alpha f(1)$, yielding the Orlicz function corresponding to $\ell_p^n$-balls up to a constant $f(1)$. Thus, Orlicz balls are an interesting and promising area of research, and they do allow for a *general* Schechtman-Zinn-type probabilistic representation as above, but the corresponding radial density does not factorize, and therefore, one cannot represent a random vector from an Orlicz ball via a random vector with i.i.d. coordinates that have some distribution analogue to $\mathbf{N}_p$.

So, overall, one can see that while the $\ell_p^n$-balls are not alone in the fact that there exist very general Schechtman-Zinn-type representations, they do, however, combine this with the specific convenient structure of their Minkowski functional, which is both absolutely homogeneous and results in the probabilistic representations to be made up of i.i.d. $p$-generalized Gaussians, making those representations specifically useful and thereby $\ell_p^n$-balls very accessible for calculations.

# Chapter 3

# Weighted $p$-radial distributions on Euclidean and matrix $p$-balls

This chapter extends the probabilistic representation result of [13] in Proposition 2.4.5 from the previous class of $p$-radial distributions $\mathbf{P}_{n,p,\mathbf{W}}$ by an additional homogeneous weight function $f$, denoted by $\mathbf{P}_{n,p,\mathbf{W},f}$. These distributions are based on weighted versions of the cone measure and the uniform distribution, $\mathbf{C}_{n,p,f}$ and $\mathbf{U}_{n,p,f}$, for which we will first derive Schechtman-Zinn-type representations as in Proposition 2.4.2.

We then turn from Euclidean space $\mathbb{R}^n$ to the spaces of self-adjoint and non-self-adjoint $(n \times n)$ matrices and consider analogues of the $\ell_p^n$-balls therein, which we will call *matrix $p$-balls*. We make use of the previously derived representation result for weighted $p$-radial distributions on Euclidean $p$-balls to derive the eigenvalue distribution and singular value distribution of random matrices with (unweighted) $p$-radial distribution on those matrix $p$-balls.

As an application, we show large deviation principles both for the empirical measure of random vectors with weighted $p$-radial distribution in Euclidean $p$-balls and for the empirical spectral measure of random matrices with $p$-radial distributions in matrix $p$-balls. For both cases, we chose $\mathbf{P}_{n,p,\mathbf{W}}$ and its matrix-analogue for a $\mathbf{W}_n$ that varies in $n \in \mathbb{N}$, for which different limiting distributions of $\mathbf{W}_n$ encapsulate several concrete distributions on Euclidean and matrix $p$-balls.

# 3.1 Preliminaries

Since the representation results for weighted $p$-radial distributions are often tailored to their further use for matrix $p$-balls, we begin by establishing matrix $p$-balls and the distributions on them which are of interest. Further, we will need quite a few Laplace integration results to show the aforementioned large deviation principles, therefore we will also present them here.

## 3.1.1 Matrix $p$-balls

Let $\mathbb{F}_\beta$ be the real numbers (if $\beta = 1$), the complex numbers (if $\beta = 2$) or the Hamiltonian quaternions (if $\beta = 4$). For $n \in \mathbb{N}$ and $\beta \in \{1, 2, 4\}$ we let $\mathscr{M}_n(\mathbb{F}_\beta)$ be the space of $(n \times n)$ matrices with entries from $\mathbb{F}_\beta$. For a matrix $A \in \mathscr{M}_n(\mathbb{F}_\beta)$ let $A^*$ be the adjoint of $A$. It is well known that, together with the scalar product $\langle A, B\rangle_{\mathscr{M}} := \operatorname{Re}\operatorname{Tr}(AB^*)$, $\mathscr{M}_n(\mathbb{F}_\beta)$ becomes a Euclidean vector space. By $\operatorname{vol}_{\beta,n}(\,\cdot\,)$ we denote the volume on $\mathscr{M}_n(\mathbb{F}_\beta)$ corresponding to this scalar product. We can now introduce the self-adjoint matrix space $\mathscr{H}_n(\mathbb{F}_\beta) := \{A \in \mathscr{M}_n(\mathbb{F}_\beta) : A = A^*\}$. For each $A \in \mathscr{H}_n(\mathbb{F}_\beta)$ we denote by $\lambda_1(A) \leq \ldots \leq \lambda_n(A)$ the (real) eigenvalues of $A$ (see [6, Appendix E] for a formal definition in the case $\beta = 4$) and define $\lambda(A) := (\lambda_1(A), \ldots, \lambda_n(A)) \in \mathbb{R}^n$. For $0 < p \leq \infty$ the self-adjoint matrix $p$-ball in $\mathscr{H}_n(\mathbb{F}_\beta)$ is defined as

$$\mathbb{B}^{n,\mathscr{H}}_{p,\beta} := \{A \in \mathscr{H}_n(\mathbb{F}_\beta) : \|\lambda(A)\|_p \leq 1\},$$

where we interpret the condition as $\max\{|\lambda_1(A)|, \ldots, |\lambda_n(A)|\} \leq 1$ if $p = \infty$. Similarly, let

$$\mathbb{S}^{n-1,\mathscr{H}}_{p,\beta} := \{A \in \mathscr{H}_n(\mathbb{F}_\beta) : \|\lambda(A)\|_p = 1\}$$

be the self-adjoint matrix $p$-sphere. The uniform distribution on $\mathbb{B}^{n,\mathscr{H}}_{p,\beta}$ and the cone probability measure on $\mathbb{S}^{n-1,\mathscr{H}}_{p,\beta}$ are denoted by $\mathbf{U}^{\mathscr{H}}_{n,p,\beta}$ and $\mathbf{C}^{\mathscr{H}}_{n,p,\beta}$, respectively.

In the self-adjoint case, one can identify the matrix $p$-balls by virtue of the eigenvalues. We now consider the non-self-adjoint case, where this will be done via the singular values. For $A \in \mathscr{M}_n(\mathbb{F}_\beta)$, $n \in \mathbb{N}$, we denote by $s_1(A) \leq \ldots \leq s_n(A)$ the singular values of $A$, that is, $s_1(A), \ldots, s_n(A)$ are the non-negative eigenvalues of $\sqrt{AA^*}$ (for the cases $\beta \in \{1, 2\}$, and if $\beta = 4$ we refer to [6, Corollary E.13] for a formal definition) and define $s(A) := (s_1(A), \ldots, s_n(A)) \in \mathbb{R}^n_+$. Additionally, set $s^2(A) := (s_1^2(A), \ldots, s_n^2(A)) \in \mathbb{R}^n_+$ to be the vector of squared ordered singular values. We do so, as the coordinates of $s^2(A)$ are the eigenvalues of $AA^*$ and can hence be treated in a fashion analogue to the vector of eigenvalues without needing to account for the root-operation. For $0 < p \leq \infty$

the non-self-adjoint matrix $p$-ball is defined as

$$\mathbb{B}_{p,\beta}^{n,\mathscr{M}} := \{A \in \mathscr{M}_n(\mathbb{F}_\beta) : \|s(A)\|_p \leq 1\},$$

once again replacing the condition by $\max\{|s_1(A)|, \ldots, |s_n(A)|\} \leq 1$ if $p = \infty$. We also denote by

$$\mathbb{S}_{p,\beta}^{n-1,\mathscr{M}} := \{A \in \mathscr{M}_n(\mathbb{F}_\beta) : \|s(A)\|_p = 1\}$$

the non-self-adjoint matrix $p$-sphere. The uniform distribution on $\mathbb{B}_{p,\beta}^{n,\mathscr{M}}$ is denoted by $\mathbf{U}_{n,p,\beta}^{\mathscr{M}}$ and we let $\mathbf{C}_{n,p,\beta}^{\mathscr{M}}$ be the cone probability measure on $\mathbb{S}_{p,\beta}^{n-1,\mathscr{M}}$. Since the singular values are non-negative, we define the non-negative parts of the $\ell_p^n$-ball and $\ell_p^n$-sphere as $\mathbb{B}_{p,+}^n := \mathbb{B}_p^n \cap \mathbb{R}_+^n$ and $\mathbb{S}_{p,+}^{n-1} := \mathbb{S}_p^{n-1} \cap \mathbb{R}_+^n$. Accordingly, we define the respective uniform distribution $\mathbf{U}_{n,p,+} := \mathbf{U}_{\mathbb{B}_{p,+}^n}$ and cone probability measure $\mathbf{C}_{n,p,+} := \mathbf{C}_{\mathbb{S}_{p,+}^{n-1}}$.

We thus identify each matrix in real, complex and quaternionic space with the $n$-dimensional vector of its ordered eigen-/singular values and define the matrix $p$-balls as the set of (non-)self-adjoint matrices, whose vector of ordered eigen-/singular values has $\ell_p^n$-norm less or equal to one, i.e., lies in the Euclidean $p$-ball. This identification of matrix $p$-balls via Euclidean $p$-balls will be a running theme of this chapter. These sets of matrices, as explained in the introduction, are the finite-dimensional analogue of the unit balls of the Schatten trace classes of compact linear operators between two Hilbert spaces with singular values forming a sequence in $\ell_p$ and their self-adjoint counterparts.

**Remark 3.1.1**

(i) Note that both $\mathscr{H}_n(\mathbb{F}_\beta)$ and $\mathscr{M}_n(\mathbb{F}_\beta)$ are Euclidean vector spaces of dimensions $\frac{\beta n(n-1)}{2} + \beta n$ and $\beta n^2$, respectively, and $\mathbb{B}_{p,\beta}^{n,\mathscr{H}}$ and $\mathbb{B}_{p,\beta}^{n,\mathscr{M}}$ both contain their respective origin in their interior and are star-shaped with respect to their origins, as $\|\lambda(\kappa A)\|_p = \kappa\|\lambda(A)\|_p \leq 1$ for $\kappa \in [0,1]$ $\big($analogue for $\mathbb{B}_{p,\beta}^{n,\mathscr{M}}\big)$. Finally, the volumes of $\mathbb{B}_{p,\beta}^{n,\mathscr{H}}$ and $\mathbb{B}_{p,\beta}^{n,\mathscr{M}}$ are non-zero and bounded (see, e.g., [63, 64]). Hence, they both satisfy the conditions of the general polar integration formula given in Lemma 2.4.6.

(ii) When referring to $\mathbb{B}_p^n$ as the "Euclidean" $\ell_p^n$-ball the term is supposed to denote the commutative setting of $\mathbb{B}_p^n$ in contrast to matrix $p$-balls $\mathbb{B}_{p,\beta}^{n,\mathscr{H}}$ and $\mathbb{B}_{p,\beta}^{n,\mathscr{M}}$ in the non-commutative setting of matrix space, despite the matrix spaces themselves being Euclidean vector spaces as well.

For a Borel probability measure $\mathbf{W}$ on $[0,\infty)$ we can now construct the analogues of the measure $\mathbf{P}_{n,p,\mathbf{W}}$ on the matrix $p$-balls $\mathbb{B}_{p,\beta}^{n,\mathscr{H}}$ and $\mathbb{B}_{p,\beta}^{n,\mathscr{M}}$ as

$$\mathbf{P}_{n,p,\mathbf{W},\beta}^{\mathscr{H}} := \mathbf{W}(\{0\})\mathbf{C}_{n,p,\beta}^{\mathscr{H}} + \Psi^{\mathscr{H}}\mathbf{U}_{n,p,\beta}^{\mathscr{H}} \text{ on } \mathbb{B}_{p,\beta}^{n,\mathscr{H}}, \tag{3.1}$$

with $\Psi^{\mathscr{H}}(A) := \psi^{\mathscr{H}}(\|\lambda(A)\|_p)$ for $A \in \mathbb{B}_{p,\beta}^{n,\mathscr{H}}$, and

$$\mathbf{P}_{n,p,\mathbf{W},\beta}^{\mathscr{M}} := \mathbf{W}(\{0\})\mathbf{C}_{n,p,\beta}^{\mathscr{M}} + \Psi^{\mathscr{M}}\mathbf{U}_{n,p,\beta}^{\mathscr{M}} \text{ on } \mathbb{B}_{p,\beta}^{n,\mathscr{M}}, \tag{3.2}$$

with $\Psi^{\mathscr{M}}(A) := \psi^{\mathscr{M}}(\|s^2(A)\|_p)$ for $A \in \mathbb{B}_{p,\beta}^{n,\mathscr{M}}$, where $\psi^{\mathscr{H}}(s)$ and $\psi^{\mathscr{M}}(s)$ are $p$-radial densities given by

$$\frac{1}{\Gamma\left(1+\frac{n+m}{p}\right)}\frac{1}{(1-s^p)^{\frac{n+m}{p}+1}}\left[\int_{(0,\infty)} w^{\frac{n+m}{p}}\, e^{-\frac{s^p}{1-s^p}w}\,\mathbf{W}(\mathrm{d}w)\right], \qquad 0 \le s \le 1,$$

with $m = \frac{\beta n(n-1)}{2}$ for $\psi^{\mathscr{H}}(s)$, and $m = \frac{\beta}{2}n^2 - n$ for $\psi^{\mathscr{M}}(s)$.

We define our distribution classes on matrix $p$-balls similarly to those on Euclidean $p$-balls via a $p$-radial distribution. Although we do not yet have a probabilistic representation for $\mathbf{P}_{n,p,\mathbf{W},\beta}^{\mathscr{H}}$ and $\mathbf{P}_{n,p,\mathbf{W},\beta}^{\mathscr{M}}$ as in Proposition 2.4.5, we still want to analyze the eigenvalue and singular value distribution of random matrices selected on $\mathbb{B}_{p,\beta}^{n,\mathscr{H}}$ and $\mathbb{B}_{p,\beta}^{n,\mathscr{M}}$ according to these distributions. We will be able to achieve this by establishing a new connection between these distributions on matrix $p$-balls and suitably weighted distributions on Euclidean $p$-balls. In contrast to the results of Proposition 2.4.5, however, we need to account for the repulsion between the eigenvalues and singular values, hence the $p$-radial densities $\psi^{\mathscr{H}}(s)$, $\psi^{\mathscr{M}}(s)$ look different than the $\psi$ in Proposition 2.4.5, insofar as the $n$ in $\psi$ is replaced by $n+m$, with $m$ being the degree of homogeneity $m$ of these repulsion factors. We will denote these repulsion factors of the eigen- and singular values by $\Delta_\beta^c$ and $\nabla_\beta^c$ (formal definitions will follow in Section 3.3), and as we will see, the two values for $m$ in the definitions (3.1) and (3.2) are their respective degrees of homogeneity. We will explain this in further detail in the following sections. Also, the fact that $\mathbf{P}_{n,p,\mathbf{W},\beta}^{\mathscr{H}}$ and $\mathbf{P}_{n,p,\mathbf{W},\beta}^{\mathscr{M}}$ are in fact probability measures will follow directly from their probabilistic representations in Theorem 3.3.1 and Theorem 3.3.5, respectively.

### 3.1.2 Asymptotic approximations for Laplace-type integrals

We will need some tools to analyze asymptotic behaviour of Laplace-type integrals to prove our large deviation results. One of them will be provided by the Laplace principle, as presented in [4, Proposition 2.10], and several useful adaptations fitting for our purposes. We begin with the former.

**Proposition 3.1.2** *Let $-\infty < a < b < +\infty$ and $f : [a,b] \to \mathbb{R}$ be a twice continuously differentiable function with a unique point $x_0 \in (a,b)$ such that $f(x_0) = \max_{x\in[a,b]} f(x)$ and $f''(x_0) < 0$. Further, let $h : [a,b] \to \mathbb{R}$ be a positive measurable function. Then*

$$\lim_{n\to\infty} \left( \int_a^b h(x)\, e^{nf(x)}\, \mathrm{d}x \right) \left( \sqrt{\frac{2\pi}{n|f''(x_0)|}}\, h(x_0)\, e^{nf(x_0)} \right)^{-1} = 1.$$

**Remark 3.1.3** Proposition 3.1.2 effectively means that

$$\lim_{n\to\infty} \frac{1}{n} \log \int_a^b h(x)\, e^{nf(x)}\, \mathrm{d}x = f(x_0).$$

However, we want to fit this result somewhat further to our needs. Assuming the set-up of Proposition 3.1.2, let $s^{(1)} := \big(s_n^{(1)}\big)_{n\in\mathbb{N}}$ and $s^{(2)} := \big(s_n^{(2)}\big)_{n\in\mathbb{N}}$ be sequences, where $s^{(1)}$ is non-negative and bounded, and $s^{(2)}$ is positive (or at least positive almost everywhere), such that

$$\lim_{n\to\infty} \frac{1}{n} \big|\log s_n^{(2)}\big| < +\infty. \tag{3.3}$$

Expanding the fraction in the Laplace principle in Proposition 3.1.2 by $s_n^{(2)}$ and adding

$$\lim_{n\to\infty} \frac{s_n^{(1)}}{s_n^{(2)} \sqrt{\frac{2\pi}{n|f''(x_0)|}}\, h(x_0)\, e^{nf(x_0)}},$$

which is zero, since $s_n^{(1)}$ is bounded, yields that

$$\lim_{n\to\infty} \left( s_n^{(1)} + s_n^{(2)} \int_a^b h(x)\, e^{nf(x)}\, \mathrm{d}x \right) \left( s_n^{(2)} \sqrt{\frac{2\pi}{n|f''(x_0)|}}\, h(x_0)\, e^{nf(x_0)} \right)^{-1} = 1.$$

Thus, we have that

$$\lim_{n\to\infty} \frac{1}{n} \log \left[ s_n^{(1)} + s_n^{(2)} \int_a^b h(x)\, e^{nf(x)}\, \mathrm{d}x \right] = \lim_{n\to\infty} \frac{1}{n} \log s_n^{(2)} + f(x_0). \tag{3.4}$$

The last tool for analyzing asymptotic integral behaviour will be the following result by Breitung and Hohenbichler [19], that provides us with asymptotic approximations of Laplace-type integrals even if the involved functions maximize on the boundary of the integration domain, specifically at the origin. Concretely, this is the result given in [19, Lemma 4] for $n = 1, k = 1$, applied to functions $h$ and $f$. The parameter $\lambda$ from [19] in our setting is replaced by the integer $n \in \mathbb{N}$. Since $n = k = 1$, the last condition in [19, Lemma 4] regarding the Hessian of $f$ at 0, that is, $f''(0)$, falls away.

**Proposition 3.1.4** *Let $F \subset \mathbb{R}$ be a compact set with $0 \in F^\circ$. If*

*(a) $f : F \to \mathbb{R}$ and $h : F \to \mathbb{R}$ are continuous functions with $h(0) \neq 0$,*

*(b) $f(x) < f(0)$ for all $x \in F \cap \mathbb{R}_+ \setminus \{0\}$,*

*(c) there exists a neighbourhood $V \subset F$ of $0$ in which $f$ is twice continuously differentiable,*

*(d) $f'(0) < 0$,*

*then it holds that*

$$\lim_{n\to\infty} \left( \int\limits_{F\cap\mathbb{R}_+} h(x)\, e^{nf(x)}\, \mathrm{d}x \right) \Big( n^{-1}\, |f'(0)|^{-1} h(0)\, e^{nf(0)} \Big) = 1.$$

We will only need the results from Proposition 3.1.4 to handle the asymptotics of one specific Laplace-type integral over the set $[0, 1]$, where the function in the exponent maximizes on the boundary at 0. Hence we will derive another asymptotic integral expansion result tailored specifically to our purposes.

**Remark 3.1.5** For functions $h$ and $f$ as described in Proposition 3.1.4 and the set $F = [-1, 1]$ it holds that

$$\lim_{n\to\infty} \frac{1}{n} \log \int\limits_0^1 h(x)\, e^{nf(x)}\, \mathrm{d}x = \lim_{n\to\infty} \frac{1}{n} \log \int\limits_{[-1,1]\cap\mathbb{R}_+} h(x)\, e^{nf(x)}\, \mathrm{d}x = f(0).$$

By the same arguments as in Remark 3.1.3 it also holds that

$$\lim_{n\to\infty} \frac{1}{n} \log \left[ s_n^{(1)} + s_n^{(2)} \int\limits_0^1 h(x)\, e^{nf(x)}\, \mathrm{d}x \right] = \lim_{n\to\infty} \frac{1}{n} \log s_n^{(2)} + f(0) \tag{3.5}$$

for sequences $s^{(1)} := \big(s_n^{(1)}\big)_{n\in\mathbb{N}}$ and $s^{(2)} := \big(s_n^{(2)}\big)_{n\in\mathbb{N}}$ as described there.

## 3.2 Weighted $p$-radial distributions on Euclidean $p$-balls

In this section, we describe a class of probability distributions on the $\ell_p^n$-ball $\mathbb{B}_p^n$ in $\mathbb{R}^n$ and its non-negative counterpart $\mathbb{B}_{p,+}^n$ in $\mathbb{R}_+^n$, generalizing the approach in [13], by allowing for an additional homogeneous weight function. To introduce our framework, we let $f : \mathbb{R}^n \to [0, \infty)$ be a measurable function, which we assume to be (positively) homogeneous of degree $m$ for some $m \geq 0$. By this we mean that $f(tx) = t^m f(x)$ for all $t \geq 0$. We also assume that $f$ is integrable with respect to the cone probability measure $\mathbf{C}_{n,p}$ on the $\ell_p^n$-sphere $\mathbb{S}_p^{n-1}$. In this chapter, we write $\mathscr{F}_m^+(\mathbb{R}^n)$ for the class of such functions (omitting its dependence on $p$ in our notation). For $p \in (0, \infty)$ and $f \in \mathscr{F}_m^+(\mathbb{R}^n)$ we let $C_{n,p,f} \in (0, \infty)$ be the normalization constant such that

$$C_{n,p,f} \int\limits_{\mathbb{R}^n} f(x)\, e^{-\|x\|_p^p}\, \mathrm{d}x = 1, \tag{3.6}$$

and denote by $\mathbf{U}_{n,p,f}$ the probability measure on $\mathbb{B}_p^n$ with density

$$x \mapsto C_{n,p,f} \operatorname{vol}_n(\mathbb{B}_p^n)\, \Gamma\left(\frac{n+m}{p} + 1\right) f(x), \qquad x \in \mathbb{B}_p^n,$$

with respect to $\mathbf{U}_{n,p}$. Similarly, let $\mathbf{C}_{n,p,f}$ be the probability measure on $\mathbb{S}_p^{n-1}$ with density

$$y \mapsto C_{n,p,f}\, n \operatorname{vol}_n(\mathbb{B}_p^n)\, p^{-1}\, \Gamma\left(\frac{n+m}{p}\right) f(y), \qquad y \in \mathbb{S}_p^{n-1},$$

with respect to $\mathbf{C}_{n,p}$. $\mathbf{U}_{n,p,f}$ and $\mathbf{C}_{n,p,f}$ are the aforementioned weighted versions of uniform distribution and cone measure, repectively.

Let us briefly show that the functions above defining $\mathbf{C}_{n,p,f}$ and $\mathbf{U}_{n,p,f}$ are in fact densities. Since $f$ is chosen to be measurable and non-negative, it only remains to check that they integrate to one over their respective domains. We start off with $\mathbf{C}_{n,p,f}$. It holds that

$$\int\limits_{\mathbb{S}_p^{n-1}} \mathbf{C}_{n,p,f}(\mathrm{d}y) = C_{n,p,f}\, n \operatorname{vol}_n(\mathbb{B}_p^n)\, p^{-1}\, \Gamma\left(\frac{n+m}{p}\right) \int\limits_{\mathbb{S}_p^{n-1}} f(y)\, \mathbf{C}_{n,p}(\mathrm{d}y).$$

Furthermore, it follows via a change of variable that

$$\int\limits_0^\infty r^{n+m-1}\, e^{r^p} \mathrm{d}r = p^{-1} \int\limits_0^\infty (r^p)^{\frac{n+m}{p}-1}\, e^{r^p} \mathrm{d}r^p = p^{-1}\, \Gamma\left(\frac{n+m}{p}\right). \tag{3.7}$$

Using (3.7) and the polar integration formula for $\ell_p^n$-balls (2.14), we get that

$$\begin{aligned}\int\limits_{\mathbb{S}_p^{n-1}} \mathbf{C}_{n,p,f}(\mathrm{d}y) &= C_{n,p,f}\, n \,\mathrm{vol}_n(\mathbb{B}_p^n) \int\limits_0^\infty r^{n+m-1}\, e^{r^p} \mathrm{d}r \int\limits_{\mathbb{S}_p^{n-1}} f(y)\, \mathbf{C}_{n,p}(\mathrm{d}y) \\ &= C_{n,p,f}\, n \,\mathrm{vol}_n(\mathbb{B}_p^n) \int\limits_0^\infty r^{n-1} \int\limits_{\mathbb{S}_p^{n-1}} f(ry)\, e^{r^p}\, \mathbf{C}_{n,p}(\mathrm{d}y)\, \mathrm{d}r \\ &= C_{n,p,f} \int\limits_{\mathbb{R}^n} f(x)\, e^{\|x\|_p^p}\, \mathrm{d}x \\ &= 1.\end{aligned}$$

Hence, the function defining $\mathbf{C}_{n,p,f}$ is in fact a density with respect to $\mathbf{C}_{n,p}$. As an auxiliary formula we get

$$\int\limits_{\mathbb{S}_p^{n-1}} f(y)\, \mathbf{C}_{n,p}(\mathrm{d}y) = \left( C_{n,p,f}\, n \,\mathrm{vol}_n(\mathbb{B}_p^n)\, p^{-1}\, \Gamma\left(\frac{n+m}{p}\right)\right)^{-1}. \tag{3.8}$$

We now proceed to $\mathbf{U}_{n,p,f}$. Again, using the polar integration formula for $\ell_p^n$-balls (2.14) and the above auxiliary result (3.8), we get that

$$\begin{aligned}\int\limits_{\mathbb{B}_p^n} \mathbf{U}_{n,p,f}(\mathrm{d}x) &= C_{n,p,f}\, \mathrm{vol}_n(\mathbb{B}_p^n)\, \Gamma\left(\frac{n+m}{p}+1\right) \int\limits_{\mathbb{B}_p^n} f(x)\, \mathbf{U}_{n,p}(\mathrm{d}x) \\ &= C_{n,p,f}\, \Gamma\left(\frac{n+m}{p}+1\right) \int\limits_{\mathbb{R}^n} f(x)\, \mathbf{1}_{\mathbb{B}_p^n}(x)\, \mathrm{d}x \\ &= C_{n,p,f}\, \Gamma\Big(\frac{n+m}{p}+1\Big)\, n\,\mathrm{vol}_n(\mathbb{B}_p^n) \int\limits_0^\infty r^{n-1} \int\limits_{\mathbb{S}_p^{n-1}} f(ry)\, \mathbf{1}_{\mathbb{B}_p^n}(ry)\, \mathbf{C}_{n,p}(\mathrm{d}y)\, \mathrm{d}r \\ &= C_{n,p,f}\, n\,\mathrm{vol}_n(\mathbb{B}_p^n)\, \Gamma\Big(\frac{n+m}{p}+1\Big) \int\limits_0^1 r^{n+m-1}\, \mathrm{d}r \int\limits_{\mathbb{S}_p^{n-1}} f(y)\, \mathbf{C}_{n,p}(\mathrm{d}y) \\ &= C_{n,p,f}\, n\,\mathrm{vol}_n(\mathbb{B}_p^n)\, p^{-1}\, \Gamma\Big(\frac{n+m}{p}\Big) \int\limits_{\mathbb{S}_p^{n-1}} f(y)\, \mathbf{C}_{n,p}(\mathrm{d}y) \\ &= 1.\end{aligned}$$

As mentioned in Section 3.1.1, the singular values of a matrix are non-negative and therefore, as we will see in Section 3.3.2, the vector of singular values is distributed on $\mathbb{B}_{p,+}^n$ and $\mathbb{S}_{p,+}^{n-1}$. For $p \in (0,\infty)$ and $f \in \mathscr{F}_m^+(\mathbb{R}_+^n)$ we define a constant $C_{n,p,f,+}$

and distributions $\mathbf{U}_{n,p,f,+}$ and $\mathbf{C}_{n,p,f,+}$ analogue to the above with respect to $\mathbb{B}^n_{p,+}$ and $\mathbb{S}^{n-1}_{p,+}$. We want to formulate all results in this section for both the classical $\ell^n_p$-balls and -spheres and their non-negative counterparts. However, as the proofs work in an entirely analogue fashion, for the sake of brevity we will use the index $\boxplus$ with all relevant quantities, indicating that any given result can be formulated with and without a "+" in the index of these quantities, i.e., for both $\mathbb{B}^n_p$, $\mathbb{S}^{n-1}_p$ and $\mathbb{B}^n_{p,+}$, $\mathbb{S}^{n-1}_{p,+}$. The relevant proof will then always be carried out for $\mathbb{B}^n_p$ and $\mathbb{S}^{n-1}_p$, and only the changes necessary in the non-negative case pointed out, if any need to be made.

In the next lemma we derive probabilistic representations of the distributions $\mathbf{U}_{n,p,f,\boxplus}$ and $\mathbf{C}_{n,p,f,\boxplus}$. This was proven in [63, Lemma 4.2] for the classical case in $\mathbb{R}^n$ and the proof here works completely analogously.

**Lemma 3.2.1** *Let $0 < p < \infty$ and $f \in \mathscr{F}^+_m(\mathbb{R}^n_\boxplus)$ for some $m \geq 0$. Further, let $X = (X_1, \ldots, X_n)$ be a random vector with joint density $C_{n,p,f,\boxplus}\, e^{-\|x\|_p^p} f(x)$, $x \in \mathbb{R}^n_\boxplus$.*

*(i) Then the random vector $\frac{X}{\|X\|_p}$ has distribution $\mathbf{C}_{n,p,f,\boxplus}$ and $\frac{X}{\|X\|_p}$ and $\|X\|_p$ are independent.*

*(ii) Independently of $X$, let $U$ be uniformly distributed on $[0,1]$. Then $U^{\frac{1}{n+m}} \frac{X}{\|X\|_p}$ has distribution $\mathbf{U}_{n,p,f,\boxplus}$.*

*Proof.* Consider a non-negative measurable function $h : \mathbb{S}^{n-1}_p \to \mathbb{R}$. We use the polar integration formula for $\ell^n_p$-balls (2.14) as well as the homogeneity of $f$ to deduce that

$$\begin{aligned}
\mathbb{E}h\left(\frac{X}{\|X\|_p}\right) &= C_{n,p,f} \int_{\mathbb{R}^n} f(x)\, e^{-\|x\|_p^p}\, h\left(\frac{x}{\|x\|_p}\right) \mathrm{d}x \\
&= C_{n,p,f}\, n \operatorname{vol}_n(\mathbb{B}^n_p) \int_0^\infty r^{n+m-1}\, e^{-r^p} \mathrm{d}r \int_{\mathbb{S}^{n-1}_p} f(y)\, h(y)\, \mathbf{C}_{n,p}(\mathrm{d}y) \\
&= C_{n,p,f}\, n \operatorname{vol}_n(\mathbb{B}^n_p)\, p^{-1}\, \Gamma\left(\frac{n+m}{p}\right) \int_{\mathbb{S}^{n-1}_p} f(y)\, h(y)\, \mathbf{C}_{n,p}(\mathrm{d}y) \\
&= \int_{\mathbb{S}^{n-1}_p} h(y)\, \mathbf{C}_{n,p,f}(\mathrm{d}y).
\end{aligned}$$

This proves the claim in (i). To show (ii), let $h : \mathbb{B}^n_p \to \mathbb{R}$ be a non-negative measurable function. We notice that if $U$ is uniformly distributed on $[0,1]$, the random variable $U^{\frac{1}{n+m}}$ has density $r \mapsto (n+m)\, r^{n+m-1}$, $r \in [0,1]$, with respect to the Lebesgue measure on $[0,1]$. Using the result from part (i), the homogeneity of $f$, and the polar integration formula for $\ell_p^n$-balls (2.14), we find that

$$
\begin{aligned}
&\mathbf{E}h\left(U^{\frac{1}{n+m}}\frac{X}{\|X\|_p}\right)\\
&\quad= C_{n,p,f}\, n \operatorname{vol}_n(\mathbb{B}_p^n)\, p^{-1}\,\Gamma\left(\frac{n+m}{p}\right)\int_0^1 (n+m)\, r^{n+m-1}\int_{\mathbb{S}_p^{n-1}} f(y)\, h(ry)\,\mathbf{C}_{n,p}(\mathrm{d}y)\,\mathrm{d}r\\
&\quad= C_{n,p,f}\,\Gamma\left(\frac{n+m}{p}\right)\frac{n+m}{p}\, n\operatorname{vol}_n(\mathbb{B}_p^n)\int_0^1 r^{n-1}\int_{\mathbb{S}_p^{n-1}} f(ry)\, h(ry)\,\mathbf{C}_{n,p}(\mathrm{d}y)\,\mathrm{d}r\\
&\quad= C_{n,p,f}\,\Gamma\left(\frac{n+m}{p}+1\right)\int_{\mathbb{B}_p^n} f(x)\, h(x)\,\mathrm{d}x\\
&\quad= C_{n,p,f}\operatorname{vol}_n(\mathbb{B}_p^n)\,\Gamma\left(\frac{n+m}{p}+1\right)\int_{\mathbb{B}_p^n} f(x)\, h(x)\,\mathbf{U}_{n,p}(\mathrm{d}x).\\
&\quad= \int_{\mathbb{B}_p^n} h(x)\,\mathbf{U}_{n,p,f}(\mathrm{d}x).
\end{aligned}
$$

The proof for $\mathbb{B}_p^n$ and $\mathbb{S}_p^{n-1}$ is thus complete. In the non-negative setting one proceeds in the same way, but applies the non-negative polar integration formula from Corollary 2.4.7 for $K=\mathbb{B}_{p,+}^n$. $\square$

**Remark 3.2.2** Note, that the distribution of a random vector $X$ with joint density $C_{n,p,f}\, e^{-\|x\|_p^p} f(x)$, $x\in\mathbb{R}^n$, is just the $n$-fold product distribution $\tilde{\mathbf{N}}_p^{\otimes n}$ of the $p$-generalized Gaussian distribution $\tilde{\mathbf{N}}_p$, weighted by the function $f$ (and appropriately renormalized). So the distribution $\tilde{\mathbf{N}}_p$ is in fact the core building block of the probabilistic representations, but is somewhat implicit in the density of the random vector. In the non-negative case $\tilde{\mathbf{N}}_p$ is replaced by the truncated and renormalized version of $\tilde{\mathbf{N}}_p$ in this role.

Next we present the main result of this section. As stated previously, it is a more general version of [13, Theorem 3] including a homogeneous weight function. Plugging in the function $f\equiv 1$, which is homogeneous of degree $m=0$, yields the original results of [13], again as in Proposition 2.4.5. The proof will work along the lines of that in [13] and relies on multiple applications of the polar integration formula.

**Theorem 3.2.3** *Let* $\mathbf{W}$ *be a Borel probability measure on* $[0,\infty)$*. Let* $0<p<\infty$ *and* $f \in \mathscr{F}_m^+(\mathbb{R}^n_{\boxplus})$ *for some* $m \geq 0$*. Let* $X=(X_1,\dots,X_n)$ *be a random vector with density* $C_{n,p,f,\boxplus}\, e^{-\|x\|_p^p} f(x)$*,* $x \in \mathbb{R}^n_{\boxplus}$*, and* $W$ *a non-negative random variable with distribution* $\mathbf{W}$*, which is independent of* $X$*. Then the random vector*

$$\frac{X}{(\|X\|_p^p + W)^{1/p}}$$

*has distribution* $\mathbf{P}_{n,p,\mathbf{W},f,\boxplus} := \mathbf{W}(\{0\})\mathbf{C}_{n,p,f,\boxplus} + \Psi_f \mathbf{U}_{n,p,f,\boxplus}$*, where* $\Psi_f(x) = \psi_f(\|x\|_p)$*,* $x \in \mathbb{B}^n_{p,\boxplus}$*, is a* $p$*-radial density with*

$$\psi_f(s) = \frac{1}{\Gamma\big(\frac{n+m}{p}+1\big)} \frac{1}{(1-s^p)^{\frac{n+m}{p}+1}} \left[\int_{(0,\infty)} w^{\frac{n+m}{p}}\, e^{-\frac{s^p}{1-s^p}w}\, \mathbf{W}(\mathrm{d}w)\right], \qquad 0 \leq s \leq 1.$$

*Proof.* Let $h:\mathbb{R}^n \to \mathbb{R}$ be an arbitrary non-negative measurable function. Then,

$$\begin{aligned}
\mathbf{E}h\left(\frac{X}{(\|X\|_p^p + W)^{1/p}}\right) &= \int_{[0,\infty)} \mathbf{E}h\left(\frac{X}{(\|X\|_p^p + w)^{1/p}}\right) \mathbf{W}(\mathrm{d}w) \\
&= \int_{[0,\infty)} \mathbf{E}h\left(\left(\frac{\|X\|_p^p}{\|X\|_p^p + w}\right)^{1/p} \frac{X}{\|X\|_p}\right) \mathbf{W}(\mathrm{d}w). \qquad (3.9)
\end{aligned}$$

For fixed $w>0$ we compute the expectation under the integral sign as follows by means of the polar integration formula for $\ell_p^n$-balls (2.14):

$$\begin{aligned}
&\mathbf{E}h\left(\left(\frac{\|X\|_p^p}{\|X\|_p^p + w}\right)^{1/p} \frac{X}{\|X\|_p}\right) \\
&\quad = C_{n,p,f} \int_{\mathbb{R}^n} e^{-\|x\|_p^p} f(x)\, h\left(\left(\frac{\|x\|_p^p}{\|x\|_p^p + w}\right)^{1/p} \frac{x}{\|x\|_p}\right) \mathrm{d}x \\
&\quad = n \operatorname{vol}_n(\mathbb{B}_p^n)\, C_{n,p,f} \int_0^\infty r^{n-1} e^{-r^p} \int_{\mathbb{S}_p^{n-1}} f(ry)\, h\left(\left(\frac{r^p}{r^p+w}\right)^{1/p} y\right) \mathbf{C}_{n,p}(\mathrm{d}y)\,\mathrm{d}r \\
&\quad = n \operatorname{vol}_n(\mathbb{B}_p^n)\, C_{n,p,f} \int_0^\infty r^{n+m-1} e^{-r^p} \int_{\mathbb{S}_p^{n-1}} f(y)\, h\left(\left(\frac{r^p}{r^p+w}\right)^{1/p} y\right) \mathbf{C}_{n,p}(\mathrm{d}y)\,\mathrm{d}r,
\end{aligned}$$

where, in addition, we used the assumption that $f$ is $m$-homogeneous. Applying the change of variables $r^p = \frac{s^p}{1-s^p}\, w$, we get

$$\mathbf{E}h\left(\left(\frac{\|X\|_p^p}{\|X\|_p^p+w}\right)^{1/p}\frac{X}{\|X\|_p}\right)$$
$$= n\,\mathrm{vol}_n(\mathbb{B}_p^n)\,C_{n,p,f}\,w^{\frac{n+m}{p}}\int_0^1\frac{s^{n+m-1}}{(1-s^p)^{\frac{n+m}{p}+1}}\,e^{-\frac{s^p}{1-s^p}w}\int_{\mathbb{S}_p^{n-1}} f(y)\,h(sy)\,\mathbf{C}_{n,p}(\mathrm{d}y)\,\mathrm{d}s.$$

Also, we know from Lemma 3.2.1 (i) that $\frac{X}{\|X\|_p}$ has distribution $\mathbf{C}_{n,p,f}$, which in turn has density

$$y\mapsto C_{n,p,f}\,n\,\mathrm{vol}_n(\mathbb{B}_p^n)\,p^{-1}\,\Gamma\left(\frac{n+m}{p}\right)\,f(y),\qquad y\in\mathbb{S}_p^{n-1},$$

with respect to $\mathbf{C}_{n,p}$. Thus,

$$\mathbf{E}h\left(\left(\frac{\|X\|_p^p}{(\|X\|_p^p+w)^{1/p}}\right)^{1/p}\frac{X}{\|X\|_p}\right)$$
$$= p\,\Gamma\left(\frac{n+m}{p}\right)^{-1}w^{\frac{n+m}{p}}\int_0^1\frac{s^{n+m-1}}{(1-s^p)^{\frac{n+m}{p}+1}}\,e^{-\frac{s^p}{1-s^p}w}\,\mathbf{E}h\left(s\frac{X}{\|X\|_p}\right)\,\mathrm{d}s.$$

As a consequence, recalling (3.9), we see that

$$\mathbf{E}h\left(\frac{X}{(\|X\|_p^p+W)^{1/p}}\right)-\mathbf{W}(\{0\})\mathbf{E}h\left(\frac{X}{\|X\|_p}\right)$$
$$= p\,\Gamma\left(\frac{n+m}{p}\right)^{-1}\int_{(0,\infty)} w^{\frac{n+m}{p}}\int_0^1\frac{s^{n+m-1}}{(1-s^p)^{\frac{n+m}{p}+1}}\,e^{-\frac{s^p}{1-s^p}w}\,\mathbf{E}h\left(s\frac{X}{\|X\|_p}\right)\,\mathrm{d}s\,\mathbf{W}(\mathrm{d}w)$$
$$= \frac{n+m}{\Gamma\left(\frac{n+m}{p}+1\right)}\int_0^1\frac{s^{n+m-1}}{(1-s^p)^{\frac{n+m}{p}+1}}\left[\int_{(0,\infty)} w^{\frac{n+m}{p}}e^{-\frac{s^p}{1-s^p}w}\,\mathbf{W}(\mathrm{d}w)\right]\mathbf{E}h\left(s\frac{X}{\|X\|_p}\right)\,\mathrm{d}s$$
$$= (n+m)\int_0^1 s^{n+m-1}\,\psi_f(s)\,\mathbf{E}h\left(s\frac{X}{\|X\|_p}\right)\,\mathrm{d}s. \tag{3.10}$$

Finally, if $\mathbf{M}$ is any probability measure on $\mathbb{B}_p^n$ with $p$-radial density $\Phi(x)=\phi(\|x\|_p)$, $x\in\mathbb{B}_p^n$, with respect to $\mathbf{U}_{n,p,f}$, the polar integration formula (2.14), together with Lemma 3.2.1 (i), yield the identity

$$\begin{aligned}
\int\limits_{\mathbb{B}_p^n} h(x)\,\mathbf{M}(\mathrm{d}x) &= \int\limits_{\mathbb{B}_p^n} h(x)\,\Phi(x)\,\mathbf{U}_{n,p,f}(\mathrm{d}x) \\
&= C_{n,p,f}\,\Gamma\left(\frac{n+m}{p}+1\right)\int\limits_{\mathbb{B}_p^n} h(x)\,\phi(\|x\|_p)\,f(x)\,\mathrm{d}x \\
&= (n+m)\,C_{n,p,f}\,p^{-1}\,\Gamma\left(\frac{n+m}{p}\right)\,n\,\mathrm{vol}_n(\mathbb{B}_p^n) \\
&\quad\times \int\limits_0^1 \phi(s)\,s^{n+m-1}\int\limits_{\mathbb{S}_p^{n-1}} f(y)\,h(sy)\,\mathbf{C}_{n,p}(\mathrm{d}y)\,\mathrm{d}s \\
&= (n+m)\int\limits_0^1 \phi(s)\,s^{n+m-1}\,\mathbf{E}h\left(s\frac{X}{\|X\|_p}\right)\,\mathrm{d}s. \qquad (3.11)
\end{aligned}$$

The claim follows by comparing (3.10) with (3.11). Again, the same follows in the non-negative setting by using the non-negative polar integration formula from Corollary 2.4.7 for $K=\mathbb{B}_{p,+}^n$. $\square$

**Remark 3.2.4** Taking $f\equiv 1$, which is homogeneous of degree $m=0$, reduces $\mathbf{U}_{n,p,f,\boxplus}$ to $\mathbf{U}_{n,p,\boxplus}$ on $\mathbb{B}_{p,\boxplus}^n$ and $\mathbf{C}_{n,p,f,\boxplus}$ to $\mathbf{C}_{n,p,\boxplus}$. As a consequence, Theorem 3.2.3 turns into [13, Theorem 3] (see Proposition 2.4.5), as already pointed out above.

Let us now consider a few specific distributions for $\mathbf{W}$ and observe the corresponding distributions $\mathbf{P}_{n,p,\mathbf{W},f,\boxplus}$ on $\mathbb{B}_{p,\boxplus}^n$.

**Example 3.2.5** Let $f\in\mathscr{F}_m^+(\mathbb{R}_\boxplus^n)$ and $\mathbf{W}=\delta_0$ be the Dirac measure at 0. Then $\Psi_f\equiv 0$ and $\mathbf{W}(\{0\})=1$, thus for $\mathbf{P}_{n,p,\mathbf{W},f,\boxplus}$ we obtain the weighted cone probability measure $\mathbf{C}_{n,p,f,\boxplus}$ on $\mathbb{B}_{p,\boxplus}^n$.

**Example 3.2.6** Let $f\in\mathscr{F}_m^+(\mathbb{R}_\boxplus^n)$ and $\mathbf{W}=\mathbf{E}(1)$ be the exponential distribution with parameter 1. In this case, we get

$$\begin{aligned}
\psi_f(s) &= \frac{1}{\Gamma\left(\frac{n+m}{p}+1\right)}\frac{1}{(1-s^p)^{\frac{n+m}{p}+1}}\left[\int\limits_{(0,\infty)} w^{\frac{n+m}{p}}e^{-\frac{s^p}{1-s^p}w}\,\mathbf{E}(1)(\mathrm{d}w)\right] \\
&= \frac{1}{\Gamma\left(\frac{n+m}{p}+1\right)}\frac{1}{(1-s^p)^{\frac{n+m}{p}+1}}\left[\int\limits_{(0,\infty)} w^{(\frac{n+m}{p}+1)-1}e^{-\frac{1}{1-s^p}w}\,\mathrm{d}w\right] = 1.
\end{aligned}$$

Thus, $\mathbf{P}_{n,p,\mathbf{W},f,\boxplus}$ is the weighted uniform distribution $\mathbf{U}_{n,p,f,\boxplus}$ on $\mathbb{B}_{p,\boxplus}^n$.

**Example 3.2.7** As a third example, we consider $f \in \mathscr{F}_m^+(\mathbb{R}^n_{\boxplus})$ and $\mathbf{W} = \mathbf{G}(\alpha, 1)$ to be a gamma distribution with shape $\alpha > 0$ and rate 1. In this situation the random variable $X/(\|X\|_p^p + W)^{1/p}$ generates a beta-type distribution $\mathbf{P}_{n,p,\mathbf{W},f,\boxplus} = \Psi_f \mathbf{U}_{n,p,f,\boxplus}$ on $\mathbb{B}^n_{p,\boxplus}$, whose density is a constant multiple of $x \mapsto (1 - \|x\|_p^p)^{\alpha-1}$, $\|x\|_p \leq 1$. To see that, we set $\mathbf{W} = \mathbf{G}(\alpha, 1)$ and compute $\psi_f(s)$ for $s \in [0, 1]$:

$$\begin{aligned}
\psi_f(s) &= \frac{1}{\Gamma\left(\frac{n+m}{p}+1\right)} \frac{1}{(1-s^p)^{\frac{n+m}{p}+1}} \left[ \int\limits_{(0,\infty)} w^{\frac{n+m}{p}} e^{-\frac{s^p}{1-s^p}w}\, \mathbf{G}(\alpha,1)(\mathrm{d}w) \right] \\
&= \frac{1}{\Gamma(\alpha)\Gamma\left(\frac{n+m}{p}+1\right)} \frac{1}{(1-s^p)^{\frac{n+m}{p}+1}} \left[ \int\limits_{(0,\infty)} w^{(\alpha+\frac{n+m}{p})-1} e^{-\frac{1}{1-s^p}w}\, \mathrm{d}w \right] \\
&= \frac{\Gamma\left(\alpha+\frac{n+m}{p}\right)}{\Gamma(\alpha)\Gamma\left(\frac{n+m}{p}+1\right)} (1-s^p)^{\alpha-1}.
\end{aligned}$$

**Remark 3.2.8** In [63, Lemma 4.2] (and Lemma 3.2.1 (ii)) we have seen a different probabilistic representation for the uniform distribution $\mathbf{U}_{n,p,f,\boxplus}$ to that in Example 3.2.7, namely $U^{\frac{1}{n+m}} \frac{X}{\|X\|_p}$, where $U$ is uniformly distributed on $[0, 1]$ and independent of $X$. However, these two representations are equivalent. Indeed, since both are $p$-radially symmetric, it is sufficient to prove that the distributions of the $p$-norms of the random variables $U^{\frac{1}{n+m}} \frac{X}{\|X\|_p}$ and $\frac{X}{(\|X\|_p^p+W)^{1/p}}$ with $W \sim \mathbf{W} = \mathbf{G}(\alpha, 1)$, $\alpha = 1$, are the same. For this we start by noticing that

$$\mathbb{P}\left(\|X\|_p^p \leq t\right) = \mathbb{P}\left(\|X\|_p \leq t^{1/p}\right) = C_{n,p,f} \int\limits_{\{x\in\mathbb{R}^n : \|x\|_p \leq t^{1/p}\}} e^{-\|x\|_p^p} f(x)\, \mathrm{d}x.$$

Using the polar integration formula for $\ell_p^n$-balls (2.14), the fact that $f$ is homogeneous of degree $m$, and the substitution $s = r^p$, we deduce that

$$\begin{aligned}
\mathbb{P}(\|X\|_p^p \leq t) &= C_{n,p,f}\, n \operatorname{vol}_n\left(\mathbb{B}_p^n\right) \int\limits_0^{t^{1/p}} r^{n-1} e^{-r^p} \int\limits_{\mathbb{S}_p^{n-1}} f(ry)\, \mathbf{C}_{n,p}(\mathrm{d}y)\, \mathrm{d}r \\
&= C_{n,p,f}\, n \operatorname{vol}_n\left(\mathbb{B}_p^n\right) \int\limits_0^{t^{1/p}} r^{n+m-1} e^{-r^p} \int\limits_{\mathbb{S}_p^{n-1}} f(y)\, \mathbf{C}_{n,p}(\mathrm{d}y)\, \mathrm{d}r \\
&= \frac{1}{\Gamma\left(\frac{n+m}{p}\right)} \int\limits_0^t s^{\frac{n+m}{p}-1} e^{-s}\, \mathrm{d}s.
\end{aligned}$$

This proves that $\|X\|_p^p \sim \mathbf{G}(\frac{n+m}{p}, 1)$. By the well-known relation between the gamma and the beta distribution (2.6), this implies that

$$\frac{\|X\|_p^p}{\|X\|_p^p + W} \sim \mathbf{B}\left(\frac{n+m}{p}, 1\right).$$

The proof is completed by noting that $U^{\frac{p}{n+m}}$ follows precisely the same beta distribution: For $t \in [0,1]$ it holds that

$$\mathbb{P}\left(U^{\frac{p}{n+m}} \leq t\right) = \mathbb{P}\left(U \leq t^{\frac{n+m}{p}}\right) = t^{\frac{n+m}{p}}.$$

Taking the derivative of the above yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}\left(U^{\frac{p}{n+m}} \leq t\right) = \frac{n+m}{p}\, t^{\frac{n+m}{p}-1} = \frac{\Gamma\left(\frac{n+m}{p}+1\right)}{\Gamma\left(\frac{n+m}{p}\right)\Gamma(1)}\, t^{\frac{n+m}{p}-1}\,(1-t)^{1-1},$$

which shows the aforementioned beta distribution of $U^{\frac{p}{n+m}}$. By analogue arguments the same holds in the non-negative setting. Note that taking the $p$-norm of $\frac{X}{(\|X\|_p^p+W)^{1/p}}$ for $\mathbf{W} = \mathbf{G}(\alpha, 1)$ for some $\alpha > 0$, by the same arguments, yields

$$\frac{\|X\|_p^p}{\|X\|_p^p + W} \sim \mathbf{B}\left(\frac{n+m}{p}, \alpha\right).$$

Again, by analogue arguments the same holds in the non-negative setting.

Choosing $\mathbf{W}$ to be a gamma distribution in the distribution $\mathbf{P}_{n,p,\mathbf{W},f,\boxplus}$ leaves us with $\mathbf{P}_{n,p,\mathbf{W},f,\boxplus} = \Psi_f \mathbf{U}_{n,p,f,\boxplus}$, as $\mathbf{W}(\{0\}) = \mathbf{G}(a,b)(\{0\}) = 0$ for all $a, b > 0$. So, in this case all probability mass is distributed within the interior of $\mathbb{B}^n_{p,\boxplus}$. But we are also interested in cases where a certain amount of probability mass remains at the boundary. For this we consider the mixture $\mathbf{P}_{n,p,\mathbf{W},f,\boxplus} = \vartheta \mathbf{C}_{n,p,f,\boxplus} + (1-\vartheta)\Psi_f \mathbf{U}_{n,p,f,\boxplus}$ for $\vartheta \in [0,1]$, which is simply a convex combination of weighted cone probability measure and weighted uniform distribution or beta-type distribution as in Example 3.2.7. This will be the main class of distributions we will consider in Section 3.4 and Section 3.5 below. In this context, the following two propositions will turn out to be useful. The first one shows that for a specific choice of $\mathbf{W}$ the random vector from Theorem 3.2.3 generates the required distribution. The second deals with the $p$-norm of that random vector.

**Proposition 3.2.9** Let $\vartheta \in [0,1]$, $\alpha \in (0,\infty)$, and consider the probability measure $\mathbf{W} = \vartheta\delta_0 + (1-\vartheta)\,\mathbf{G}(\alpha,1)$. Other than that, we assume the set-up of Theorem 3.2.3. Then the random vector $X/\big(\|X\|_p^p + W\big)^{1/p}$ generates the distribution $\mathbf{P}_{n,p,\mathbf{W},f,\boxplus} = \vartheta\,\mathbf{C}_{n,p,f,\boxplus} + (1-\vartheta)\Psi_f \mathbf{U}_{n,p,f,\boxplus}$, with $\Psi_f \mathbf{U}_{n,p,f,\boxplus}$ on $\mathbb{B}^n_{p,\boxplus}$ being a beta-type distribution.

*Proof.* Let $h : \mathbb{R}^n \to \mathbb{R}$ be a non-negative measurable function. Then, following the arguments in the proof of Theorem 3.2.3 and using the results from Lemma 3.2.1 (i) and Example 3.2.7, we get

$$
\begin{aligned}
&\mathbf{E}h\left(\frac{X}{(\|X\|_p^p + W)^{1/p}}\right)\\
&= \int\limits_{[0,\infty)} \mathbf{E}h\left(\frac{X}{(\|X\|_p^p + w)^{1/p}}\right)\mathbf{W}(\mathrm{d}w)\\
&= \mathbf{E}h\left(\frac{X}{\|X\|_p}\right)\mathbf{W}(\{0\}) + \int\limits_{(0,\infty)} \mathbf{E}h\left(\frac{X}{(\|X\|_p^p + w)^{1/p}}\right)\mathbf{W}(\mathrm{d}w)\\
&= \vartheta \int\limits_{\mathbb{S}_p^{n-1}} h(x)\,\mathbf{C}_{n,p,f}(\mathrm{d}x) + (1-\vartheta)\int\limits_{(0,\infty)} \mathbf{E}h\left(\frac{X}{(\|X\|_p^p + w)^{1/p}}\right)\mathbf{G}(\alpha,1)(\mathrm{d}w)\\
&= \vartheta \int\limits_{\mathbb{S}_p^{n-1}} h(x)\,\mathbf{C}_{n,p,f}(\mathrm{d}x) + (1-\vartheta)\int\limits_{\mathbb{B}_p^n} h(x)\,\Psi_f\mathbf{U}_{n,p,f}(\mathrm{d}x).
\end{aligned}
$$

This completes the proof. □

**Proposition 3.2.10** We assume the same set-up as in Theorem 3.2.3 for the specific choice $\mathbf{W} = \vartheta\delta_0 + (1-\vartheta)\,\mathbf{G}(\alpha,1)$, where $\vartheta \in [0,1]$ and $\alpha \in (0,\infty)$. Then the random variable $B := \|X\|_p^p/\big(\|X\|_p^p + W\big)$ has distribution $\vartheta\delta_1 + (1-\vartheta)\mathbf{B}\big(\frac{n+m}{p},\alpha\big)$.

*Proof.* Let $A \subset \mathbb{R}$ be a Borel set. Then, by the same arguments as in Remark 3.2.8, we get

$$
\begin{aligned}
\mathbb{P}(B \in A) &= \mathbb{P}\left(\frac{\|X\|_p^p}{\|X\|_p^p + W} \in A\right)\\
&= \int\limits_{[0,\infty)} \mathbb{P}\left(\frac{\|X\|_p^p}{\|X\|_p^p + w} \in A\right)\mathbf{W}(\mathrm{d}w)\\
&= \mathbb{P}(1 \in A)\,\vartheta\delta_0(\{0\}) + \int\limits_{(0,\infty)} \mathbb{P}\left(\frac{\|X\|_p^p}{\|X\|_p^p + w} \in A\right)(1-\vartheta)\,\mathbf{G}(\alpha,1)(\mathrm{d}w)\\
&= \vartheta\delta_1(A) + (1-\vartheta)\,\mathbf{B}\left(\frac{n+m}{p},\alpha\right)(A).
\end{aligned}
$$

The proof is thus complete. □

## 3.3 Eigenvalue & singular value distributions on matrix $p$-balls

After having studied weighted $p$-radial distributions in Euclidean $p$-balls, we now turn to the eigenvalue and singular value distributions for self-adjoint and non-self-adjoint random matrices in matrix $p$-balls. To do so, we present two versions of the Weyl integration formula, one for self-adjoint matrices and one for non-self-adjoint matrices. For matrix-functions that only depend on their eigen-/singular values, they allow us to rewrite their integral over matrix space as an integral over Euclidean space with an additional (weight-)function in the integral.

The main results of this section provide the explicit eigen-/singular value distributions for random matrices with distributions $\mathbf{P}^{\mathscr{H}}_{n,p,\mathbf{W},\beta}$ and $\mathbf{P}^{\mathscr{M}}_{n,p,\mathbf{W},\beta}$ on matrix $p$-balls by representing them as weighted $p$-radial distributions on Euclidean $p$-balls, thereby generalizing the probabilistic representations in [63, Corollary 4.3], using a similar method of proof to do so, based on polar integration and the Weyl integration formula.

### 3.3.1 Eigenvalue distribution for self-adjoint random matrices in matrix $p$-balls

We begin by presenting the Weyl integration formula for $\mathscr{H}_n(\mathbb{F}_\beta)$, see [6, Proposition 4.1.1] and also [6, Proposition 4.1.14]. It states that for any measurable function $f : \mathscr{H}_n(\mathbb{F}_\beta) \to [0, \infty)$, such that $f(A)$ only depends on the eigenvalues of $A$, we have

$$\int\limits_{\mathscr{H}_n(\mathbb{F}_\beta)} f(A)\,\mathrm{vol}_{\beta,n}(\mathrm{d}A) = c^{\mathscr{H}}_{n,\beta} \int\limits_{\mathbb{R}^n} f(\lambda) \prod_{1\le i<j\le n} |\lambda_i - \lambda_j|^\beta \,\mathrm{d}\lambda, \tag{3.12}$$

where for every $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$ we write $f(\lambda) = f(A)$ for any self-adjoint matrix $A \in \mathscr{H}_n(\mathbb{F}_\beta)$ with (unordered) eigenvalues $\lambda_1, \dots, \lambda_n$, and the constant $c^{\mathscr{H}}_{n,\beta}$ is given by

$$c^{\mathscr{H}}_{n,\beta} := \frac{1}{n!}\left(\frac{2\pi^{\beta/2}}{\Gamma\big(\frac{\beta}{2}\big)}\right)^{-n} \prod_{k=1}^{n} \frac{2(2\pi)^{\beta k/2}}{2^{\beta/2}\Gamma\big(\frac{\beta k}{2}\big)}.$$

To distinguish between the distributions of random eigenvalues in the standard increasing order and in unordered form, we will use the following version of the Weyl integration formula

$$\int\limits_{\mathscr{H}_n(\mathbb{F}_\beta)} f(A)\,\mathrm{vol}_{\beta,n}(\mathrm{d}A) = n!\, c^{\mathscr{H}}_{n,\beta} \int\limits_{\mathbb{R}^n} f(\lambda) \prod_{1\le i<j\le n} |\lambda_i - \lambda_j|^\beta\, \mathbf{1}_{\{x\in\mathbb{R}^n : x_1 \le \dots \le x_n\}}(\lambda)\,\mathrm{d}\lambda. \tag{3.13}$$

We do so to carry out most of the proof of the main theorem of this section in the more canonical increasingly ordered setting, so we only need to apply an appropriate permutation argument at the very end.

The following theorem shows how the distribution $\mathbf{P}^{\mathscr{H}}_{n,p,\mathbf{W},\beta}$ in matrix $p$-balls is connected to the weighted $p$-radial distribution $\mathbf{P}_{n,p,\mathbf{W},f}$ in Euclidean $p$-balls studied in the previous section. It is derived by application of Weyl's integration formula in connection with the polar integration formula. In the case that $\mathbf{W}(\{0\}) = 0$ it could also be deduced from a classical formula in [104], which is essentially based on the same ingredients, see also [18, Lemma 4.3.1]. However, we present a detailed argument for completeness.

The following functions and normalization terms are needed for said result: For $x \in \mathbb{R}^n$, set

$$\Delta_\beta(x) := \prod_{1\le i<j\le n} |x_i - x_j|^\beta,$$

which is the repulsion factor of the eigenvalues of a random matrix given by the Weyl integration formula (3.12). For a matrix with eigenvalues $\lambda$ note that $\Delta_\beta(\lambda)$ is also its Vandermonde determinant. Additionally, in the spirit of (3.6), define a constant $C_{n,p,\Delta_\beta}$ such that

$$C_{n,p,\Delta_\beta} \int_{\mathbb{R}^n} \Delta_\beta(x)\, e^{-\|x\|_p^p}\, \mathrm{d}x = 1.$$

Further, define the function $\Delta_\beta^c(x) := C_{\Delta_\beta}\, \Delta_\beta(x)$ with a more elaborate normalization factor

$$C_{\Delta_\beta} := \frac{c^{\mathscr{H}}_{n,\beta}}{\mathrm{vol}_{\beta,n}\big(\mathbb{B}^{n,\mathscr{H}}_{p,\beta}\big)\, C_{n,p,\Delta_\beta}\, \Gamma\Big(\frac{n+m}{p}+1\Big)},$$

where $m = \frac{1}{2}\beta n(n-1)$ is the degree of homogeneity of $\Delta_\beta$. Lastly, we define another normalization constant $C_{n,p,\Delta^c_\beta}$ in the spirit of (3.6) satisfying

$$C_{n,p,\Delta^c_\beta} \int_{\mathbb{R}^n} \Delta^c_\beta(x)\, e^{-\|x\|_p^p}\, \mathrm{d}x = 1.$$

**Theorem 3.3.1** *Let $0 < p < \infty$, $\beta \in \{1,2,4\}$ and $\mathbf{W}$ be a probability measure on $[0,\infty)$. Let $W$ be a real-valued random variable with distribution $\mathbf{W}$ and, independently of $W$, $X$ be a random vector with density $C_{n,p,\Delta^c_\beta}\, e^{-\|x\|_p^p}\, \Delta^c_\beta(x)$, $x \in \mathbb{R}^n$, with respect to the Lebesgue measure. Let $Z$ be a random matrix with distribution*

$$\mathbf{P}^{\mathscr{H}}_{n,p,\mathbf{W},\beta} := \mathbf{W}(\{0\})\mathbf{C}^{\mathscr{H}}_{n,p,\beta} + \Psi^{\mathscr{H}}\mathbf{U}^{\mathscr{H}}_{n,p,\beta}$$

*on $\mathbb{B}^{n,\mathscr{H}}_{p,\beta}$, where $\Psi^{\mathscr{H}}(A) := \Psi_{\Delta^c_\beta}(\lambda(A)) = \psi_{\Delta^c_\beta}(\|\lambda(A)\|_p)$ for $A \in \mathbb{B}^{n,\mathscr{H}}_{p,\beta}$, and $\psi_{\Delta^c_\beta}$ is defined as in Theorem 3.2.3 for $f = \Delta^c_\beta$. Independently, let $\sigma$ be a uniform random permutation in the symmetric group on $n$ elements. Then*

$$\lambda_\sigma(Z) := \big(\lambda_{\sigma(1)}(Z), \dots, \lambda_{\sigma(n)}(Z)\big) \quad \textit{and} \quad \frac{X}{(\|X\|_p^p + W)^{1/p}}$$

*are identically distributed on $\mathbb{B}^n_p$ with distribution*

$$\mathbf{P}_{n,p,\mathbf{W},\Delta^c_\beta} := \mathbf{W}(\{0\})\mathbf{C}_{n,p,\Delta^c_\beta} + \Psi_{\Delta^c_\beta}\mathbf{U}_{n,p,\Delta^c_\beta}.$$

**Remark 3.3.2**

(i) If the probability measure $\mathbf{W}$ is the Dirac measure at 0 (that is, $\mathbf{P}^{\mathscr{H}}_{n,p,\mathbf{W},\beta} = \mathbf{C}^{\mathscr{H}}_{n,p,\beta}$) or the exponential distribution with parameter 1 (that is, $\mathbf{P}^{\mathscr{H}}_{n,p,\mathbf{W},\beta} = \mathbf{U}^{\mathscr{H}}_{n,p,\beta}$) the result was previously obtained in [63].

(ii) We can see that the definition $\Psi^{\mathscr{H}}(A) := \psi_{\Delta^c_\beta}(\|\lambda(A)\|_p)$ coincides with that of $\Psi^{\mathscr{H}}(A)$ from (3.1), as the degree of homogeneity $m = \frac{1}{2}\beta n(n-1)$ is the same.

*Proof of Theorem 3.3.1.* Let $h : \mathbb{R}^n \to \mathbb{R}$ be a non-negative measurable function and $\tilde{h} : \mathscr{H}_n(\mathbb{F}_\beta) \to \mathbb{R}$ given by $\tilde{h}(A) := h(\lambda(A))$ for $A \in \mathscr{H}_n(\mathbb{F}_\beta)$. We now compute $\mathbf{E}\tilde{h}(Z)$:

$$\begin{aligned}\mathbf{E}\tilde{h}(Z) &= \int\limits_{\mathscr{H}_n(\mathbb{F}_\beta)} \tilde{h}(A)\ \big(\mathbf{W}(\{0\})\mathbf{C}^{\mathscr{H}}_{n,p,\beta} + \Psi^{\mathscr{H}}\mathbf{U}^{\mathscr{H}}_{n,p,\beta}\big)\,(\mathrm{d}A)\\ &= \mathbf{W}(\{0\}) \int\limits_{\mathbb{S}^{n-1,\mathscr{H}}_{p,\beta}} h(\lambda(A))\,\mathbf{C}^{\mathscr{H}}_{n,p,\beta}(\mathrm{d}A) + \int\limits_{\mathbb{B}^{n,\mathscr{H}}_{p,\beta}} h(\lambda(A))\,\Psi^{\mathscr{H}}\mathbf{U}^{\mathscr{H}}_{n,p,\beta}(\mathrm{d}A). \qquad (3.14)\end{aligned}$$

Consider the radial extension $h\big(\lambda(A)/\|\lambda(A)\|_p\big)$, $A \in \mathscr{H}_n(\mathbb{F}_\beta)$, of $h(\lambda(A))$ from $\mathbb{S}^{n-1,\mathscr{H}}_{p,\beta}$ onto $\mathscr{H}_n(\mathbb{F}_\beta)$. By Remark 3.1.1 (i) we can apply the polar integration formula from Lemma 2.4.6 to $\mathbb{B}^{n,\mathscr{H}}_{p,\beta}$ to get

$$\begin{aligned}&\int\limits_{\mathbb{B}^{n,\mathscr{H}}_{p,\beta}} h\left(\frac{\lambda(A)}{\|\lambda(A)\|_p}\right)\mathbf{U}^{\mathscr{H}}_{n,p,\beta}(\mathrm{d}A)\\ &= \left(\frac{\beta n(n-1)}{2} + \beta n\right)\int\limits_0^1 r^{\frac{\beta n(n-1)}{2}+\beta n-1} \int\limits_{\mathbb{S}^{n-1,\mathscr{H}}_{p,\beta}} h(\lambda(A))\,\mathbf{C}^{\mathscr{H}}_{n,p,\beta}(\mathrm{d}A)\,\mathrm{d}r\\ &= \int\limits_{\mathbb{S}^{n-1,\mathscr{H}}_{p,\beta}} h(\lambda(A))\,\mathbf{C}^{\mathscr{H}}_{n,p,\beta}(\mathrm{d}A). \qquad (3.15)\end{aligned}$$

With (3.15) it follows that (3.14) can be rewritten as:

$$\begin{aligned}\mathbf{E}\tilde{h}(Z) &= \mathbf{W}(\{0\})\,\mathrm{vol}_{\beta,n}\left(\mathbb{B}^{n,\mathscr{H}}_{p,\beta}\right)^{-1}\int\limits_{\mathbb{B}^{n,\mathscr{H}}_{p,\beta}} h\left(\frac{\lambda(A)}{\|\lambda(A)\|_p}\right)\mathrm{vol}_{\beta,n}(\mathrm{d}A)\\ &\quad+\mathrm{vol}_{\beta,n}\left(\mathbb{B}^{n,\mathscr{H}}_{p,\beta}\right)^{-1}\int\limits_{\mathbb{B}^{n,\mathscr{H}}_{p,\beta}} h(\lambda(A))\Psi^{\mathscr{H}}(A)\mathrm{vol}_{\beta,n}(\mathrm{d}A).\end{aligned}$$

To both of those terms on the right-hand side we can now apply the "ordered" Weyl integration formula (3.13) with respect to the functions $f_1(A) = h\big(\lambda(A)/\|\lambda(A)\|_p\big)$ and $f_2(A) = h(\lambda(A))\,\Psi^{\mathscr{H}}(\lambda(A))$. It follows from $\Delta_\beta \in \mathscr{F}^+_m(\mathbb{R}^n)$ with $m = \frac{1}{2}\beta n(n-1)$ and $\Psi^{\mathscr{H}} = \Psi_{\Delta_\beta} = \Psi_{\Delta^c_\beta}$ that

$$\begin{aligned}\mathbf{E}\tilde{h}(Z) &= \frac{n!\,c^{\mathscr{H}}_{n,\beta}}{\mathrm{vol}_{\beta,n}\left(\mathbb{B}^{n,\mathscr{H}}_{p,\beta}\right)}\quad\mathbf{W}(\{0\})\int\limits_{\mathbb{B}^n_p} h\Big(\frac{\lambda}{\|\lambda\|_p}\Big)\,\Delta_\beta(\lambda)\,\mathbf{1}_{\{x\in\mathbb{R}^n:x_1\le\ldots\le x_n\}}(\lambda)\,\mathrm{d}\lambda\\ &\quad+\frac{n!\,c^{\mathscr{H}}_{n,\beta}}{\mathrm{vol}_{\beta,n}\left(\mathbb{B}^{n,\mathscr{H}}_{p,\beta}\right)}\int\limits_{\mathbb{B}^n_p} h(\lambda)\,\Psi_{\Delta^c_\beta}(\lambda)\,\Delta_\beta(\lambda)\,\mathbf{1}_{\{x\in\mathbb{R}^n:x_1\le\ldots\le x_n\}}(\lambda)\,\mathrm{d}\lambda\\ &= \frac{n!\,c^{\mathscr{H}}_{n,\beta}}{\mathrm{vol}_{\beta,n}\big(\mathbb{B}^{n,\mathscr{H}}_{p,\beta}\big)}\quad\mathbf{W}(\{0\})\int\limits_{\mathbb{B}^n_p} h\Big(\frac{\lambda}{\|\lambda\|_p}\Big)\,\Delta_\beta\Big(\frac{\lambda}{\|\lambda\|_p}\Big)\|\lambda\|_p^m\,\mathbf{1}_{\{x\in\mathbb{R}^n:x_1\le\ldots\le x_n\}}(\lambda)\,\mathrm{d}\lambda\\ &\quad+\frac{n!\,c^{\mathscr{H}}_{n,\beta}}{\mathrm{vol}_{\beta,n}\big(\mathbb{B}^{n,\mathscr{H}}_{p,\beta}\big)}\int\limits_{\mathbb{B}^n_p} h(\lambda)\,\Psi_{\Delta^c_\beta}(\lambda)\,\Delta_\beta(\lambda)\,\mathbf{1}_{\{x\in\mathbb{R}^n:x_1\le\ldots\le x_n\}}(\lambda)\,\mathrm{d}\lambda.\end{aligned}$$

Applying now the polar integration formula from Lemma 2.4.6, we conclude that the last expression is equal to

$$\begin{aligned}&\frac{n!\,c^{\mathscr{H}}_{n,\beta}\,n\,\mathrm{vol}_n(\mathbb{B}^n_p)}{\mathrm{vol}_{\beta,n}\left(\mathbb{B}^{n,\mathscr{H}}_{p,\beta}\right)}\quad\mathbf{W}(\{0\})\int_0^1 r^{n+m-1}\mathrm{d}r\int\limits_{\mathbb{S}^{n-1}_p} h(\lambda)\,\Delta_\beta(\lambda)\,\mathbf{1}_{\{x\in\mathbb{R}^n:x_1\le\ldots\le x_n\}}(\lambda)\,\mathbf{C}_{n,p}(\mathrm{d}\lambda)\\ &\quad+\frac{n!\,c^{\mathscr{H}}_{n,\beta}}{\mathrm{vol}_{\beta,n}\left(\mathbb{B}^{n,\mathscr{H}}_{p,\beta}\right)}\int\limits_{\mathbb{B}^n_p} h(\lambda)\,\Psi_{\Delta^c_\beta}(\lambda)\,\Delta_\beta(\lambda)\,\mathbf{1}_{\{x\in\mathbb{R}^n:x_1\le\ldots\le x_n\}}(\lambda)\,\mathrm{d}\lambda\\ &= \frac{n!\,c^{\mathscr{H}}_{n,\beta}\,n\,\mathrm{vol}_n(\mathbb{B}^n_p)}{\mathrm{vol}_{\beta,n}\left(\mathbb{B}^{n,\mathscr{H}}_{p,\beta}\right)(n+m)}\quad\mathbf{W}(\{0\})\int\limits_{\mathbb{S}^{n-1}_p} h(\lambda)\,\Delta_\beta(\lambda)\,\mathbf{1}_{\{x\in\mathbb{R}^n:x_1\le\ldots\le x_n\}}(\lambda)\,\mathbf{C}_{n,p}(\mathrm{d}\lambda)\\ &\quad+\quad\frac{n!\,c^{\mathscr{H}}_{n,\beta}\,\mathrm{vol}_n(\mathbb{B}^n_p)}{\mathrm{vol}_{\beta,n}\left(\mathbb{B}^{n,\mathscr{H}}_{p,\beta}\right)}\int\limits_{\mathbb{B}^n_p} h(\lambda)\,\Psi_{\Delta^c_\beta}(\lambda)\,\Delta_\beta(\lambda)\,\mathbf{1}_{\{x\in\mathbb{R}^n:x_1\le\ldots\le x_n\}}(\lambda)\,\mathbf{U}_{n,p}(\mathrm{d}\lambda).\end{aligned}$$

Next, we use the definition of $\mathbf{U}_{n,p,f}$ and $\mathbf{C}_{n,p,f}$ for $f = \Delta_\beta$ and the definition of $\Delta_\beta^c = C_{\Delta_\beta}\,\Delta_\beta$. This gives

$$\begin{aligned}
&\mathbf{E}\tilde{h}(Z) \\
&= \frac{n!\,c^{\mathscr{H}}_{n,\beta}}{\mathrm{vol}_{\beta,n}\left(\mathbb{B}^{n,\mathscr{H}}_{p,\beta}\right)\, C_{n,p,\Delta_\beta}\,\Gamma\!\left(\frac{n+m}{p}+1\right)}\mathbf{W}(\{0\}) \int\limits_{\mathbb{S}_p^{n-1}} h(\lambda)\,\mathbf{1}_{\{x\in\mathbb{R}^n : x_1\le\ldots\le x_n\}}(\lambda)\,\mathbf{C}_{n,p,\Delta_\beta}(\mathrm{d}\lambda) \\
&\quad + \frac{n!\,c^{\mathscr{H}}_{n,\beta}}{\mathrm{vol}_{\beta,n}\left(\mathbb{B}^{n,\mathscr{H}}_{p,\beta}\right)\, C_{n,p,\Delta_\beta}\,\Gamma\!\left(\frac{n+m}{p}+1\right)} \int\limits_{\mathbb{B}_p^n} h(\lambda)\,\Psi_{\Delta_\beta^c}(\lambda)\,\mathbf{1}_{\{x\in\mathbb{R}^n : x_1\le\ldots\le x_n\}}(\lambda)\,\mathbf{U}_{n,p,\Delta_\beta}(\mathrm{d}\lambda) \\
&= n!\,\mathbf{W}(\{0\}) \int\limits_{\mathbb{S}_p^{n-1}} h(\lambda)\,\mathbf{1}_{\{x\in\mathbb{R}^n : x_1\le\ldots\le x_n\}}(\lambda)\,\mathbf{C}_{n,p,\Delta_\beta^c}(\mathrm{d}\lambda) \\
&\quad + n! \int\limits_{\mathbb{B}_p^n} h(\lambda)\,\Psi_{\Delta_\beta^c}(\lambda)\,\mathbf{1}_{\{x\in\mathbb{R}^n : x_1\le\ldots\le x_n\}}(\lambda)\,\mathbf{U}_{n,p,\Delta_\beta^c}(\mathrm{d}\lambda).
\end{aligned}$$

As a consequence, when applying a uniform random permutation $\sigma$ in the symmetric group on $n$ elements, we get by Theorem 3.2.3

$$\begin{aligned}
\mathbf{E}(\tilde{h}\circ\sigma)(Z) &= \mathbf{W}(\{0\}) \int\limits_{\mathbb{S}_p^{n-1}} h(\lambda)\,\mathbf{C}_{n,p,\Delta_\beta^c}(\mathrm{d}\lambda) + \int\limits_{\mathbb{B}_p^n} h(\lambda)\,\Psi_{\Delta_\beta^c}(\lambda)\,\mathbf{U}_{n,p,\Delta_\beta^c}(\mathrm{d}\lambda) \\
&= \mathbf{E}h\left(\frac{X}{(\|X\|_p^p + W)^{1/p}}\right).
\end{aligned}$$

This proves the claim. □

**Remark 3.3.3** Let us briefly show that for $\Delta_\beta^c$ the degree of homogeneity is in fact $m = \frac{\beta n(n-1)}{2}$. Let $s \in [0,\infty), x \in \mathbb{R}^n, \beta \in \{1,2,4\}$, then

$$\begin{aligned}
\Delta_\beta^c(sx) &= C_{\Delta_\beta} \prod_{1\le i<j\le n} |sx_i - sx_j|^\beta \\
&= C_{\Delta_\beta} \prod_{i=1}^{n}\prod_{j=i}^{n} s^\beta |x_i - x_j|^\beta \\
&= C_{\Delta_\beta} \prod_{i=1}^{n} s^{(n-i)\beta} \prod_{j=i}^{n} |x_i - x_j|^\beta \\
&= s^{\beta\sum_{i=0}^{n-1} i}\, C_{\Delta_\beta} \prod_{i=1}^{n}\prod_{j=i}^{n} |x_i - x_j|^\beta \\
&= s^{\beta\frac{(n-1)n}{2}}\, \Delta_\beta^c(x).
\end{aligned}$$

**Remark 3.3.4** Since the degree of homogeneity of $\Delta_\beta^c$ is $m = \frac{\beta n(n-1)}{2}$, if one considers the set-up of Theorem 3.2.3 for $f = \Delta_\beta^c$ and $\mathbf{W} = \mathbf{G}(\alpha, 1)$ for $\alpha > 0$, by the results outlined in Remark 3.2.8, we have that

$$\frac{\|X\|_p^p}{\|X\|_p^p + W} \sim \mathbf{B}\left(\frac{\beta n^2}{2p} - \frac{\beta n}{2p} + \frac{n}{p}\,,\, \alpha\right),$$

and for $\mathbf{W} = \vartheta\delta_0 + (1-\vartheta)\,\mathbf{G}(\alpha, 1)$, where $\vartheta \in [0,1]$ and $\alpha \in (0,\infty)$, by the arguments from Proposition 3.2.10, we have that

$$\frac{\|X\|_p^p}{\|X\|_p^p + W} \sim \vartheta\delta_1 + (1-\vartheta)\mathbf{B}\left(\frac{\beta n^2}{2p} - \frac{\beta n}{2p} + \frac{n}{p}\,,\, \alpha\right).$$

### 3.3.2 Singular value distribution for non-self-adjoint random matrices in matrix $p$-balls

Let us now consider the non-self-adjoint case, where the singular values take over the role of the eigenvalues. The following result is proven by almost literally repeating the proof of Theorem 3.3.1 (or, at least in the case that $\mathbf{W}(\{0\}) = 0$, by applying a formula from [104], which corresponds to [18, Lemma 4.3.1] as we explained before Theorem 3.3.1). However, this time the argument is based on the Weyl-type integration formula from [6, Proposition 4.1.3], which replaces (3.12). This formula primarily changes the repulsion factor from $\Delta_\beta$ to an appropriate $\nabla_\beta$ and the normalization constant from $c_{n,\beta}^{\mathscr{H}}$ to $c_{n,\beta}^{\mathscr{M}}$ as follows: It says that for any non-negative measurable function $f : \mathscr{M}_n(\mathbb{F}_\beta) \to [0,\infty)$, such that $f(A)$ only depends on the singular values of $A$, we have that

$$\int\limits_{\mathscr{M}_n(\mathbb{F}_\beta)} f(A)\,\mathrm{vol}_{\beta,n}(\mathrm{d}A) = c_{n,\beta}^{\mathscr{M}} \int\limits_{\mathbb{R}_+^n} f(s) \prod_{1\le i<j\le n} \left|s_i^2 - s_j^2\right|^\beta \prod_{i=1}^n s_i^{\,\beta-1}\,\mathrm{d}s, \tag{3.16}$$

writing $f(s) = f(A)$ for any matrix $A \in \mathscr{M}_n(\mathbb{F}_\beta)$ with (unordered) singular values $(s_1, \ldots, s_n) \in \mathbb{R}_+^n$, and where

$$c_{n,\beta}^{\mathscr{M}} := \frac{1}{n!}\frac{1}{2^{\frac{\beta}{2}n(n-1)}}\left(\frac{2\pi^{\beta/2}}{\Gamma\left(\frac{\beta}{2}\right)}\right)^{-n} \prod_{k=1}^n \left(\frac{2(2\pi)^{\beta k/2}}{2^{\beta/2}\Gamma\left(\frac{\beta k}{2}\right)}\right)^2.$$

To apply the same arguments as previously, we only consider the non-negative measurable function $f : \mathscr{M}_n(\mathbb{F}_\beta) \to [0,\infty)$, such that $f(A)$ only depends on the vector $s^2(A) := (s_1^2(A), \ldots, s_n^2(A))$ of *squared* singular values of $A$, i.e., $f(A) = f(s^2(A))$. For such a function, we derive an "ordered version" of the Weyl integration formula to shift the necessity for permutations to the end of the proof, and additionally apply some change of variables argument to get:

$$\int_{\mathscr{M}_n(\mathbb{F}_\beta)} f(A)\,\mathrm{vol}_{\beta,n}(\mathrm{d}A)$$

$$= n!\, c^{\mathscr{M}}_{n,\beta}\, 2^{-n} \int_{\mathbb{R}^n_+} f\big(s^2\big) \prod_{1\le i<j\le n} \big|s_i^2 - s_j^2\big|^\beta \prod_{i=1}^n \big(s_i^2\big)^{\frac{\beta}{2}-1} \mathbf{1}_{\big\{x\in\mathbb{R}^n_+ : x_1\le\ldots\le x_n\big\}}\big(s^2\big)\,\mathrm{d}s^2. \quad (3.17)$$

As discussed in Section 3.1.1, the vectors $s(A)$ and $s^2(A)$ of (squared) singular values of a matrix $A \in \mathscr{M}_n(\mathbb{F}_\beta)$ live in the non-negative orthant $\mathbb{B}^n_{p,+}$ of the $\ell^n_p$-ball $\mathbb{B}^n_p$. Furthermore, the matrix $p$-ball $\mathbb{B}^{n,\mathscr{M}}_{p,\beta}$ will be represented in Euclidean space via $\mathbb{B}^n_{p/2,+}$, not $\mathbb{B}^n_{p,+}$, due to the structure of the Weyl integration formula for singular values in (3.17). Since it uses the squares of the singular values in its repulsion factor, we adapt our representation appropriately, such that the same arguments as for the eigenvalues are applicable. Thus, we reformulate the defining condition of $\mathbb{B}^{n,\mathscr{M}}_{p,\beta}$ from $\sum_{i=1}^n |s_i(A)|^p \le 1$ to $\sum_{i=1}^n |s_i^2(A)|^{p/2} \le 1$, and apply the same arguments as before to the vector $s^2(A)$, which then in turn lies in $\mathbb{B}^n_{p/2,+}$.

As in the self-adjoint setting, we need to define some functions and normalization terms to formulate the next result. For $x \in \mathbb{R}^n_+$ we set

$$\nabla_\beta(x) := \prod_{1\le i<j\le n} |x_i - x_j|^\beta \prod_{i=1}^n x_i^{\frac{\beta}{2}-1},$$

which again is the repulsion factor of singular values from the Weyl integration formula (3.17), and define $C_{n,p,\nabla_\beta,+}$ to be the normalization constant such that

$$C_{n,p,\nabla_\beta} \int_{\mathbb{R}^n_+} \nabla_\beta(x)\, e^{-\|x\|_p^p}\,\mathrm{d}x = 1.$$

Based on this definition, we further set $\nabla^c_\beta(x) := C_{\nabla_\beta}\,\nabla_\beta(x)$ for $x \in \mathbb{R}^n_+$ with

$$C_{\nabla_\beta} := \frac{c^{\mathscr{M}}_{n,\beta}}{\mathrm{vol}_{\beta,n}\Big(\mathbb{B}^{n,\mathscr{M}}_{p,\beta}\Big)\; C_{n,p/2,\nabla_\beta}\,\Gamma\Big(\frac{n+m}{p/2}+1\Big)2^n},$$

where $m = \frac{\beta}{2}n^2 - n$ is the degree of homogeneity of $\nabla^c_\beta$. A final normalization constant $C_{n,p,\nabla^c_\beta,+}$ is defined by

$$C_{n,p,\nabla^c_\beta} \int_{\mathbb{R}^n_+} \nabla^c_\beta(x)\, e^{-\|x\|_p^p}\,\mathrm{d}x = 1.$$

**Theorem 3.3.5** *Let $0 < p < \infty$, $\beta \in \{1,2,4\}$ and $\mathbf{W}$ be a probability measure on $[0,\infty)$. Let $W$ be a real-valued random variable with density $\mathbf{W}$ and, independently of $W$, $X$ be a random vector with distribution given by the density $C_{n,p/2,\nabla^c_\beta,+}\, e^{-\|x\|_{p/2}^{p/2}}\, \nabla^c_\beta(x)$, $x \in \mathbb{R}^n_+$, with respect to the Lebesgue measure . Let $\sigma$ be a uniform random permutation in the symmetric group on $n$ elements and $Z$ be a random matrix with distribution*

$$\mathbf{P}^{\mathscr{M}}_{n,p,\mathbf{W},\beta} := \mathbf{W}(\{0\})\mathbf{C}^{\mathscr{M}}_{n,p,\beta} + \Psi^{\mathscr{M}}\mathbf{U}^{\mathscr{M}}_{n,p,\beta}$$

*on $\mathbb{B}^{n,\mathscr{M}}_{p,\beta}$, where $\Psi^{\mathscr{M}}(A) := \Psi_{\nabla^c_\beta}(s(A)) = \psi_{\nabla^c_\beta}(\|s(A)\|_p)$ for $A \in \mathbb{B}^{n,\mathscr{M}}_{p,\beta}$, and $\psi_{\nabla^c_\beta}$ is defined as in Theorem 3.2.3 for $f = \nabla^c_\beta$. Then*

$$s^2_\sigma(Z) := \big(s^2_{\sigma(1)}(Z), \dots, s^2_{\sigma(n)}(Z)\big) \qquad \textit{and} \qquad \frac{X}{(\|X\|_{p/2}^{p/2} + W)^{2/p}}$$

*are identically distributed on $\mathbb{B}^n_{p/2,+}$ with distribution*

$$\mathbf{P}_{n,p/2,\mathbf{W},\nabla^c_\beta,+} = \mathbf{W}(\{0\})\mathbf{C}_{n,p/2,\nabla^c_\beta,+} + \Psi_{\nabla^c_\beta}\mathbf{U}_{n,p/2,\nabla^c_\beta,+}\,.$$

The proof of this goes along the same lines as that of Theorem 3.3.1, just using the representation results from Theorem 3.2.3 in the non-negative setting and the Weyl integration formula from (3.17) instead of (3.13). We do, however, include it for the self-containedness of this chapter.

*Proof of Theorem 3.3.5.* Let $h : \mathbb{R}^n \to \mathbb{R}$ be a non-negative and measurable function and consider $\tilde{h} : \mathscr{M}_n(\mathbb{F}_\beta) \to \mathbb{R}$ with $\tilde{h}(A) := h(s^2(A))$. We proceed to compute $\mathbf{E}\tilde{h}(Z)$:

$$\begin{aligned}
\mathbf{E}\tilde{h}(Z) &= \int\limits_{\mathscr{M}_n(\mathbb{F}_\beta)} \tilde{h}(A)\,\big(\mathbf{W}(\{0\})\mathbf{C}^{\mathscr{M}}_{n,p,\beta} + \Psi^{\mathscr{M}}\mathbf{U}^{\mathscr{M}}_{n,p,\beta}\big)(\mathrm{d}A) \\
&= \mathbf{W}(\{0\}) \int\limits_{\mathbb{S}^{n-1,\mathscr{M}}_{p,\beta}} h\big(s^2(A)\big)\,\mathbf{C}^{\mathscr{M}}_{n,p,\beta}(\mathrm{d}A) + \int\limits_{\mathbb{B}^{n,\mathscr{M}}_{p,\beta}} h\big(s^2(A)\big)\,\Psi^{\mathscr{M}}\mathbf{U}^{\mathscr{M}}_{n,p,\beta}(\mathrm{d}A).
\end{aligned}$$

Again, one considers the radial extension of $h(s^2(A))$ from $\mathbb{S}^{n-1,\mathscr{M}}_{p,\beta}$ onto $\mathscr{M}_n(\mathbb{F}_\beta)$, which has the form $h\big(s^2(A)/\|s^2(A)\|_{p/2}\big)$, $A \in \mathscr{M}_n(\mathbb{F}_\beta)$, since we are considering $s^2(A)$ instead of $s(A)$. Following Remark 3.1.1 (i), the polar integration formula from Lemma 2.4.6 applied to $\mathbb{B}^{n,\mathscr{M}}_{p,\beta}$ gives

$$\begin{aligned}
\int\limits_{\mathbb{B}^{n,\mathscr{M}}_{p,\beta}} h\left(\frac{s^2(A)}{\|s^2(A)\|_{p/2}}\right)\mathbf{U}^{\mathscr{M}}_{n,p,\beta}(\mathrm{d}A) &= \beta n^2 \int\limits_0^1 r^{\beta n^2-1} \int\limits_{\mathbb{S}^{n-1,\mathscr{M}}_{p,\beta}} h(s^2(A))\,\mathbf{C}^{\mathscr{M}}_{n,p,\beta}(\mathrm{d}A)\mathrm{d}r \\
&= \int\limits_{\mathbb{S}^{n-1,\mathscr{M}}_{p,\beta}} h\big(s^2(A)\big)\,\mathbf{C}^{\mathscr{M}}_{n,p,\beta}(\mathrm{d}A). \qquad (3.18)
\end{aligned}$$

Rewriting $\mathbf{E}\tilde{h}(Z)$ via (3.18) implies that

$$\begin{aligned}\mathbf{E}\tilde{h}(Z) =&\mathbf{W}(\{0\})\,\mathrm{vol}_{\beta,n}\left(\mathbb{B}^{n,\mathscr{M}}_{p,\beta}\right)^{-1}\int\limits_{\mathbb{B}^{n,\mathscr{M}}_{p,\beta}} h\left(\frac{s^2(A)}{\|s^2(A)\|_{p/2}}\right)\mathrm{vol}_{\beta,n}(\mathrm{d}A)\\ &+\mathrm{vol}_{\beta,n}\left(\mathbb{B}^{n,\mathscr{M}}_{p,\beta}\right)^{-1}\int\limits_{\mathbb{B}^{n,\mathscr{M}}_{p,\beta}} h(s^2(A))\Psi^{\mathscr{M}}(A)\mathrm{vol}_{\beta,n}(\mathrm{d}A).\end{aligned}$$

We apply the Weyl integration formula for non-self-adjoint matrices (3.17) with respect to the functions $f_1(A) = h\big(s^2(A)/\|s^2(A)\|_{p/2}\big)$ and $f_2(A) = h(s^2(A))\,\Psi^{\mathscr{M}}(s^2(A))$. The repulsion factor therein is given by $\nabla_\beta$. Further, it holds that $\nabla_\beta \in \mathscr{F}^+_m(\mathbb{R}^n_+)$ with $m = \frac{1}{2}\beta n^2 - n$ and $\Psi^{\mathscr{M}} = \Psi_{\nabla_\beta} = \Psi_{\nabla^c_\beta}$ (since they only depend on $m$), from which follows

$$\begin{aligned}\mathbf{E}\tilde{h}(Z) &= \frac{n!\,c^{\mathscr{M}}_{n,\beta}}{\mathrm{vol}_{\beta,n}\left(\mathbb{B}^{n,\mathscr{M}}_{p,\beta}\right)2^n}\,\mathbf{W}(\{0\})\int\limits_{\mathbb{B}^n_{p/2}} h\left(\frac{s^2}{\|s^2\|_{p/2}}\right)\,\nabla_\beta\big(s^2\big)\,\mathbf{1}_{\{x\in\mathbb{R}^n_+ : x_1<\ldots<x_n\}}\big(s^2\big)\,\mathrm{d}s^2\\ &\quad+\frac{n!\,c^{\mathscr{M}}_{n,\beta}}{\mathrm{vol}_{\beta,n}\left(\mathbb{B}^{n,\mathscr{H}}_{p,\beta}\right)2^n}\int\limits_{\mathbb{B}^n_{p/2}} h\big(s^2\big)\,\Psi_{\nabla^c_\beta}\big(s^2\big)\,\nabla_\beta(s)\,\mathbf{1}_{\{x\in\mathbb{R}^n_+ : x_1<\ldots<x_n\}}\big(s^2\big)\mathrm{d}s^2\\ &= \frac{n!\,c^{\mathscr{M}}_{n,\beta}}{\mathrm{vol}_{\beta,n}\left(\mathbb{B}^{n,\mathscr{H}}_{p,\beta}\right)2^n}\,\mathbf{W}(\{0\})\\ &\qquad\int\limits_{\mathbb{B}^n_{p/2}} h\left(\frac{s^2}{\|s^2\|_{p/2}}\right)\,\nabla_\beta\left(\frac{s^2}{\|s^2\|_{p/2}}\right)\|s^2\|^m_{p/2}\,\mathbf{1}_{\{x\in\mathbb{R}^n : x_1<\ldots<x_n\}}(s)\,\mathrm{d}s^2\\ &\quad+\frac{n!\,c^{\mathscr{M}}_{n,\beta}}{\mathrm{vol}_{\beta,n}\left(\mathbb{B}^{n,\mathscr{M}}_{p,\beta}\right)2^n}\int\limits_{\mathbb{B}^n_{p/2}} h\big(s^2\big)\,\Psi_{\nabla^c_\beta}\big(s^2\big)\,\nabla_\beta\big(s^2\big)\,\mathbf{1}_{\{x\in\mathbb{R}^n_+ : x_1<\ldots<x_n\}}\big(s^2\big)\,\mathrm{d}s^2.\end{aligned}$$

We use the polar integration formula from Lemma 2.4.6 for $K = \mathbb{B}^n_{p/2}$, by which it follows that the above is equal to

$$\begin{aligned}&\frac{n!\,c^{\mathscr{M}}_{n,\beta}\,n\,\mathrm{vol}_n(\mathbb{B}^n_{p/2})}{\mathrm{vol}_{\beta,n}\left(\mathbb{B}^{n,\mathscr{M}}_{p,\beta}\right)2^n}\mathbf{W}(\{0\})\int\limits_0^1 r^{n+m-1}\mathrm{d}r\\ &\qquad\times\int\limits_{\mathbb{S}^{n-1}_{p/2}} h\big(s^2\big)\,\nabla_\beta\big(s^2\big)\,\mathbf{1}_{\{x\in\mathbb{R}^n_+ : x_1<\ldots<x_n\}}\big(s^2\big)\,\mathbf{C}_{n,p/2}\big(\mathrm{d}s^2\big)\\ &+\frac{n!\,c^{\mathscr{M}}_{n,\beta}}{\mathrm{vol}_{\beta,n}\left(\mathbb{B}^{n,\mathscr{H}}_{p,\beta}\right)2^n}\int\limits_{\mathbb{B}^n_{p/2}} h\big(s^2\big)\,\Psi_{\nabla^c_\beta}\big(s^2\big)\,\nabla_\beta\big(s^2\big)\,\mathbf{1}_{\{x\in\mathbb{R}^n_+ : x_1<\ldots<x_n\}}\big(s^2\big)\,\mathrm{d}s^2\end{aligned}$$

$$
\begin{aligned}
&= \frac{n!\, c_{n,\beta}^{\mathscr{M}}\, n\, \mathrm{vol}_n(\mathbb{B}_{p/2}^n)}{\mathrm{vol}_{\beta,n}\left(\mathbb{B}_{p,\beta}^{n,\mathscr{M}}\right)(n+m)2^n}\, \mathbf{W}(\{0\}) \int\limits_{\mathbb{S}_{p/2}^{n-1}} h\big(s^2\big)\, \nabla_\beta\big(s^2\big)\, \mathbf{1}_{\left\{x\in\mathbb{R}_+^n : x_1<\ldots<x_n\right\}}\big(s^2\big)\, \mathbf{C}_{n,p/2}\big(\mathrm{d}s^2\big) \\
&\quad + \frac{n!\, c_{n,\beta}^{\mathscr{M}}\, \mathrm{vol}_n(\mathbb{B}_{p/2}^n)}{\mathrm{vol}_{\beta,n}\left(\mathbb{B}_{p,\beta}^{n,\mathscr{M}}\right)2^n} \int\limits_{\mathbb{B}_{p/2}^n} h\big(s^2\big)\, \Psi_{\nabla_\beta^c}\big(s^2\big)\, \nabla_\beta\big(s^2\big)\, \mathbf{1}_{\left\{x\in\mathbb{R}_+^n : x_1<\ldots<x_n\right\}}\big(s^2\big)\, \mathbf{U}_{n,p/2}\big(\mathrm{d}s^2\big).
\end{aligned}
$$

By the definition of $\mathbf{U}_{n,p/2,f}$ and $\mathbf{C}_{n,p/2,f}$ for $f = \nabla_\beta$ and $\nabla_\beta^c := C_{\nabla_\beta} \nabla_\beta$ it thereby follows that

$$
\begin{aligned}
\mathbf{E}\tilde{h}(Z) = \quad & \frac{n!\, c_{n,\beta}^{\mathscr{M}}}{\mathrm{vol}_{\beta,n}\left(\mathbb{B}_{p,\beta}^{n,\mathscr{M}}\right) C_{n,p,\nabla_\beta}\, \Gamma\Big(\frac{n+m}{p/2}+1\Big)2^n} \\
& \times \mathbf{W}(\{0\}) \int\limits_{\mathbb{S}_{p/2}^{n-1}} h\big(s^2\big)\, \mathbf{1}_{\{x\in\mathbb{R}_+^n : x_1<\ldots<x_n\}}\big(s^2\big)\, \mathbf{C}_{n,p/2,\nabla_\beta}\big(\mathrm{d}s^2\big) \\
+ \quad & \frac{n!\, c_{n,\beta}^{\mathscr{M}}}{\mathrm{vol}_{\beta,n}\left(\mathbb{B}_{p,\beta}^{n,\mathscr{M}}\right) C_{n,p,\nabla_\beta}\, \Gamma\Big(\frac{n+m}{p/2}+1\Big)2^n} \\
& \times \int\limits_{\mathbb{B}_{p/2}^n} h\big(s^2\big)\, \Psi_{\nabla_\beta^c}\big(s^2\big)\, \mathbf{1}_{\{x\in\mathbb{R}_+^n : x_1<\ldots<x_n\}}\big(s^2\big)\, \mathbf{U}_{n,p/2,\nabla_\beta}\big(\mathrm{d}s^2\big) \\
= n!\, \mathbf{W}(\{0\}) & \int\limits_{\mathbb{S}_{p/2}^{n-1}} h\big(s^2\big)\, \mathbf{1}_{\{x\in\mathbb{R}_+^n : x_1<\ldots<x_n\}}\big(s^2\big)\, \mathbf{C}_{n,p/2,\nabla_\beta^c}\big(\mathrm{d}s^2\big) \\
+ n! & \int\limits_{\mathbb{B}_{p/2}^n} h\big(s^2\big)\, \Psi_{\nabla_\beta^c}\big(s^2\big)\, \mathbf{1}_{\{x\in\mathbb{R}_+^n : x_1<\ldots<x_n\}}\big(s^2\big)\, \mathbf{U}_{n,p/2,\nabla_\beta^c}\big(\mathrm{d}s^2\big).
\end{aligned}
$$

Finally, when permuting the coordinates of $s^2(A)$ by a uniform random permutation $\sigma$ from the symmetric group on $n$ elements, our probabilistic representation in Theorem 3.2.3 yields

$$
\begin{aligned}
\mathbf{E}(\tilde{h}\circ\sigma)(Z) &= \mathbf{W}(\{0\}) \int\limits_{\mathbb{S}_{p/2}^{n-1}} h\big(s^2\big)\, \mathbf{C}_{n,p/2,\nabla_\beta^c}\big(\mathrm{d}s^2\big) + \int\limits_{\mathbb{B}_{p/2}^n} h\big(s^2\big)\, \Psi_{\nabla_\beta^c}\big(s^2\big)\, \mathbf{U}_{n,p/2,\nabla_\beta^c}\big(\mathrm{d}s^2\big) \\
&= \mathbf{E}h\left(\frac{X}{(\|X\|_{p/2}^{p/2}+W)^{2/p}}\right),
\end{aligned}
$$

which concludes the proof. □

**Remark 3.3.6** We now show that the degree of homogeneity of $\nabla_\beta^c$ is $m = \frac{\beta}{2}n^2 - n$ as assumed in Theorem 3.3.5: Let $s \in [0,\infty), x \in \mathbb{R}^n, \beta \in \{1,2,4\}$, then

$$\begin{aligned}
\nabla_\beta^c(sx) &= C_{\nabla_\beta} \prod_{1\le i<j\le n} |sx_i - sx_j|^\beta \prod_{i=1}^n (sx_i)^{\frac{\beta}{2}-1} \\
&= s^{\frac{\beta}{2}n-n}\, C_{\nabla_\beta} \prod_{i=1}^n \prod_{j=i}^n s^\beta |x_i - x_j|^\beta \prod_{i=1}^n x_i^{\frac{\beta}{2}-1} \\
&= s^{\frac{\beta}{2}n-n}\, C_{\nabla_\beta} \prod_{i=1}^n s^{(n-i)\beta} \prod_{j=i}^n |x_i - x_j|^\beta \prod_{i=1}^n x_i^{\frac{\beta}{2}-1} \\
&= s^{\frac{\beta}{2}n-n}\, s^{\beta\sum_{i=0}^{n-1} i}\, C_{\nabla_\beta} \prod_{i=1}^n \prod_{j=i}^n |x_i - x_j|^\beta \prod_{i=1}^n x_i^{\frac{\beta}{2}-1} \\
&= s^{\frac{\beta}{2}n^2-n}\, \nabla_\beta^c(x).
\end{aligned}$$

Hence, our assumption was indeed justified.

**Remark 3.3.7** For $\nabla_\beta^c$ the degree of homogeneity is $m = \frac{\beta}{2}n^2 - n$. Thus, if we chose $\mathbf{W} = \mathbf{G}(\alpha, 1)$ for $\alpha > 0$, analogue arguments as in Remark 3.2.8 for a random vector $X$ distributed on $\mathbb{R}_+^n$ as in Theorem 3.3.5 yield that

$$\frac{\|X\|_{p/2}^{p/2}}{\|X\|_{p/2}^{p/2} + W} \sim \mathbf{B}\left(\frac{\beta}{p}n^2, \alpha\right),$$

and for $\mathbf{W} = \vartheta\delta_0 + (1-\vartheta)\,\mathbf{G}(\alpha, 1)$, where $\vartheta \in [0,1]$ and $\alpha \in (0,\infty)$, we have by the arguments from Proposition 3.2.10 that

$$\frac{\|X\|_{p/2}^{p/2}}{\|X\|_{p/2}^{p/2} + W} \sim \vartheta\delta_1 + (1-\vartheta)\mathbf{B}\left(\frac{\beta}{p}n^2, \alpha\right).$$

## 3.4 Sanov-type LDPs for $p$-radial distributions on Euclidean $p$-balls

In [74] an LDP was derived for the empirical measure of the (suitably scaled) coordinates of a random vector that is distributed according to the cone probability measure $\mathbf{C}_{n,p}$ on $\mathbb{B}_p^n$. In this section, we prove a similar large deviation principle with the random vectors chosen according to one of the more general distributions $\mathbf{P}_{n,p,\mathbf{W}}$. We restrict ourselves to the following situation: for each $n \in \mathbb{N}$ we consider $\mathbf{W}_n := \vartheta_n\delta_0 + (1-\vartheta_n)\mathbf{G}(\alpha_n, 1)$ with $\vartheta_n \in [0,1]$ and $\alpha_n \ge 0$. This way, the distribution is specific enough to compute a concrete rate function, yet broad enough to still encapsulate many interesting distributions for the corresponding $\mathbf{P}_{n,p,\mathbf{W}_n}$, such as those outlined in Remark 2.4.3. As we will see, the large deviation behaviour of the empirical measure will be dependent both on the limits $\lim_{n\to\infty}\vartheta_n =: \vartheta \in [0,1]$ and $\lim_{n\to\infty}\alpha_n n^{-1} =: \alpha \in [0,\infty)$ of the parameter sequences and their speed of convergence, and thus will be universal to all distributions who have the same parameter limits and parameter convergence speeds. We shall appropriately write $\Psi_{f,n}$ for the $p$-radial density associated with $\mathbf{W}_n$ as defined in Theorem 3.2.3 (however, the weighting function will not be needed in this section, i.e., it can be set to $f \equiv 1$). For probability measures $\nu, \mu \in \mathcal{M}_1(\mathbb{R})$ we define the relative entropy of $\nu$ with respect to $\mu$ as

$$H(\nu\|\mu) := \begin{cases} \displaystyle\int_{\mathbb{R}} \log\frac{\nu(\mathrm{d}x)}{\mu(\mathrm{d}x)}\,\nu(\mathrm{d}x) & : \nu \ll \mu \\ +\infty & : \textit{otherwise}, \end{cases} \tag{3.19}$$

where $\nu \ll \mu$ denotes absolute continuity of $\nu$ with respect to $\mu$ and $\frac{\nu(\mathrm{d}x)}{\mu(\mathrm{d}x)}$ denotes the corresponding Radon-Nikodým derivative. For a random vector $Z^{(n)} := \big(Z_1^{(n)}, \dots, Z_n^{(n)}\big)$ in $\mathbb{R}^n$ the empirical measure of its coordinates is defined as

$$\nu_n := \frac{1}{n}\sum_{i=1}^n \delta_{Z_i^{(n)}}.$$

In the following result, the random vector $Z^{(n)}$ will have distribution $\mathbf{C}_{n,p}$ on $\mathbb{B}_p^n$, thus we will consider the empirical measure of the coordinates scaled by the factor $n^{1/p}$, i.e.,

$$\mu_n := \frac{1}{n}\sum_{i=1}^n \delta_{n^{1/p}Z_i^{(n)}}.$$

The scaling is necessary to receive non-trivial results and can be derived by the following reasoning. Since the defining condition of $\mathbb{S}_p^{n-1}$ restricts the $n$-fold sum of $p$-th powers of the coordinates of a random vector to be equal to one, it follows that the typical coordinate of that vector must be of order $n^{-1/p}$, which the rescaling counteracts (cf. [74, Proposition 2.2]). The same scaling will be applied for all other distributions on $\ell_p^n$-balls as well, as they all have $p$-radial components that are less or equal to that of the cone probability measure. We will often just call $\mu_n$ the empirical measure of a random vector $Z^{(n)}$. As mentioned in the introduction, Rachev and Rüschendorf [99] showed that the (one dimensional) marginal distributions of $\mathbf{C}_{n,p}$ asymptotically are $p$-generalized Gaussian distributions $\tilde{\mathbf{N}}_p$, thus the expectation of the $\mu_n$ is $\tilde{\mathbf{N}}_p$. In [74, Proposition 3.6] Kim and Ramanan derived the following Sanov-type LDP for the empirical measure of a random vector in $\mathbb{B}_p^n$ with distribution $\mathbf{C}_{n,p}$.

**Proposition 3.4.1** *Let $0 < p < \infty$ and let $(Z^{(n)})_{n\in\mathbb{N}}$ be a sequence of random vectors $Z^{(n)} = \big(Z_1^{(n)}, \dots, Z_n^{(n)}\big)$ in $\mathbb{B}_p^n$ with distribution $\mathbf{C}_{n,p}$. For $\mu \in \mathcal{M}_1(\mathbb{R})$ denote by $m_p(\mu)$ its $p$-th absolute moment as in* (2.5)*. Then the sequence of random probability measures $(\mu_n)_{n\in\mathbb{N}}$ with $\mu_n := \frac{1}{n}\sum_{i=1}^n \delta_{n^{1/p}Z_i^{(n)}}$ satisfies a large deviation principle on $\mathcal{M}_1(\mathbb{R})$ with speed $n$ and good rate function*

$$\mathcal{I}_{\text{cone}}(\mu) = \begin{cases} H\big(\mu \| \tilde{\mathbf{N}}_p\big) + (1 - m_p(\mu)) & : m_p(\mu) \le 1 \\ +\infty & : \textit{otherwise.} \end{cases}$$

Thus, in accordance with the theorem of Sanov, the rate function is given by the relative entropy with respect to the asymptotic coordinate distribution $\tilde{\mathbf{N}}_p$, but with an additional "penalty term" based on the $p$-th absolute moment of a measure.

**Remark 3.4.2**

(i) The original version of this result in [74, Proposition 3.6] was only formulated for $p \in [1, \infty]$, but can be expanded to $p \in (0, \infty]$, since all the probabilistic representations used in the proof also hold for $p \in (0, 1)$, and neither the convexity of $\mathbb{B}_p^n$ nor the norm-property of $\|\cdot\|_p$ was used in the proof. We exclude the case $p = \infty$ in this chapter though, hence we only present results for $p \in (0, \infty)$.

(ii) As mentioned in Remark 2.2.3, the scale of the $p$-generalized Gaussians in [74] is $p^{1/p}$ instead of 1, that is, one considers $\mathbf{N}_p$ instead of $\tilde{\mathbf{N}}_p$. Accordingly, the rate function in Proposition 3.4.1 had to be adjusted to compensate for the different parametrization in this chapter, which was chosen to keep the main results more in line with those for matrix $p$-balls from [62], which employ $\tilde{\mathbf{N}}_p$.

(iii) The above Sanov-type LDP for Euclidean $p$-balls of Kim and Ramanan [74] has been recently generalized to a Sanov-type LDP for Orlicz balls by Frühwirth and Prochno in [35]. Despite being proven differently, due to the lack of Schechtman-Zinn-type probabilistic representations, their results still exhibit a similarity to those in [74] with the rate function of the LDP being given by a relative entropy term and a generalization of the moment penalty.

We now extend Proposition 3.4.1 to random vectors with distribution $\mathbf{P}_{n,p,\mathbf{W}_n}$ on $\mathbb{B}_p^n$. It will turn out that the rate function will again be based on the relative entropy, this time perturbed by some more elaborate $p$-th moment penalty.

**Theorem 3.4.3** *Let $0 < p < \infty$ and let $(\vartheta_n)_{n\in\mathbb{N}}$ be a sequence in $[0,1]$ such that $\lim_{n\to\infty} \vartheta_n = \vartheta \in [0,1]$ and denote by $k(\vartheta) \geq 1$ the smallest number for which $\lim_{n\to\infty} n^{-k(\vartheta)} |\log(1-\vartheta_n)| < +\infty$. Also let $(\alpha_n)_{n\in\mathbb{N}}$ be a positive, real sequence such that $\lim_{n\to\infty} \alpha_n n^{-1} = \alpha \in [0,\infty)$. For each $n \in \mathbb{N}$ let $\mathbf{W}_n = \vartheta_n \delta_0 + (1-\vartheta_n) \boldsymbol{G}(\alpha_n, 1)$, and let $(Z^{(n)})_{n\in\mathbb{N}}$ be a sequence of random vectors $Z^{(n)} = \big(Z_1^{(n)}, \ldots, Z_n^{(n)}\big)$ in $\mathbb{B}_p^n$ chosen according to the distribution $\mathbf{P}_{n,p,\mathbf{W}_n}$. Then the sequence of random probability measures $(\mu_n)_{n\in\mathbb{N}}$ with $\mu_n := \frac{1}{n}\sum_{i=1}^n \delta_{n^{1/p} Z_i^{(n)}}$ satisfies a large deviation principle on $\mathcal{M}_1(\mathbb{R})$ with speed $n$ and good rate function*

$$\mathcal{I}_{\mathrm{emp}}(\mu) = \begin{cases} \mathcal{I}_{\mathrm{cone}}(\mu) - c_{(1-\vartheta)} & : \begin{array}{l} m_p(\mu) \leq 1, k(\vartheta) \geq 1, \\ \alpha = 0 \end{array} \\ \mathcal{I}_{\mathrm{cone}}(\mu) + \frac{1}{p}\log\left(\frac{1}{p}\right) - \left(\frac{1}{p} + \alpha\right)\log\left(\frac{1}{p}+\alpha\right) & : \begin{array}{l} m_p(\mu) < 1, k(\vartheta) = 1, \\ \alpha > 0 \end{array} \\ -\alpha \log\left(\frac{1-m_p(\mu)}{\alpha}\right) - c_{(1-\vartheta)} & \\ +\infty & : \textit{otherwise}, \end{cases}$$

*where $\mathcal{I}_{\mathrm{cone}}$ is the same as in Proposition 3.4.1 and*

$$c_{(1-\vartheta)} := \begin{cases} \lim\limits_{n\to\infty} n^{-1} \log(1-\vartheta_n) & : k(\vartheta) = 1 \\ 0 & : k(\vartheta) > 1. \end{cases}$$

**Remark 3.4.4**

(i) The term $c_{(1-\vartheta)}$ serves as a correction term that is only positive if $\vartheta_n$ tends to 1 in such a way that both $n^{-1}\log(1-\vartheta_n)$ and $(\alpha_n n^{-1})_{n\in\mathbb{N}}$ share the same speed of convergence. For the case $\vartheta \in [0,1)$ we always have that $k(\vartheta) = 1$ and $c_{(1-\vartheta)} = \lim_{n\to\infty} n^{-1} \log(1-\vartheta_n) = 0$, hence the rate function simplifies accordingly. For $k(\vartheta) > 1$ (which implies that $\vartheta_n$ tends to $\vartheta = 1$ faster than $\alpha_n n^{-1}$ tends to $\alpha$), the term $c_{(1-\vartheta)}$ also vanishes and the sequence $(\mu_n)_{n\in\mathbb{N}}$ from Theorem 3.4.3 based on $\mathbf{P}_{n,p,\mathbf{W}_n}$ shares its rate function with that for the cone measure $\mathbf{C}_{n,p}$ from Proposition 3.4.1. Any convergence speeds slower than $k(\vartheta) = 1$ would only yield trivial results, as the resulting LDP for the $p$-radial component of our probabilistic representation (see Lemma 3.4.5) would have a speed slower than the LDP of the directional component (see Proposition 3.4.1). However, overall we see that for many parameter sequences $(\vartheta_n)_{n\in\mathbb{N}}$, i.e., for many distributions $\mathbf{P}_{n,p,\mathbf{W}_n}$, the rate functions of the corresponding LDPs are universal.

(ii) We need to consider $k(\vartheta)$ such that $\lim_{n\to\infty} n^{-k(\vartheta)} |\log(1-\vartheta_n)| < \infty$ in order to analyze the interplay between the convex combination of measures in $\mathbf{W}_n = \vartheta_n\delta_0 + (1-\vartheta_n)\mathbf{G}(\alpha_n, 1)$ and the parameter sequence $(\alpha_n)_{n\in\mathbb{N}}$ of the involved gamma distributions. The value of $k(\vartheta)$ and the limiting behavior of $n^{-k(\vartheta)} |\log(1-\vartheta_n)|$ determine if the convex combination in $\mathbf{W}_n$ "drowns out" the involved gamma distributions $\mathbf{G}(\alpha_n, 1)$ faster than their parameter sequence $(\alpha_n)_{n\in\mathbb{N}}$ can grow and have an influence on the large deviation behavior.

The strategy of the proof of Theorem 3.4.3 will be the following: for a given random vector in $\mathbb{B}_p^n$ with distribution $\mathbf{P}_{n,p,\mathbf{W}_n}$ we apply the probabilistic representation from Proposition 2.4.5 for the specific $\mathbf{W}_n$. We split that representation into two components, one representing the directional component and the other representing the $p$-radial component of the random vector and we will derive LDPs for them separately. However, the LDP for the directional component (which has distribution $\mathbf{C}_{n,p}$) has been obtained in [74] (see Proposition 3.4.1). So, only the LDP for the $p$-radial component has to be established. Applying the contraction principle will then conclude the proof.

**Lemma 3.4.5** *Let $0 < p < \infty$ and let $(\vartheta_n)_{n\in\mathbb{N}}$ be a sequence in $[0,1]$ such that $\lim_{n\to\infty} \vartheta_n = \vartheta \in [0,1]$ and denote by $k(\vartheta) \geq 1$ the smallest number for which $\lim_{n\to\infty} n^{-k(\vartheta)} |\log(1-\vartheta_n)| < +\infty$ holds. Also let $(\alpha_n)_{n\in\mathbb{N}}$ be a positive, real sequence such that $\lim_{n\to\infty} \alpha_n n^{-1} = \alpha \in [0,\infty)$. For each $n \in \mathbb{N}$ let $X^{(n)} = \big(X_1^{(n)}, \dots, X_n^{(n)}\big)$ be a random vector with independent coordinates such that $X_i \sim \tilde{\mathbf{N}}_p$. Independently of $\big(X^{(n)}\big)_{n\in\mathbb{N}}$, let $\big(W^{(n)}\big)_{n\in\mathbb{N}}$ be a sequence of random variables with $W^{(n)} \sim \mathbf{W}_n = \vartheta_n\delta_0 + (1-\vartheta_n)\,\boldsymbol{G}(\alpha_n, 1)$. Then the sequence of random variables $\big(B^{(n)}\big)_{n\in\mathbb{N}}$ with $B^{(n)} := \big\|X^{(n)}\big\|_p^p / \big(\big\|X^{(n)}\big\|_p^p + W^{(n)}\big)$ satisfies a large deviation principle on $[0,1]$ with speed $n$ and good rate function*

$$\mathcal{I}_{\text{beta}}(x) = \begin{cases} 0 & : k(\vartheta) > 1, x = 1 \\ -\frac{1}{p}\log(x) - c_{(1-\vartheta)} & : \begin{array}{l} k(\vartheta) = 1, \alpha = 0, \\ x \in (0,1] \end{array} \\ \begin{array}{l} -\dfrac{1}{p}\log(xp) - \alpha\log\left(\dfrac{1-x}{\alpha}\right) - \left(\dfrac{1}{p}+\alpha\right)\log\left(\dfrac{1}{p}+\alpha\right) \\ -c_{(1-\vartheta)} \end{array} & : \begin{array}{l} k(\vartheta) = 1, \alpha > 0, \\ x \in (0,1) \end{array} \\ +\infty & : \textit{otherwise}, \end{cases}$$

*where*

$$c_{(1-\vartheta)} := \begin{cases} \lim_{n\to\infty} n^{-1} \log(1-\vartheta_n) & : k(\vartheta) = 1 \\ 0 & : k(\vartheta) > 1. \end{cases}$$

*Proof.* We have seen in Proposition 3.2.10 that $B^{(n)} \sim \vartheta_n \delta_1 + (1-\vartheta_n)\mathbf{B}\big(\frac{n}{p}, \alpha_n\big)$ (for $f \equiv 1$ with $m = 0$). We intend to apply the theorem of Gärtner-Ellis (Proposition 2.3.7) to show the above LDP, somewhat following along the proof of Lemma 4.1 in [4], and thus consider the following limit for $t \in \mathbb{R}$:

$$\begin{aligned} \Lambda(t) &:= \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[e^{ntB^{(n)}}\right] \\ &= \lim_{n\to\infty} \frac{1}{n} \log \left[\int_0^1 e^{ntx} \left(\vartheta_n\delta_1 + (1-\vartheta_n)\mathbf{B}\left(\frac{n}{p}, \alpha_n\right)\right)(\mathrm{d}x)\right] \\ &= \lim_{n\to\infty} \frac{1}{n} \log \left[\vartheta_n e^{nt} + (1-\vartheta_n) \int_0^1 e^{ntx}\, \mathbf{B}\left(\frac{n}{p}, \alpha_n\right)(\mathrm{d}x)\right] \\ &= t + \lim_{n\to\infty} \frac{1}{n} \log \left[\vartheta_n + (1-\vartheta_n) \int_0^1 e^{nt(x-1)}\, \mathbf{B}\left(\frac{n}{p}, \alpha_n\right)(\mathrm{d}x)\right], \end{aligned}$$

which yields

$$\Lambda(t) = t + \lim_{n\to\infty} \frac{1}{n} \log \left[\vartheta_n + (1-\vartheta_n) \frac{1}{B\left(\frac{n}{p}, \alpha_n\right)} \int_0^1 e^{nt(x-1)}\, x^{\frac{n}{p}-1}(1-x)^{\alpha_n - 1}\, \mathrm{d}x\right].$$

The change of variables $y = 1 - x$ then gives us

$$\begin{aligned} &\Lambda(t) \\ &= t + \lim_{n\to\infty} \frac{1}{n} \log \left[\vartheta_n + \frac{1-\vartheta_n}{B\left(\frac{n}{p}, \alpha_n\right)} \int_0^1 e^{-nty}\, (1-y)^{\frac{n}{p}-1} y^{\alpha_n - 1}\, \mathrm{d}y\right] \\ &= t + \lim_{n\to\infty} \frac{1}{n} \log \left[\vartheta_n + \frac{1-\vartheta_n}{B\left(\frac{n}{p}, \alpha_n\right)} \int_0^1 e^{n(-ty + \frac{n/p-1}{n}\log(1-y) + \frac{\alpha_n - 1}{n}\log(y))}\, \mathrm{d}y\right]. \qquad (3.20) \end{aligned}$$

At this point, we need to distinguish the cases $k(\vartheta) = 1$ and $k(\vartheta) > 1$, and the cases $\alpha = 0$ and $\alpha > 0$. The method of the proof will be mostly the same for those cases with $k(\vartheta) = 1$, which is to give upper and lower bounds for the integrand in the above expression, such that only the initial coefficient in the exponent remains dependent on $n$, whereby we can apply one of the asymptotic integral expansion results presented in Section 3.1.2. After some explicit calculations, we will then let our upper and lower estimates approach our initial integrand, and thereby give the sought-after limit explicitly. For $k(\vartheta) > 1$ the proof follows from rather straightforward calculations.

We begin with $k(\vartheta) = 1$ and $\alpha > 0$. For any $\epsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for $n \geq n_0$ and $y \in (0,1)$ we have that

$$e^{n\left(-ty+\left(\frac{1}{p}-\frac{1}{n}\right)\log(1-y)+\frac{\alpha_n-1}{n}\log(y)\right)} \leq e^{n\left(-ty+\left(\frac{1}{p}-\epsilon\right)\log(1-y)+(\alpha-\epsilon)\log(y)\right)} \tag{3.21}$$

and

$$e^{n\left(-ty+\left(\frac{1}{p}-\frac{1}{n}\right)\log(1-y)+\frac{\alpha_n-1}{n}\log(y)\right)} \geq e^{n\left(-ty+\left(\frac{1}{p}+\epsilon\right)\log(1-y)+(\alpha+\epsilon)\log(y)\right)}. \tag{3.22}$$

Thus, the term in (3.20) for $\alpha > 0$ is bounded from above by

$$t + \lim_{n\to\infty}\frac{1}{n}\log\left[\vartheta_n + (1-\vartheta_n)\frac{1}{B\left(\frac{n}{p},\alpha_n\right)}\int_0^1 e^{n(-ty+\left(\frac{1}{p}-\epsilon\right)\log(1-y)+(\alpha-\epsilon)\log(y))}\,\mathrm{d}y\right], \tag{3.23}$$

and bounded from below by

$$t + \lim_{n\to\infty}\frac{1}{n}\log\left[\vartheta_n + (1-\vartheta_n)\frac{1}{B\left(\frac{n}{p},\alpha_n\right)}\int_0^1 e^{n(-ty+\left(\frac{1}{p}+\epsilon\right)\log(1-y)+(\alpha+\epsilon)\log(y))}\,\mathrm{d}y\right], \tag{3.24}$$

which we denote as $\Lambda_{-\epsilon}(t)$ and $\Lambda_{+\epsilon}(t)$, respectively. We want to apply the adapted Laplace principle from Remark 3.1.3 to the terms in limits of the above expressions, and thus denote

$$\varrho_{-\epsilon,t}(y) := -ty + \left(\frac{1}{p}-\epsilon\right)\log(1-y) + (\alpha-\epsilon)\log(y),$$

and

$$\varrho_{+\epsilon,t}(y) := -ty + \left(\frac{1}{p}+\epsilon\right)\log(1-y) + (\alpha+\epsilon)\log(y).$$

We already have that $(\vartheta_n)_{n\in\mathbb{N}}$ is bounded and non-negative. Also, $(1-\vartheta_n)B\left(\frac{n}{p},\alpha_n\right)^{-1}$ is positive for all $n \in \mathbb{N}$ bigger than some $N \in \mathbb{N}$, since $k(\vartheta) = 1$ implies that $\vartheta_n \neq 1$ for $n \in \mathbb{N}$ bigger than some $N \in \mathbb{N}$.

Furthermore, both $\varrho_{-\epsilon,t}$ and $\varrho_{+\epsilon,t}$ are twice continuously differentiable on $(0,1)$. It remains to show that (3.3) holds for the sequence $\big(s^{(2)}\big)_{n\in\mathbb{N}}$ with

$$s_n^{(2)} := (1-\vartheta_n)B\left(\frac{n}{p},\alpha_n\right)^{-1},$$

and that the maximum conditions of the Laplace principle are met by $\varrho_{-\epsilon,t}$ and $\varrho_{+\epsilon,t}$. We begin with the former. It follows from $\alpha > 0$ that $\alpha_n \to +\infty$, thus Stirling's formula (2.4) tells us that, for increasing $n$, $B\big(\frac{n}{p},\alpha_n\big)$ behaves like

$$\sqrt{2\pi}\frac{\left(\frac{n}{p}\right)^{n/p\,-\,1/2}\alpha_n^{\alpha_n-1/2}}{\left(\frac{n}{p}+\alpha_n\right)^{n/p\,+\,\alpha_n\,-\,1/2}}.$$

Hence, we have

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}\log\frac{1}{B\left(\frac{n}{p},\alpha_n\right)}\\
=\;&-\lim_{n\to\infty}\Bigg[\frac{\log\sqrt{2\pi}}{n}+\frac{\frac{n}{p}-\frac{1}{2}}{n}\left(\log n+\log\frac{n/p}{n}\right)+\frac{\alpha_n-\frac{1}{2}}{n}\left(\log n+\log\frac{\alpha_n}{n}\right)\\
&\qquad-\frac{\frac{n}{p}+\alpha_n-\frac{1}{2}}{n}\left(\log n+\log\frac{\frac{n}{p}+\alpha_n}{n}\right)\Bigg]\\
=\;&-\frac{1}{p}\log\frac{1}{p}-\alpha\log\alpha+\left(\frac{1}{p}+\alpha\right)\log\left(\frac{1}{p}+\alpha\right). \qquad (3.25)
\end{aligned}$$

Thus, with $k(\vartheta)=1$ and the above, it follows for the sequence $\big(s_n^{(2)}\big)_{n\in\mathbb{N}}$, given by $s_n^{(2)}=(1-\vartheta_n)B\big(\frac{n}{p},\alpha_n\big)^{-1}$, that

$$\lim_{n\to\infty}\frac{1}{n}\log s_n^{(2)} = c_{(1-\vartheta)}-\frac{1}{p}\log\frac{1}{p}-\alpha\log\alpha+\left(\frac{1}{p}+\alpha\right)\log\left(\frac{1}{p}+\alpha\right)<+\infty. \qquad (3.26)$$

Regarding the maximum conditions of the Laplace principle, direct calculation yields that for $\epsilon<\min\big\{\alpha,\frac{1}{p}\big\}$ and $t\in\mathbb{R}\setminus\{0\}$ we have

$$\sup_{y\in(0,1)}\varrho_{-\epsilon,t}(y) \;=\; \sup_{y\in(0,1)}\left[-ty+\left(\left(\frac{1}{p}-\epsilon\right)\log(1-y)+(\alpha-\epsilon)\log(y)\right)\right]$$

$$= \frac{1}{2}\left[ -t - \left(\alpha + \frac{1}{p} - 2\epsilon\right) - \sqrt{\left(\alpha + \frac{1}{p} - 2\epsilon + t\right)^2 - 4(\alpha-\epsilon)t} \right]$$

$$+ \left(\frac{1}{p} - \epsilon\right) \log \frac{t - \left(\alpha + \frac{1}{p} - 2\epsilon\right) - \sqrt{\left(\alpha + \frac{1}{p} - 2\epsilon + t\right)^2 - 4(\alpha-\epsilon)t}}{2t}$$

$$+ (\alpha - \epsilon) \log \frac{t + \left(\alpha + \frac{1}{p} - 2\epsilon\right) + \sqrt{\left(\alpha + \frac{1}{p} - 2\epsilon + t\right)^2 - 4(\alpha-\epsilon)t}}{2t},$$

and for $t = 0$ it holds that

$$\sup_{y\in(0,1)} \varrho_{-\epsilon,0}(y) = \left(\frac{1}{p} - \epsilon\right) \log \frac{\frac{1}{p}}{\alpha + \frac{1}{p} - 2\epsilon} + (\alpha-\epsilon) \log \frac{\alpha - \epsilon}{\alpha + \frac{1}{p} - 2\epsilon}.$$

The analogue of the above holds for the maximum of $\varrho_{+\epsilon,t}$, simply replacing $-\epsilon$ by $+\epsilon$ (also, in this latter calculation the condition $\epsilon < \min\left\{\alpha, \frac{1}{p}\right\}$ is not required). By the above, it follows that the suprema of $\varrho_{-\epsilon,t}$ and $\varrho_{+\epsilon,t}$ are not attained on the boundary of the interval $[0,1]$, hence the Laplace principle can be applied to both. But before doing so, by setting

$$\Psi_{-\epsilon}(y) := -\left(\frac{1}{p} - \epsilon\right) \log(1-y) - (\alpha - \epsilon)\log(y),$$

and

$$\Psi_{+\epsilon}(y) := -\left(\frac{1}{p} + \epsilon\right) \log(1-y) - (\alpha + \epsilon)\log(y),$$

we see that

$$\sup_{y\in(0,1)} \varrho_{-\epsilon,t}(y) = \sup_{y\in(0,1)} \Big[(-t)y - \Psi_{-\epsilon}(y)\Big] = \Psi^*_{-\epsilon}(-t), \tag{3.27}$$

and

$$\sup_{y\in(0,1)} \varrho_{+\epsilon,t}(y) = \sup_{y\in(0,1)} \Big[(-t)y - \Psi_{+\epsilon}(y)\Big] = \Psi^*_{+\epsilon}(-t), \tag{3.28}$$

i.e., the suprema of $\varrho_{-\epsilon,t}$ and $\varrho_{+\epsilon,t}$ can be written as Legendre-Fenchel transforms of $\Psi_{-\epsilon}$ and $\Psi_{+\epsilon}$ at $(-t)$, respectively. Now, using the adapted Laplace principle from (3.4), and the identities from (3.26), (3.27), and (3.28), we can reformulate the respective upper and lower bounds $\Lambda_{-\epsilon}(t), \Lambda_{+\epsilon}(t)$ from (3.23) and (3.24) as

$$\Lambda_{-\epsilon}(t) = t + c_{(1-\vartheta)} - \frac{1}{p}\log\frac{1}{p} - \alpha\log\alpha + \left(\frac{1}{p} + \alpha\right)\log\left(\frac{1}{p} + \alpha\right) + \Psi^*_{-\epsilon}(-t)$$

and

$$\Lambda_{+\epsilon}(t) = t + c_{(1-\vartheta)} - \frac{1}{p}\log\frac{1}{p} - \alpha\log\alpha + \left(\frac{1}{p}+\alpha\right)\log\left(\frac{1}{p}+\alpha\right) + \Psi^*_{+\epsilon}(-t).$$

As the above holds for every sufficiently small $\epsilon > 0$, considering the limit of $\Lambda_{-\epsilon}(t)$ and $\Lambda_{+\epsilon}(t)$ as $\epsilon$ tends to 0 yields that

$$\Lambda(t) = t + c_{(1-\vartheta)} - \frac{1}{p}\log\frac{1}{p} - \alpha\log\alpha + \left(\frac{1}{p}+\alpha\right)\log\left(\frac{1}{p}+\alpha\right) + \Psi^*(-t),$$

where $\Psi^*$ is the Legendre-Fenchel transform of $\Psi$ with $\Psi(y) := -\frac{1}{p}\log(1-y) - \alpha\log(y)$, which is the limit of both $\Psi_{-\epsilon}$ and $\Psi_{+\epsilon}$ as $\epsilon$ tends to 0. Since $\Lambda$ is finite in an open neighbourhood of the origin and is lower semi-continuous and differentiable, it now follows via the theorem of Gärtner-Ellis (Proposition 2.3.7) that the sequence $\left(B^{(n)}\right)_{n\in\mathbb{N}}$ satisfies an LDP with speed $n$ and rate function $\Lambda^*$. Setting

$$c_{\vartheta,p,\alpha} := c_{(1-\vartheta)} - \frac{1}{p}\log\frac{1}{p} - \alpha\log\alpha + \left(\frac{1}{p}+\alpha\right)\log\left(\frac{1}{p}+\alpha\right)$$

for notational brevity, we get that for $x \in (0,1)$

$$\begin{aligned}
\Lambda^*(x) &= \sup_{t\in\mathbb{R}}\Big[tx - \Lambda(t)\Big] \\
&= \sup_{t\in\mathbb{R}}\Big[tx - t - \Psi^*(-t)\Big] - c_{\vartheta,p,\alpha} \\
&= \sup_{t\in\mathbb{R}}\Big[t(x-1) - \Psi^*(-t)\Big] - c_{\vartheta,p,\alpha}.
\end{aligned}$$

Again, using the change of variables $z = 1 - x$ as in (3.20), we get

$$\Lambda^*(x) = \sup_{t\in\mathbb{R}}\Big[(-t)z + \Psi^*(-t)\Big] - c_{\vartheta,p,\alpha} = \sup_{\tilde{t}\in\mathbb{R}}\Big[\tilde{t}z - \Psi^*(\tilde{t})\Big] - c_{\vartheta,p,\alpha} = (\Psi^*)^*(z) - c_{\vartheta,p,\alpha}.$$

By Lemma 2.1.1 (2), the Legendre-Fenchel transform is an involution on $(0,1)$, hence

$$\Lambda^*(x) = (\Psi^*)^*(z) - c_{\vartheta,p,\alpha} = \Psi(z) - c_{\vartheta,p,\alpha}.$$

Plugging in the definition of $\Psi$ and rolling back the previous change of variables, we have that

$$
\begin{aligned}
\Lambda^*(x) &= -\frac{1}{p}\log(1-z) - \alpha\log(z) + \frac{1}{p}\log\frac{1}{p} + \alpha\log\alpha - \left(\frac{1}{p}+\alpha\right)\log\left(\frac{1}{p}+\alpha\right) - c_{(1-\vartheta)} \\
&= -\frac{1}{p}\log(x) - \alpha\log(1-x) + \frac{1}{p}\log\frac{1}{p} + \alpha\log\alpha - \left(\frac{1}{p}+\alpha\right)\log\left(\frac{1}{p}+\alpha\right) - c_{(1-\vartheta)} \\
&= -\frac{1}{p}\log(xp) - \alpha\log\left(\frac{1-x}{\alpha}\right) - \left(\frac{1}{p}+\alpha\right)\log\left(\frac{1}{p}+\alpha\right) - c_{(1-\vartheta)},
\end{aligned}
$$

yielding the first case of the rate function. For $x \in \{0,1\}$ direct computation yields that $\Lambda^*(x) = +\infty$ in these cases.

For $k(\vartheta) = 1$ and $\alpha = 0$ we need to slightly adapt some of the steps in the proof of the previous case. We again provide upper and lower bounds for the integrand, where the lower bound will be handled completely analogously to the previous case via the adapted Laplace principle (3.4), but the upper bound needs to be approached via the asymptotic integral results from (3.5). Let $\alpha = 0$, then there exists $n_0 \in \mathbb{N}$ such that for $n \geq n_0$ and $y \in (0,1)$ we have that

$$
\begin{aligned}
e^{n\left(-ty+\left(\frac{1}{p}+\epsilon\right)\log(1-y)+\epsilon\log(y)\right)} &\leq e^{n\left(-ty+\left(\frac{1}{p}-\frac{1}{n}\right)\log(1-y)+\frac{\alpha_n-1}{n}\log(y)\right)} \\
&\leq e^{n\left(-ty+\left(\frac{1}{p}-\epsilon\right)\log(1-y)\right)}.
\end{aligned}
$$

We choose a different upper bound here than in the previous case in (3.21), since for $\alpha = 0$ and $(-1) \leq t$ the function $-ty + \left(\frac{1}{p}-\epsilon\right)\log(1-y) + (\alpha-\epsilon)\log(y)$ is strictly decreasing in $y$ and attains its maximum over $[0,1]$ on the boundary of the interval at 0. Since this is not the case for the lower bound in (3.22), we can use its analogue for $\alpha = 0$ here as well. With these bounds we get the following respective upper and lower bounds for the term in (3.20):

$$
t + \lim_{n\to\infty}\frac{1}{n}\log\left[\vartheta_n + (1-\vartheta_n)\frac{1}{B\left(\frac{n}{p},\alpha_n\right)}\int_0^1 e^{n\left(-ty+\left(\frac{1}{p}-\epsilon\right)\log(1-y)\right)}\,\mathrm{d}y\right] \tag{3.29}
$$

and

$$
t + \lim_{n\to\infty}\frac{1}{n}\log\left[\vartheta_n + (1-\vartheta_n)\frac{1}{B\left(\frac{n}{p},\alpha_n\right)}\int_0^1 e^{n\left(-ty+\left(\frac{1}{p}+\epsilon\right)\log(1-y)+\epsilon\log(y)\right)}\,\mathrm{d}y\right], \tag{3.30}
$$

again denoting these as $\Lambda_{-\epsilon}(t)$ and $\Lambda_{+\epsilon}(t)$, repectively. We again need to consider the behaviour of $\left(s^{(2)}\right)_{n\in\mathbb{N}}$ with $s_n^{(2)} := (1-\vartheta_n)B\left(\frac{n}{p},\alpha_n\right)^{-1}$ and check the conditions of the relevant asymptotic integral expansions for the functions in the respective integrands, denoted as

$$\tilde{\varrho}_{-\epsilon,t}(y) := -ty + \left(\frac{1}{p} - \epsilon\right)\log(1-y) \quad \text{and} \quad \tilde{\varrho}_{+\epsilon,t}(y) := -ty + \left(\frac{1}{p} + \epsilon\right)\log(1-y) + \epsilon\log(y).$$

If, on the one hand, both $\alpha_n \to +\infty$ and $\alpha = 0$ hold simultaneously, applying Stirling's formula as in (3.25) and interpreting the expression $0\log(0)$ as 0 yields that

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{1}{B\left(\frac{n}{p}, \alpha_n\right)} = 0. \tag{3.31}$$

If, on the other hand, $\alpha_n$ is bounded, again, by Stirling's formula (2.4), it follows that $B\left(\frac{n}{p}, \alpha_n\right)$ behaves like $\Gamma(\alpha_n)\left(\frac{n}{p}\right)^{-\alpha_n}$ for large $n \in \mathbb{N}$, which implies

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{1}{B\left(\frac{n}{p}, \alpha_n\right)} = -\lim_{n\to\infty} \frac{\log(\Gamma(\alpha_n))}{n} - \lim_{n\to\infty} \frac{\alpha_n}{n} \log\left(\frac{n}{p}\right) = 0. \tag{3.32}$$

The positivity of $s_n^{(2)}$ (at least almost everywhere) follows again by $k(\vartheta) = 1$. The function $\tilde{\varrho}_{+\epsilon,t}$ satisfies the conditions of the Laplace principle (Proposition 3.1.2) by the same arguments as in the previous case. We again set

$$\tilde{\Psi}_{-\epsilon}(y) := -\left(\frac{1}{p} - \epsilon\right)\log(1-y), \quad \text{and} \quad \tilde{\Psi}_{+\epsilon}(y) := -\left(\frac{1}{p} + \epsilon\right)\log(1-y) - \epsilon\log(y),$$

such that

$$\sup_{y\in(0,1)} \tilde{\varrho}_{-\epsilon,t}(y) = \tilde{\Psi}^*_{-\epsilon}(-t) \quad \text{and} \quad \sup_{y\in(0,1)} \tilde{\varrho}_{+\epsilon,t}(y) = \tilde{\Psi}^*_{+\epsilon}(-t), \tag{3.33}$$

as in (3.27) and (3.28). Now, using (3.31), (3.32), (3.33), and applying the adapted Laplace principle from (3.4) to the limit in $\Lambda_{+\epsilon}(t)$ in (3.30), we get

$$\Lambda_{+\epsilon}(t) = t + c_{(1-\vartheta)} + \tilde{\Psi}^*_{+\epsilon}(-t).$$

Again, we consider the limit of $\Lambda_{+\epsilon}$ as $\epsilon$ tends to zero, giving the lower bound for $\Lambda(t)$:

$$\Lambda_{+0}(t) = t + c_{(1-\vartheta)} + \tilde{\Psi}^*_{+0}(-t) = t + c_{(1-\vartheta)} + \sup_{y\in(0,1)} \left[-ty + \frac{1}{p}\log(1-y)\right]. \tag{3.34}$$

As to the upper bound, for $(-1) \le t$ the function $\tilde{\Psi}^*_{-\epsilon}$ satisfies conditions $(a) - (d)$ from Proposition 3.1.4, thus, by (3.5) from Remark 3.1.5, from (3.29) we get the upper bound for $\Lambda(t)$:

$$\Lambda_{-\epsilon}(t) = t + c_{(1-\vartheta)} + \tilde{\Psi}^*_{-\epsilon}(0) = t + c_{(1-\vartheta)} + \sup_{y\in[0,1]} \left[-\left(\frac{1}{p} - \epsilon\right)\log(1-y)\right] = t + c_{(1-\vartheta)}.$$

For $t < (-1)$ the function $\tilde{\Psi}_{-\epsilon}$ is again strictly concave and attains its supremum on $(0,1)$, so applying the adapted Laplace principle (3.4) to (3.29) yields

$$\Lambda_{-\epsilon}(t) = t + c_{(1-\vartheta)} + \tilde{\Psi}^*_{-\epsilon}(-t).$$

Combining the two cases and again considering the limit for $\epsilon$ tending to zero, we see that overall it holds for $t \in \mathbb{R}$,

$$\Lambda_{-0}(t) = t + c_{(1-\vartheta)} + \tilde{\Psi}^*_{-0}(t) = t + c_{(1-\vartheta)} + \sup_{y\in(0,1)} \left[-ty + \frac{1}{p}\log(1-y)\right],$$

which together with (3.34) yields that

$$\Lambda(t) = t + c_{(1-\vartheta)} + \sup_{y\in(0,1)} \left[-ty + \frac{1}{p}\log(1-y)\right] = t + c_{(1-\vartheta)} + \tilde{\Psi}^*(-t),$$

with $\tilde{\Psi}^*$ being the Legendre-Fenchel transform of $\tilde{\Psi}$ with $\tilde{\Psi}(y) := -\frac{1}{p}\log(1-y)$. By the theorem of Gärtner-Ellis (Proposition 2.3.7) and the same involution and change of variables arguments as in the previous case, we get that for $\alpha = 0$ the sequence $(B^{(n)})_{n\in\mathbb{N}}$ thus satisfies an LDP with speed $n$ and rate function

$$\Lambda^*(x) = -\frac{1}{p}\log(x) - c_{(1-\vartheta)}.$$

Lastly, let $k(\vartheta) > 1$. This implies on the one hand that $\vartheta = 1$ and on the other hand that $(1-\vartheta_n)$ tends to zero faster than the integral expression in (3.20) tends to infinity, i.e., the product of both tends to zero. Hence, the expression in (3.20) simplifies to

$$\begin{aligned}\Lambda(t) &= t + \lim_{n\to\infty}\frac{1}{n}\log\left[\vartheta_n + (1-\vartheta_n)\frac{1}{B\left(\frac{n}{p},\alpha_n\right)}\int_0^1 e^{n\left(-ty+\frac{n/p-1}{n}\log(1-y)+\frac{\alpha n-1}{n}\log(y)\right)}\,\mathrm{d}y\right] \\ &= t + \lim_{n\to\infty}\frac{1}{n}\log\left[\vartheta_n\right] = t,\end{aligned}$$

which implies via the theorem of Gärtner-Ellis (Proposition 2.3.7) that the sequence $(B^{(n)})_{n\in\mathbb{N}}$ satisfies an LDP on $[0,1]$ with speed $n$ and rate function

$$\mathcal{I}_{\text{beta}}(x) = \Lambda^*(x) = \sup_{t\in\mathbb{R}}\left[tx - t\right] = \sup_{t\in\mathbb{R}}\left[t(x-1)\right] = \begin{cases} 0 & : x = 1 \\ +\infty & : \textit{otherwise}. \end{cases}$$

This finishes the proof of Lemma 3.4.5. □

*Proof of Theorem 3.4.3.* By the probabilistic representation results from Proposition 2.4.5 and Lemma 3.2.1 (i), and the distributional identities from Proposition 3.2.9 and Proposition 3.2.10 (all for $f \equiv 1$), we have that

$$\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{n^{1/p} Z_i^{(n)}} \overset{\mathcal{D}}{=} \frac{1}{n}\sum_{i=1}^n \delta_{n^{1/p} \frac{X_i^{(n)}}{\left(\|X^{(n)}\|_p^p + W^{(n)}\right)^{1/p}}} = \frac{1}{n}\sum_{i=1}^n \delta_{n^{1/p} B^{(n)1/p} \frac{X_i^{(n)}}{\|X^{(n)}\|_p}},$$

where $X^{(n)}$ is a random vector with density $C_{n,p,1}\, e^{-\|x\|_p^p}$, $x \in \mathbb{R}^n$, (i.e., with distribution $\tilde{\mathbf{N}}_p^{\otimes n}$), $W^{(n)}$ is a random variable on $[0,\infty)$, independent of $X^{(n)}$, with distribution $\mathbf{W}_n = \vartheta_n \delta_0 + (1-\vartheta_n)\mathbf{G}(\alpha_n, 1)$, and $B^{(n)} := \|X^{(n)}\|_p^p / \left(\|X^{(n)}\|_p^p + W^{(n)}\right)$ as in Lemma 3.4.5. Let us define a sequence of random probability measures $(\xi_n)_{n\in\mathbb{N}}$ by

$$\xi_n := \frac{1}{n}\sum_{i=1}^n \delta_{n^{1/p} \frac{X_i^{(n)}}{\|X^{(n)}\|_p}}.$$

Then, since $X^{(n)}/\|X^{(n)}\|_p$ is independent from $\|X^{(n)}\|_p$ by Lemma 3.2.1, it follows from Proposition 2.3.4, Proposition 3.4.1, and Lemma 3.4.5 that the sequence $(\xi_n, B^{(n)})_{n\in\mathbb{N}}$ satisfies a large deviation principle on $\mathcal{M}_1(\mathbb{R}) \times [0,1]$ with speed $n$ and good rate function

$$\mathcal{I}_1(\xi, z) = \mathcal{I}_{\text{cone}}(\xi) + \mathcal{I}_{\text{beta}}(z), \qquad (\xi, z) \in \mathcal{M}_1(\mathbb{R}) \times [0,1].$$

In the case $z = 0$, we can see by Lemma 3.4.5, that $\mathcal{I}_{\text{beta}}(0) = +\infty$ and thereby $\mathcal{I}_1(\xi, 0) = \mathcal{I}_{\text{cone}}(\xi) + \infty = +\infty$ for all $\xi \in \mathcal{M}_1(\mathbb{R})$. Thus, we confine ourselves to $z \in (0,1]$. Next, we introduce the continuous map $F_p : \mathcal{M}_1(\mathbb{R}) \times (0,1] \to \mathcal{M}_1(\mathbb{R})$ with $(\xi, z) \mapsto \xi(z^{-1/p}\,\cdot\,)$ and notice that for each $n \in \mathbb{N}$ and for any Borel set $A \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned} F_p\left(\xi_n, B^{(n)}\right)(A) &= F_p\left(\frac{1}{n}\sum_{i=1}^n \delta_{n^{1/p} \frac{X_i^{(n)}}{\|X^{(n)}\|_p}}, B^{(n)}\right)(A) \\ &= \frac{1}{n}\sum_{i=1}^n \delta_{n^{1/p} \frac{X_i^{(n)}}{\|X^{(n)}\|_p}}\left(B^{(n)-1/p} A\right) \\ &= \frac{1}{n}\sum_{i=1}^n \delta_{n^{1/p} B^{(n)1/p} \frac{X_i^{(n)}}{\|X^{(n)}\|_p}}(A) \\ &= \mu_n(A). \end{aligned}$$

By the contraction principle in Proposition 2.3.5, the sequence of random probability measures $(\mu_n)_{n\in\mathbb{N}}$ thus satisfies a large deviation principle with speed $n$ and good rate function $\mathcal{I}_2 : \mathcal{M}_1(\mathbb{R}) \times (0,1] \to [0,\infty]$ given by

$$\mathcal{I}_2(\mu) = \inf_{\xi(z^{-1/p}\,\cdot\,) = \mu(\cdot)} \left[\mathcal{I}_{\text{cone}}(\xi) + \mathcal{I}_{\text{beta}}(z)\right], \qquad \mu \in \mathcal{M}_1(\mathbb{R}), z \in (0,1].$$

It remains to show that $\mathcal{I}_2$ in fact coincides with the rate function $\mathcal{I}_{\mathrm{emp}}$ stated in the theorem. The rate functions $\mathcal{I}_{\mathrm{cone}}$ and $\mathcal{I}_{\mathrm{beta}}$ each depend on their respective parameters $m_p(\mu) \in [0, \infty]$, $k(\vartheta) \geq 1$, and $\alpha \in [0, \infty)$, so we need to check for which parameter configurations they remain finite.

***Case 1.*** Let $\mu \in \mathcal{M}_1(\mathbb{R})$ be such that $m_p(\mu) > 1$. Then, by $\xi(z^{-1/p}\,\cdot\,) = \mu(\,\cdot\,)$, we know that $m_p(\xi) = z^{-1}\, m_p(\mu)$, so $m_p(\xi) > 1$. Therefore $\mathcal{I}_{\mathrm{cone}}(\xi) = +\infty$ and $\mathcal{I}_{\mathrm{emp}}(\mu) = \mathcal{I}_2(\mu) = +\infty$.

***Case 2.*** Let $\mu \in \mathcal{M}_1(\mathbb{R})$ be such that $m_p(\mu) \leq 1$ and $\mathbf{W}_n$ be such that $\alpha = 0$. By $\xi(z^{-1/p}\,\cdot\,) = \mu(\,\cdot\,)$, we again know that $m_p(\xi) = z^{-1}\, m_p(\mu)$. Now we have to distinguish between the cases $m_p(\xi) > 1$ and $m_p(\xi) \leq 1$. In the first case, $\mathcal{I}_{\mathrm{cone}}(\xi) = +\infty$ and therefore $\mathcal{I}_{\mathrm{emp}}(\mu) = \mathcal{I}_2(\mu) = +\infty$. If $m_p(\xi) \leq 1$, then $z$ is restricted to the non-empty interval $[m_p(\mu), 1]$. Hence, $z \in [m_p(\mu), 1] \cap (0, 1]$. If $k(\vartheta) > 1$, we know by Lemma 3.4.5 that $\mathcal{I}_{\mathrm{beta}}(z)$ is only finite for $z = 1$, in which case it follows that $\xi = \mu$ and $\mathcal{I}_2(\mu) = \mathcal{I}_{\mathrm{cone}}(\mu) = \mathcal{I}_{\mathrm{cone}}(\mu) - c_{(1-\vartheta)}$. If $k(\vartheta) = 1$, by Proposition 3.4.1 and Lemma 3.4.5, we get

$$
\begin{aligned}
\mathcal{I}_2(\mu) &= \inf_{\xi(z^{-1/p}\,\cdot\,) = \mu(\cdot)} \left[ H(\xi \| \tilde{\mathbf{N}}_p) + (1 - m_p(\xi)) - \frac{1}{p} \log(z) - c_{(1-\vartheta)} \right] \\
&= \inf_{\xi(z^{-1/p}\,\cdot\,) = \mu(\cdot)} \left[ \int_{\mathbb{R}} \log \frac{\xi(\mathrm{d}x)}{\tilde{\mathbf{N}}_p(\mathrm{d}x)}\, \xi(\mathrm{d}x) + (1 - z^{-1} m_p(\mu)) - \frac{1}{p} \log(z) \right] - c_{(1-\vartheta)}.
\end{aligned}
$$

The change of variables $y = z^{1/p} x$ then gives us $\xi(\mathrm{d}x) = \xi(\mathrm{d}z^{-1/p} y) = \mu(\mathrm{d}y)$, and

$$
\tilde{\mathbf{N}}_p(\mathrm{d}x) = \tilde{\mathbf{N}}_p\left(\mathrm{d}z^{-1/p} y\right) = \left(2 z^{1/p} \Gamma(1 + 1/p)\right)^{-1} e^{-z^{-1}|y|^p} \mathrm{d}y =: \tilde{\mathbf{N}}_{p,z}(\mathrm{d}y).
$$

Thus,

$$
\mathcal{I}_2(\mu) = \inf_{z \in [m_p(\mu), 1] \cap (0, 1]} \left[ \int_{\mathbb{R}} \log \frac{\mu(\mathrm{d}y)}{\tilde{\mathbf{N}}_{p,z}(\mathrm{d}y)}\, \mu(\mathrm{d}y) + \left(1 - z^{-1} m_p(\mu)\right) - \frac{1}{p} \log(z) \right] - c_{(1-\vartheta)},
$$

which is only dependent on $z \in [m_p(\mu), 1] \cap (0, 1]$. We further compute

$$
\begin{aligned}
\int_{\mathbb{R}} \log \frac{\mu(\mathrm{d}y)}{\tilde{\mathbf{N}}_{p,z}(\mathrm{d}y)}\, \mu(\mathrm{d}y) &= \int_{\mathbb{R}} \log \left( \frac{\mu(\mathrm{d}y)}{\tilde{\mathbf{N}}_p(\mathrm{d}y)} \frac{\tilde{\mathbf{N}}_p(\mathrm{d}y)}{\tilde{\mathbf{N}}_{p,z}(\mathrm{d}y)} \right) \mu(\mathrm{d}y) \\
&= \int_{\mathbb{R}} \log \frac{\mu(\mathrm{d}y)}{\tilde{\mathbf{N}}_p(\mathrm{d}y)}\, \mu(\mathrm{d}y) + \int_{\mathbb{R}} \log \frac{\tilde{\mathbf{N}}_p(\mathrm{d}y)}{\tilde{\mathbf{N}}_{p,z}(\mathrm{d}y)}\, \mu(\mathrm{d}y)
\end{aligned}
$$

$$= H(\mu\|\tilde{\mathbf{N}}_p) + \int_{\mathbb{R}} \log \frac{\tilde{\mathbf{N}}_p(\mathrm{d}y)}{\tilde{\mathbf{N}}_{p,z}(\mathrm{d}y)}\, \mu(\mathrm{d}y).$$

Since

$$\frac{\tilde{\mathbf{N}}_p(\mathrm{d}y)}{\tilde{\mathbf{N}}_{p,z}(\mathrm{d}y)} = e^{(z^{-1}-1)|y|^p} z^{1/p},$$

we conclude that

$$\begin{aligned}
\int_{\mathbb{R}} \log \frac{\mu(\mathrm{d}y)}{\tilde{\mathbf{N}}_{p,z}(\mathrm{d}y)} \mu(\mathrm{d}y) &= H(\mu\|\tilde{\mathbf{N}}_p) + \int_{\mathbb{R}} \log \left( e^{(z^{-1}-1)|y|^p} z^{1/p} \right) \mu(\mathrm{d}y) \\
&= H(\mu\|\tilde{\mathbf{N}}_p) + (z^{-1}-1) \int_{\mathbb{R}} |y|^p\, \mu(\mathrm{d}y) + \frac{1}{p} \log(z) \int_{\mathbb{R}} \mu(\mathrm{d}y) \\
&= H(\mu\|\tilde{\mathbf{N}}_p) + (z^{-1}-1) m_p(\mu) + \frac{1}{p} \log(z), \qquad (3.35)
\end{aligned}$$

which minimizes for $z = 1$. Hence, the rate function is of the form

$$\mathcal{I}_2(\mu) = H(\mu\|\tilde{\mathbf{N}}_p) + (1 - m_p(\mu)) - c_{(1-\vartheta)} = \mathcal{I}_{\mathrm{cone}}(\mu) - c_{(1-\vartheta)}.$$

***Case 3.*** Let $\mu \in \mathcal{M}_1(\mathbb{R})$ be such that $m_p(\mu) \leq 1$ and $\mathbf{W}_n$ be such that $\alpha > 0$. By the same arguments as above, we assume that $m_p(\xi) \leq 1$ and $z \in [m_p(\mu), 1] \cap (0, 1)$, where we exclude $z = 1$ due to Lemma 3.4.5. Then, by Proposition 3.4.1 and Lemma 3.4.5, we get

$$\begin{aligned}
\mathcal{I}_2(\mu) &= \inf_{\xi(z^{-1/p}\,\cdot\,)=\mu(\cdot)} \Bigg[ H(\xi\|\tilde{\mathbf{N}}_p) + (1 - m_p(\xi)) - \frac{1}{p}\log(pz) - \alpha \log \frac{1-z}{\alpha} \\
&\qquad - \left(\frac{1}{p} + \alpha\right) \log\left(\frac{1}{p} + \alpha\right) - c_{(1-\vartheta)} \Bigg] \\
&= \inf_{\xi(z^{-1/p}\,\cdot\,)=\mu(\cdot)} \left[ \int_{\mathbb{R}} \log \frac{\xi(\mathrm{d}x)}{\tilde{\mathbf{N}}_p(\mathrm{d}x)} \xi(\mathrm{d}x) + (1 - z^{-1} m_p(\mu)) - \frac{1}{p}\log(pz) - \alpha \log \frac{1-z}{\alpha} \right] \\
&\qquad - \left(\frac{1}{p} + \alpha\right) \log\left(\frac{1}{p} + \alpha\right) - c_{(1-\vartheta)}.
\end{aligned}$$

The change of variables $y = z^{1/p} x$ as in Case 2 then lets us reformulate the above to

$$\begin{aligned}
\inf_{z\in[m_p(\mu),1]\cap(0,1)} &\left[ \int_{\mathbb{R}} \log \frac{\mu(\mathrm{d}y)}{\tilde{\mathbf{N}}_{p,z}(\mathrm{d}y)} \mu(\mathrm{d}y) + (1 - z^{-1} m_p(\mu)) - \frac{1}{p}\log(pz) - \alpha \log \frac{1-z}{\alpha} \right] \\
&- \left(\frac{1}{p} + \alpha\right) \log\left(\frac{1}{p} + \alpha\right) - c_{(1-\vartheta)}.
\end{aligned}$$

Using the argument from (3.35) it follows that

$$\begin{aligned}
\mathcal{I}_2(\mu) = \inf_{z\in[m_p(\mu),1]\cap(0,1)} & \Bigg[ H(\mu\|\tilde{\mathbf{N}}_p) + (z^{-1}-1)m_p(\mu) + \frac{1}{p}\log(z) + (1-z^{-1}m_p(\mu)) \\
& \qquad - \frac{1}{p}\log(pz) - \alpha\log\frac{1-z}{\alpha}\Bigg] - \left(\frac{1}{p}+\alpha\right)\log\left(\frac{1}{p}+\alpha\right) - c_{(1-\vartheta)} \\
= {} & H(\mu\|\tilde{\mathbf{N}}_p) + (1-m_p(\mu)) - \frac{1}{p}\log(p) + \alpha\log(\alpha) - \left(\frac{1}{p}+\alpha\right)\log\left(\frac{1}{p}+\alpha\right) \\
& - c_{(1-\vartheta)} + \inf_{z\in[m_p(\mu),1]\cap(0,1)}\Big[-\alpha\log(1-z)\Big] \\
= {} & \mathcal{I}_{\text{cone}}(\mu) + \frac{1}{p}\log\left(\frac{1}{p}\right) - \left(\frac{1}{p}+\alpha\right)\log\left(\frac{1}{p}+\alpha\right) - \alpha\log\left(\frac{1-m_p(\mu)}{\alpha}\right) - c_{(1-\vartheta)},
\end{aligned}$$

where the last equality only holds for $m_p(\mu) < 1$, since for $m_p(\mu) = 1$, we have $z \in [m_p(\mu),1] \cap (0,1) = \emptyset$, and thus $\mathcal{I}_2(\mu) = +\infty$. Hence, $m_p(\mu) = 1$ can only be permitted if $\alpha = 0$.

Thus, we have shown that $\mathcal{I}_2$ in fact coincides with $\mathcal{I}_{\text{emp}}$ as given in Theorem 3.4.3, finishing its proof. □

**Example 3.4.6** If $\vartheta = 0$, we get the large deviation behaviour of the empirical measure of a random vector distributed according to some beta-type distribution $\Psi_{f,n}\mathbf{U}_{n,p,f}$ as discussed in Example 3.2.7 for $f \equiv 1$. Note that this could be any distribution $\vartheta_n\mathbf{C}_{n,p} + (1-\vartheta_n)\Psi_{f,n}\mathbf{U}_{n,p,f}$ with $\vartheta_n \to \vartheta = 0$. Since $\vartheta$ only influences the rate function via $c_{(1-\vartheta)}$ and $c_{(1-\vartheta)} = 0$ for $\vartheta \in [0,1)$, any distribution $\vartheta_n\mathbf{C}_{n,p} + (1-\vartheta_n)\Psi_{f,n}\mathbf{U}_{n,p,f}$ with $\vartheta_n \to \vartheta \in [0,1)$ exhibits the same large deviation behaviour, i.e., shares the same universal rate function for the sequence $(\mu_n)_{n\in\mathbb{N}}$ of corresponding empirical measures

$$\mathcal{I}_{\text{emp}}(\mu) = \begin{cases} \mathcal{I}_{\text{cone}}(\mu) & : m_p(\mu) \le 1,\ \alpha = 0 \\[2ex] \mathcal{I}_{\text{cone}}(\mu) + \frac{1}{p}\log\frac{1}{p} - \left(\frac{1}{p}+\alpha\right)\log\left(\frac{1}{p}+\alpha\right) & : m_p(\mu) < 1,\ \alpha > 0 \\ -\alpha\log\left(\frac{1-m_p(\mu)}{\alpha}\right) & \\[2ex] +\infty & : \textit{otherwise}. \end{cases}$$

## 3.5 Sanov-type LDPs for $p$-radial distributions on matrix $p$-balls

In this section, we want to use the tools from the previous sections to analyze the large deviation behaviours of random matrices in $\mathbb{B}_{p,\beta}^{n,\mathscr{H}}$ and $\mathbb{B}_{p,\beta}^{n,\mathscr{M}}$ distributed according to $\mathbf{P}_{n,p,\mathbf{W},\beta}^{\mathscr{H}}$ and $\mathbf{P}_{n,p,\mathbf{W},\beta}^{\mathscr{M}}$, respectively. We will use the probabilistic representations from Theorem 3.3.1 and Theorem 3.3.5 regarding the eigenvalue and singular value distributions together with further large deviation results for their $p$-radial component in the spirit of Lemma 3.4.5 to derive LDPs for $\mathbb{B}_{p,\beta}^{n,\mathscr{H}}$ and $\mathbb{B}_{p,\beta}^{n,\mathscr{M}}$.

### 3.5.1 Self-adjoint matrix $p$-balls

In the case of the matrix $p$-balls our goal is to derive an LDP for the so-called empirical spectral measure of a random matrix in $\mathbb{B}_{p,\beta}^{n,\mathscr{H}}$. Let $Z^{(n)} \in \mathscr{H}_n(\mathbb{F}_\beta)$ be a self-adjoint random matrix with eigenvalues $\lambda_1\big(Z^{(n)}\big) \le \ldots \le \lambda_n\big(Z^{(n)}\big)$. We then define the empirical spectral measure as the random measure

$$\nu_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(Z^{(n)})},$$

i.e., the empirical measure with respect to the eigenvalues. We will again consider the suitably scaled version

$$\mu_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{n^{1/p}\lambda_i(Z^{(n)})}$$

and refer to it as the empirical spectral measure of the random matrix $Z^{(n)}$. In [63] a large deviation principle for the empirical spectral measure of random matrices chosen according to either $\mathbf{U}_{n,p,\beta}^{\mathscr{H}}$ or $\mathbf{C}_{n,p,\beta}^{\mathscr{H}}$ was proven. In this section, we generalize this result by proving a large deviation principle for random matrices chosen according to one of the more general distributions $\mathbf{P}_{n,p,\mathbf{W},\beta}^{\mathscr{H}} := \mathbf{W}(\{0\})\mathbf{C}_{n,p,\beta}^{\mathscr{H}} + \Psi^{\mathscr{H}}\mathbf{U}_{n,p,\beta}^{\mathscr{H}}$ on $\mathbb{B}_{p,\beta}^{n,\mathscr{H}}$ introduced in Section 3.3. We again consider distributions $\mathbf{W}_n := \vartheta_n\delta_0 + (1-\vartheta_n)\mathbf{G}(\alpha_n, 1)$ with weight sequence $(\vartheta_n)_{n\in\mathbb{N}}$ in $[0,1]$ and parameter sequence $(\alpha_n)_{n\in\mathbb{N}}$ in $[0,\infty)$, and thus write $\mathbf{P}_{n,p,\mathbf{W}_n,\beta}^{\mathscr{H}}$ and $\Psi_n^{\mathscr{H}}$ (and $\Psi_{f,n}$ in the Euclidean representation).

**Theorem 3.5.1** *Let $0 < p < \infty$, $\beta \in \{1,2,4\}$, and let $(\vartheta_n)_{n\in\mathbb{N}}$ be a sequence in $[0,1]$ with $\lim_{n\to\infty}\vartheta_n = \vartheta \in [0,1]$ and denote by $k(\vartheta) \ge 2$ the smallest number such that $\lim_{n\to\infty} n^{-k(\vartheta)}\,|\log(1-\vartheta_n)| < +\infty$. Further, let $(\alpha_n)_{n\in\mathbb{N}}$ be a positive, real sequence such that $\lim_{n\to\infty}\alpha_n n^{-2} = \alpha \in [0,\infty)$. For each $n \in \mathbb{N}$ let $\mathbf{W}_n = \vartheta_n\delta_0 + (1-\vartheta_n)\boldsymbol{G}(\alpha_n, 1)$, and let $Z^{(n)}$ be a random matrix in $\mathbb{B}_{p,\beta}^{n,\mathscr{H}}$ chosen according to the distribution $\mathbf{P}_{n,p,\mathbf{W}_n,\beta}^{\mathscr{H}}$.*

*Then the sequence of random probability measures* $\mu_n := \frac{1}{n}\sum_{i=1}^n \delta_{n^{1/p}\lambda_i(Z^{(n)})}$ *satisfies a large deviation principle on* $\mathcal{M}_1(\mathbb{R})$ *with speed* $n^2$ *and good rate function*

$$\mathcal{I}_{\mathrm{emp}}^{\mathscr{H}}(\mu) = \begin{cases} \mathcal{I}_{\mathrm{cone}}^{\mathscr{H}}(\mu) - c_{(1-\vartheta)}^{\mathscr{H}} & : \begin{array}{l} m_p(\mu) \leq 1, \\ k(\vartheta) \geq 2, \\ \alpha = 0 \end{array} \\ \mathcal{I}_{\mathrm{cone}}^{\mathscr{H}}(\mu) + \dfrac{\beta}{2p}\log\left(\dfrac{\beta}{2p}\right) - \left(\dfrac{\beta}{2p}+\alpha\right)\log\left(\dfrac{\beta}{2p}+\alpha\right) & : \begin{array}{l} m_p(\mu) < 1, \\ k(\vartheta) = 2, \\ \alpha > 0 \end{array} \\ \quad -\alpha\log\left(\frac{1-m_p(\mu)}{\alpha}\right) - c_{(1-\vartheta)}^{\mathscr{H}} & \\ +\infty & : \textit{otherwise}, \end{cases}$$

*where*

$$\mathcal{I}_{\mathrm{cone}}^{\mathscr{H}}(\mu) = \begin{cases} -\dfrac{\beta}{2}\displaystyle\int_{\mathbb{R}}\int_{\mathbb{R}} \log|x-y|\,\mu(\mathrm{d}x)\mu(\mathrm{d}y) + \dfrac{\beta}{2p}\log\left(\dfrac{\sqrt{\pi}\,p\,\Gamma(\frac{p}{2})}{2^p\sqrt{e}\,\Gamma\left(\frac{p+1}{2}\right)}\right) & : m_p(\mu) \leq 1 \\ +\infty & : \textit{otherwise}, \end{cases}$$

*and*

$$c_{(1-\vartheta)}^{\mathscr{H}} := \begin{cases} \lim\limits_{n\to\infty} n^{-2}\log(1-\vartheta_n) & : k(\vartheta) = 2 \\ 0 & : k(\vartheta) > 2. \end{cases}$$

The proof of this result is rather similar to that of Theorem 3.4.3, with the main difference that it uses the probabilistic representation from Theorem 3.3.1, which is weighted by the repulsion factor $\Delta_\beta$ of the eigenvalues ($\Delta_\beta^c$ with normalizing constants). We again split that probabilistic representation into two components, one directional component with distribution $\mathbf{C}_{n,p,\beta}^{\mathscr{H}}$ on the matrix $p$-ball and the other reflecting the $p$-radial component. The main difference will be that the degree of homogeneity $m$ of the weight function $f$ is non-zero if $f = \Delta_\beta^c$, but $m = \frac{\beta n(n-1)}{2}$. Therefore, as outlined in Remark 3.3.4, the first parameter of the beta distribution involved in the distribution of the $p$-radial component (compare with Lemma 3.4.5) will have a different limit behaviour, affecting both the speed (via the order of convergence) and the rate function (via the limit).

We now present two results outlining the large deviation behaviour of the aforementioned two components of the probabilistic representation of a random matrix with distribution $\mathbf{P}_{n,p,\mathbf{W}_n,\beta}^{\mathscr{H}}$. One will do so for the empirical spectral measure of random matrices with distribution $\mathbf{C}_{n,p,\beta}^{\mathscr{H}}$ on $\mathbb{B}_{p,\beta}^{n,\mathscr{H}}$ and the other for the $p$-radial component of the probabilistic representation. We start with the latter.

**Lemma 3.5.2** *Let $0 < p < \infty$, $\beta \in \{1, 2, 4\}$, and let $(\vartheta_n)_{n\in\mathbb{N}}$ be a sequence in $[0, 1]$ with $\lim_{n\to\infty} \vartheta_n = \vartheta \in [0, 1]$ and denote by $k(\vartheta) \geq 2$ the smallest number such that $\lim_{n\to\infty} n^{-k(\vartheta)} |\log(1 - \vartheta_n)| < +\infty$. Also let $(\alpha_n)_{n\in\mathbb{N}}$ be a positive, real sequence such that $\lim_{n\to\infty} \alpha_n n^{-2} = \alpha \in [0, \infty)$. For each $n \in \mathbb{N}$ let $X^{(n)} = \big(X_1^{(n)}, \ldots, X_n^{(n)}\big)$ be a random vector with density $C_{n,p,\Delta_\beta^c}\, e^{-\|x\|_p^p}\, \Delta_\beta^c(x)$, $x \in \mathbb{R}^n$, with $\Delta_\beta^c$ defined as in Theorem 3.3.1. Independently of the sequence $\big(X^{(n)}\big)_{n\in\mathbb{N}}$, let $\big(W^{(n)}\big)_{n\in\mathbb{N}}$ be a sequence of random variables with $W^{(n)} \sim \mathbf{W}_n = \vartheta_n\delta_0 + (1 - \vartheta_n)\boldsymbol{G}(\alpha_n, 1)$. Then the sequence of random variables $\big(B^{(n)}\big)_{n\in\mathbb{N}}$ with $B^{(n)} := \big\|X^{(n)}\big\|_p^p / \big(\big\|X^{(n)}\big\|_p^p + W^{(n)}\big)$ satisfies a large deviation principle on $[0, \infty)$ with speed $n^2$ and good rate function*

$$\mathcal{I}_{\text{beta}}^{\mathscr{H}}(x) = \begin{cases} 0 & : k(\vartheta) > 2, x = 1 \\ -\frac{\beta}{2p}\log(x) - c_{(1-\vartheta)}^{\mathscr{H}}(x) & : \begin{array}{l} k(\vartheta) = 2, \alpha = 0, \\ x \in (0, 1] \end{array} \\ \begin{array}{l} -\frac{\beta}{2p}\log\left(\frac{2xp}{\beta}\right) - \alpha\log\left(\frac{1-x}{\alpha}\right) - \left(\frac{\beta}{2p} + \alpha\right)\log\left(\frac{\beta}{2p} + \alpha\right) \\ - c_{(1-\vartheta)}^{\mathscr{H}}(x) \end{array} & : \begin{array}{l} k(\vartheta) = 2, \alpha > 0, \\ x \in (0, 1) \end{array} \\ +\infty & : \textit{otherwise}, \end{cases}$$

*where*

$$c_{(1-\vartheta)}^{\mathscr{H}} := \begin{cases} \lim\limits_{n\to\infty} n^{-2}\log(1 - \vartheta_n) & : k(\vartheta) = 2 \\ 0 & : k(\vartheta) > 2. \end{cases}$$

This is proven in the same way as Lemma 3.4.5 with only a few differences. Since we are dealing with matrix $p$-balls here, the weight function is $\Delta_\beta^c$, which is homogeneous of degree $m = \frac{1}{2}\beta n(n - 1)$. We use the probabilistic representation of the $\ell_p^n$-norm of the eigenvalue-vector via the distributional convex combination given in Remark 3.3.4. We have seen in the proof of Lemma 3.4.5 that the LDP of the latter is heavily dependent on the limits and orders of convergence of the involved parameter sequences. It holds for the first parameter of the involved beta distribution from Remark 3.3.4 that $\lim_{n\to\infty} \frac{n+m}{p} n^{-2} = \frac{\beta}{2p}$. This explains the appearance of $n^2$ instead of $n$ for the speed and the factor $\frac{\beta}{2p}$ instead of $\frac{1}{p}$ in the rate function.

The second lemma is a large deviation principle for the sequence of empirical spectral measures of a random matrix in $\mathbb{B}_{p,\beta}^{n,\mathscr{H}}$ with distribution $\mathbf{C}_{n,p,\beta}^{\mathscr{H}}$ from [62, Theorem 1.1].

**Lemma 3.5.3** *Let $0 < p < \infty$, $\beta \in \{1, 2, 4\}$ and $n \in \mathbb{N}$. Further, let $Z^{(n)}$ be a random matrix in $\mathbb{B}_{p,\beta}^{n,\mathscr{H}}$ with distribution $\mathbf{C}_{n,p,\beta}^{\mathscr{H}}$ and eigenvalues $\lambda_i(Z^{(n)})$, $i \in \{1, \dots, n\}$. Then the sequence of random probability measures $\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{n^{1/p}\lambda_i(Z^{(n)})}$ satisfies a large deviation principle on $\mathcal{M}_1(\mathbb{R})$ with speed $n^2$ and good rate function*

$$\mathcal{I}_{\mathrm{cone}}^{\mathscr{H}}(\mu) = \begin{cases} -\frac{\beta}{2} \int\limits_{\mathbb{R}} \int\limits_{\mathbb{R}} \log|x-y| \, \mu(\mathrm{d}x)\mu(\mathrm{d}y) + \frac{\beta}{2p} \log\left( \frac{\sqrt{\pi}\, p\, \Gamma\left(\frac{p}{2}\right)}{2^p \sqrt{e}\, \Gamma\left(\frac{p+1}{2}\right)} \right) & : m_p(\mu) \leq 1 \\ +\infty & : \textit{otherwise.} \end{cases}$$

*Proof of Theorem 3.5.1.* Since this proof is again quite similar to that of Theorem 3.4.3, we reduce it to the essential differences. We use the probabilistic representations from Theorem 3.3.1 and Lemma 3.2.1 (i), and the distributional identities from Proposition 3.2.9 and Proposition 3.2.10 to get

$$\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{n^{1/p}\lambda_i(Z^{(n)})} \overset{\mathcal{D}}{=} \frac{1}{n} \sum_{i=1}^n \delta_{n^{1/p} \frac{X_i^{(n)}}{\left(\|X^{(n)}\|_p^p + W^{(n)}\right)^{1/p}}} = \frac{1}{n} \sum_{i=1}^n \delta_{n^{1/p} B^{(n)\,1/p} \frac{X_i^{(n)}}{\|X^{(n)}\|_p}},$$

where $X^{(n)}$ is a random vector with density $C_{n,p,\Delta_\beta^c}\, e^{-\|x\|_p^p}\, \Delta_\beta^c(x)$, $x \in \mathbb{R}^n$, $W^{(n)}$ a random variable on $[0, \infty)$ with distribution $\mathbf{W}_n = \vartheta_n \delta_0 + (1 - \vartheta_n)\mathbf{G}(\alpha_n, 1)$, and with $B^{(n)} := \|X^{(n)}\|_p^p / \left(\|X^{(n)}\|_p^p + W^{(n)}\right)$. Note, that while Theorem 3.3.1 makes a distributional statement for the randomly permuted eigenvalue vector $\lambda_\sigma(Z)$, the above statement holds for the empirical measure of the ordered eigenvalue vector $\lambda(Z)$ as well, since we are considering the Dirac measures of its coordinates within a sum, in which the order of the summands is irrelevant. Using Lemma 3.5.2 and Lemma 3.5.3, by the same arguments as in the proof of Theorem 3.4.3, we get that $(\mu_n)_{n\in\mathbb{N}}$ satisfies an LDP with speed $n^2$ and good rate function $\mathcal{I}_2^{\mathscr{H}} : \mathcal{M}_1(\mathbb{R}) \times (0, 1] \to [0, \infty)$ given by

$$\mathcal{I}_2^{\mathscr{H}}(\mu) = \inf_{\xi(z^{-1/p}\,\cdot) = \mu(\cdot)} \Big[ \mathcal{I}_{\mathrm{cone}}^{\mathscr{H}}(\xi) + \mathcal{I}_{\mathrm{beta}}^{\mathscr{H}}(z) \Big], \qquad \mu \in \mathcal{M}_1(\mathbb{R}), z \in (0, 1].$$

It remains to show that $\mathcal{I}_2^{\mathscr{H}}$ is just the rate function $\mathcal{I}_{\mathrm{emp}}^{\mathscr{H}}$ stated in the theorem. However, this is done analogously to the Euclidean setting by a case-by-case analysis of parameter configurations $m_p(\mu) \in [0, \infty]$, $k(\vartheta) \geq 2$ and $\alpha \in [0, \infty)$, such that the rate functions $\mathcal{I}_{\mathrm{cone}}^{\mathscr{H}}$ and $\mathcal{I}_{\mathrm{beta}}^{\mathscr{H}}$ remain finite, hence we omit the details. □

**Example 3.5.4** Similar to Example 3.4.6, it follows that empirical spectral measure of a random matrix $Z^{(n)}$ of any distribution $\vartheta_n \mathbf{C}_{n,p,\beta}^{\mathscr{H}} + (1 - \vartheta_n)\Psi_n^{\mathscr{H}} \mathbf{U}_{n,p,\beta}^{\mathscr{H}}$ with $\vartheta_n \to \vartheta \in [0, 1)$ exhibits large deviation behaviour described by the rate function

$$\mathcal{I}_{\mathrm{emp}}^{\mathscr{H}}(\mu) = \begin{cases} \mathcal{I}_{\mathrm{cone}}^{\mathscr{H}}(\mu) & : \begin{array}{l} m_p(\mu) \le 1, \\ \alpha = 0 \end{array} \\ \mathcal{I}_{\mathrm{cone}}^{\mathscr{H}}(\mu) + \frac{\beta}{2p}\log\left(\frac{\beta}{2p}\right) - \left(\frac{\beta}{2p}+\alpha\right)\log\left(\frac{\beta}{2p}+\alpha\right) & : \begin{array}{l} m_p(\mu) < 1, \\ \alpha \in (0,\infty) \end{array} \\ -\alpha\log\left(\frac{1-m_p(\mu)}{\alpha}\right) & \\ +\infty & : \textit{otherwise}, \end{cases}$$

hence the rate function and the corresponding large deviation behaviour is universal for all these distributions.

### 3.5.2 Non-self-adjoint matrix $p$-balls

If the matrix is not self-adjoint, we define the empirical spectral measure of a matrix $Z^{(n)} \in \mathscr{M}_n(\mathbb{F}_\beta)$ with respect to the squared singular values $s_1^2\big(Z^{(n)}\big) \le \ldots \le s_n^2\big(Z^{(n)}\big)$ as

$$\mu_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{n^{2/p} s_i^2(Z^{(n)})}.$$

Note that, just as before, the coordinates of the vector $\big(s_1^2\big(Z^{(n)}\big), \ldots, s_n^2\big(Z^{(n)}\big)\big) \in \mathbb{R}_+^n$ are suitably scaled. In the non-self-adjoint case, we refer to the rescaled empirical spectral measure with respect to the squared singular values when we talk of the empirical spectral measure. As in the previous section, a large deviation principle for the empirical spectral measure of a sequence of random matrices with distribution $\mathbf{U}_{n,p,\beta}^{\mathscr{M}}$ or $\mathbf{C}_{n,p,\beta}^{\mathscr{M}}$ on $\mathbb{B}_{p,\beta}^{n,\mathscr{M}}$ was proven in [62]. Especially, it was observed that the rate function in both cases is the same up to a constant. Slightly adapting the proof of Theorem 3.5.1, we can show that this phenomenon occurs in a more general context. The proof is now based on Theorem 3.3.5 and the norm distribution outlined in Remark 3.3.7 instead of Theorem 3.3.1 and Remark 3.3.4, but this time also on [62, Theorem 1.5] instead of [62, Theorem 1.1], the latter of which we stated as Lemma 3.5.3 above.

**Theorem 3.5.5** *Let $0 < p < \infty$, $\beta \in \{1,2,4\}$, and let $(\vartheta_n)_{n\in\mathbb{N}}$ be a sequence in $[0,1]$ with $\lim_{n\to\infty}\vartheta_n = \vartheta \in [0,1]$ and denote by $k(\vartheta) \ge 2$ the smallest number such that $\lim_{n\to\infty} n^{-k(\vartheta)}\,|\log(1-\vartheta_n)| < +\infty$. Also, let $(\alpha_n)_{n\in\mathbb{N}}$ be a positive, real sequence such that $\lim_{n\to\infty}\alpha_n n^{-2} = \alpha \in [0,\infty)$. For each $n \in \mathbb{N}$ let $\mathbf{W}_n = \vartheta_n\delta_0 + (1-\vartheta_n)\boldsymbol{G}(\alpha_n,1)$, and let $Z^{(n)}$ be a random matrix in $\mathbb{B}_{p,\beta}^{n,\mathscr{M}}$ chosen according to the distribution $\mathbf{P}_{n,p,\mathbf{W}_n,\beta}^{\mathscr{M}}$. Then the sequence of random probability measures $\mu_n := \frac{1}{n}\sum_{i=1}^{n}\delta_{n^{2/p}s_i^2(Z^{(n)})}$ satisfies a large deviation principle on $\mathcal{M}_1(\mathbb{R}_+)$ with speed $n^2$ and good rate function*

$$\mathcal{I}^{\mathscr{M}}_{\text{emp}}(\mu) = \begin{cases} \mathcal{I}^{\mathscr{M}}_{\text{cone}}(\mu) - c^{\mathscr{M}}_{(1-\vartheta)} & : \begin{array}{l} m_p(\mu) \le 1, k(\vartheta) \ge 2, \\ \alpha = 0 \end{array} \\ \mathcal{I}^{\mathscr{M}}_{\text{cone}}(\mu) + \frac{\beta}{p}\log\left(\frac{\beta}{p}\right) - \left(\frac{\beta}{p}+\alpha\right)\log\left(\frac{\beta}{p}+\alpha\right) & : \begin{array}{l} m_p(\mu) < 1, k(\vartheta) = 2, \\ \alpha > 0 \end{array} \\ -\alpha\log\left(\frac{1-m_p(\mu)}{\alpha}\right) - c^{\mathscr{M}}_{(1-\vartheta)} & \\ +\infty & : \textit{otherwise}, \end{cases}$$

*where*

$$\mathcal{I}^{\mathscr{M}}_{\text{cone}}(\mu) = \begin{cases} -\frac{\beta}{2}\int\limits_{\mathbb{R}}\int\limits_{\mathbb{R}} \log|x-y|\,\mu(\mathrm{d}x)\mu(\mathrm{d}y) + \frac{\beta}{p}\log\left(\frac{\sqrt{\pi}\,p\,\Gamma\left(\frac{p}{2}\right)}{2^p\sqrt{e}\,\Gamma\left(\frac{p+1}{2}\right)}\right) & : m_{p/2}(\mu) \le 1 \\ +\infty & : \textit{otherwise}, \end{cases}$$

*and*

$$c^{\mathscr{M}}_{(1-\vartheta)} := \begin{cases} \lim\limits_{n\to\infty} n^{-2}\log(1-\vartheta_n) & : k(\vartheta) = 2 \\ 0 & : k(\vartheta) > 2. \end{cases}$$

The proof of Theorem 3.5.5 is completely analogous to the one of Theorem 3.5.1, thus we will only point out the changes in the auxiliary results that need to be made.

**Lemma 3.5.6** *Let $0 < p < \infty$, $\beta \in \{1,2,4\}$, and let $(\vartheta_n)_{n\in\mathbb{N}}$ be a sequence in $[0,1]$ with $\lim_{n\to\infty}\vartheta_n = \vartheta \in [0,1]$ and denote by $k(\vartheta) \ge 2$ the smallest number such that $\lim_{n\to\infty} n^{-k(\vartheta)}\,|\log(1-\vartheta_n)| < +\infty$. Also, let $(\alpha_n)_{n\in\mathbb{N}}$ be a positive, real sequence such that $\lim_{n\to\infty}\alpha_n n^{-2} = \alpha \in [0,\infty)$. For each $n \in \mathbb{N}$ let $X^{(n)} = (X_1^{(n)},\dots,X_n^{(n)})$ be a random vector with density $C_{n,p/2,\nabla^c_\beta}\, e^{-\|x\|^{p/2}_{p/2}}\,\nabla^c_\beta(x)$, $x \in \mathbb{R}^n_+$, with $\nabla^c_\beta$ defined as in Theorem 3.3.5. Independently of the sequence $(X^{(n)})_{n\in\mathbb{N}}$, let $(W^{(n)})_{n\in\mathbb{N}}$ be a sequence of random variables with $W^{(n)} \sim \mathbf{W}_n = \vartheta_n\delta_0 + (1-\vartheta_n)\,\boldsymbol{G}(\alpha_n,1)$. Then the sequence of random variables $(B^{(n)})_{n\in\mathbb{N}}$ with $B^{(n)} := \left\|X^{(n)}\right\|^{p/2}_{p/2}/\left(\left\|X^{(n)}\right\|^{p/2}_{p/2} + W^{(n)}\right)$ satisfies a large deviation principle on $[0,\infty)$ with speed $n^2$ and good rate function*

$$\mathcal{I}^{\mathscr{M}}_{\text{beta}}(x) = \begin{cases} 0 & : k(\vartheta) > 2, x = 1 \\ -\frac{\beta}{p}\log(x) - c^{\mathscr{M}}_{(1-\vartheta)}(x) & : \begin{array}{l} k(\vartheta) = 2, \alpha = 0, \\ x \in (0,1] \end{array} \\ -\frac{\beta}{p}\log\left(\frac{xp}{\beta}\right) - \alpha\log\left(\frac{1-x}{\alpha}\right) - \left(\frac{\beta}{p}+\alpha\right)\log\left(\frac{\beta}{p}+\alpha\right) & : \begin{array}{l} k(\vartheta) = 2, \alpha > 0, \\ x \in (0,1) \end{array} \\ \quad - c^{\mathscr{M}}_{(1-\vartheta)}(x) & \\ +\infty & : \textit{otherwise}, \end{cases}$$

*where*

$$c^{\mathscr{M}}_{(1-\vartheta)} := \begin{cases} \lim\limits_{n\to\infty} n^{-2}\,\log(1-\vartheta_n) & : k(\vartheta) = 2 \\ 0 & : k(\vartheta) > 2. \end{cases}$$

This first lemma establishes an LDP for the beta distributed $\ell^n_{p/2}$-norm of the random vector $X^{(n)}/\big(\|X^{(n)}\|^{p/2}_{p/2} + W^{(n)}\big)^{2/p}$. This is proven in the same way as Lemma 3.4.5. In the non-self-adjoint case nothing changes in comparison to the self-adjoint case, besides the value for $p$ (which becomes $\frac{p}{2}$) and the density of the random vector $X^{(n)}$ underlying that representation. For the singular value distribution in non-self-adjoint matrix $p$-balls a different weight function $\nabla^c_\beta$ is needed with a different degree of homogeneity $m = (\frac{\beta}{2})n^2 - n$. This $m$ only plays a role in the first parameter of the beta distribution involved in the distribution of the $p$-radial component (see Remark 3.3.7). It affects the large deviation behaviour of the random variable $B^{(n)}$ only insofar as the limit of the first parameter changes from $\frac{\beta}{2p}$ to $\lim_{n\to\infty} n^{-2}(n+m)/(\frac{p}{2}) = \frac{\beta}{p}$.

The second lemma is the analogue of Lemma 3.5.3 and gives a large deviation principle for the empirical spectral measure of a non-self-adjoint random matrix in $\mathbb{B}^{n,\mathscr{M}}_{p,\beta}$ with distribution $\mathbf{C}^{\mathscr{M}}_{n,p,\beta}$. This result can be found in [62, Theorem 1.5].

**Lemma 3.5.7** *For $n \in \mathbb{N}$ let $Z^{(n)}$ be a random matrix in $\mathbb{B}^{n,\mathscr{M}}_{p,\beta}$ with distribution $\mathbf{C}^{\mathscr{M}}_{n,p,\beta}$ and singular values $s_i\big(Z^{(n)}\big)$, $i \in \{1,\dots,n\}$. Then the sequence of random probability measures $\mu_n := \frac{1}{n}\sum_{i=1}^n \delta_{n^{2/p}s_i^2(Z^{(n)})}$ satisfies a large deviation principle on $\mathcal{M}_1(\mathbb{R}_+)$ with speed $n^2$ and good rate function*

$$\mathcal{I}^{\mathscr{M}}_{\mathrm{cone}}(\mu) = \begin{cases} -\dfrac{\beta}{2}\displaystyle\int\limits_{\mathbb{R}_+}\int\limits_{\mathbb{R}_+} \log|x-y|\,\mu(\mathrm{d}x)\mu(\mathrm{d}y) + \dfrac{\beta}{p}\log\left(\dfrac{\sqrt{\pi}\,p\,\Gamma\left(\frac{p}{2}\right)}{2^p\sqrt{e}\,\Gamma\left(\frac{p+1}{2}\right)}\right) & : m_{p/2}(\mu) \le 1 \\ +\infty & : \textit{otherwise.} \end{cases}$$

From here on the proof will be completely analogous to that of Theorem 3.5.1, with the difference being that one uses the rate functions from Lemma 3.5.6 and Lemma 3.5.7 instead of those from Lemma 3.4.5 and Lemma 3.5.3, and the probabilistic representation from Theorem 3.3.5 instead of the one from Theorem 3.3.1. We thus again omit the details.

# Chapter 4

# Large deviations for random projections of $\ell_p^n$-balls

The objects of study in this chapter are projections of probability distributions on $\ell_p^n$-balls onto uniform random $k$-dimensional subspaces for $k \leq n$. The main goal is to show large deviation principles for these projected distributions on the space $\mathcal{M}_1(\mathbb{R}^k)$ of probability measures on $\mathbb{R}^k$. The distributions that are projected will be taken from the class of $p$-radial distributions $\mathbf{P}_{n,p,\mathbf{W}}$. The uniform random projection will be facilitated by a random variable $\mathbb{V}_{n,k}$ uniformly distributed on the so-called Stiefel manifold of orthonormal $k$-frames in $\mathbb{R}^n$.

The specifics of the set-up, which was already sketched in Section 1.2, shall be layed out in the following Section 4.1, followed by a formulation of the chapter's main results in Section 4.2 and their subsequent proofs in Section 4.3. The idea of both proofs can be formulated in three steps: First, one shows that the distance of any projected distribution (with respect to the Lévy–Prokhorov metric, see (4.4)) to some more simplified distribution tends to zero with increasing dimension $n \in \mathbb{N}$. For this simplified distribution a weak LDP already follows by results of Kabluchko and Prochno from [59], and hence we infer a weak LDP for said projected distribution as a second step. The third step then is to ameliorate this weak LDP to a full LDP by some compactness arguments.

## 4.1 Preliminaries

For $n, k \in \mathbb{N}, k \leq n$, a tuple of $k$ linearly independent vectors in $\mathbb{R}^n$ is called a $k$-frame. The set of all such $k$-frames whose constituent (row-)vectors are additionally orthonormal is referred to as the Stiefel manifold, denoted as

$$\mathbb{V}_{n,k} := \left\{ V \in \mathbb{R}^{k\times n} \colon VV^T = I_k \right\},$$

with $I_k$ denoting the $(k \times k)$ identity matrix and $V$ being a $(k \times n)$ matrix consisting row-wise of the $k$ orthonormal vectors from $\mathbb{R}^n$. As outlined in the introduction, we equip $\mathbb{V}_{n,k}$ with the uniform distribution (i.e., the invariant Haar probability measure, cf. [82, Chapter 3]) $\mathbf{U}_{n,k,\mathbb{V}}$, writing $V_{n,k}$ for the corresponding random variable, and recall that $V_{n,k}$ is characterized by the following invariance property: for any orthogonal matrices $O \in \mathbb{R}^{k\times k}$ and $O' \in \mathbb{R}^{n\times n}$, $OV_{n,k}O'$ has the same distribution as $V_{n,k}$. Also, the multiplication of a vector $X^{(n)}$ in $\mathbb{R}^n$ with an element $V \in \mathbb{V}_{n,k}$ from the Stiefel manifold can be interpreted geometrically as a projection of $X^{(n)}$ onto a $k$-dimensional subspace. Hence, for a random variable $V_{n,k}$ with distribution $\mathbf{U}_{n,k,\mathbb{V}}$ the multiplication of a vector $X^{(n)} \in \mathbb{R}^n$ with $V_{n,k}$ corresponds to a projection of $X^{(n)}$ onto a uniform random subspace of dimension $k$. For a random vector $X^{(n)}$ on $\mathbb{B}_p^n$ with distribution $\mathbf{P}_{n,p,\mathbf{W}}$ for some Borel probability measure $\mathbf{W}$ on $[0, \infty)$ as in (2.13) the distribution of $V_{n,k}X^{(n)}$ on $\mathbb{R}^k$, given by

$$\mu_{V_{n,k}X^{(n)}}(A) := \mathbb{P}\left(V_{n,k}X^{(n)} \in A\right), \qquad A \in \mathcal{B}(\mathbb{R}^d), \tag{4.1}$$

can be interpreted as the uniform random $k$-dimensional projection of $\mathbf{P}_{n,p,\mathbf{W}}$. This distribution is the object of study in this chapter. Viewed as a random variable on the space $\mathcal{M}_1(\mathbb{R}^k)$ of probability measures on $\mathbb{R}^k$, we want to analyze its large deviation behaviour. This will be done by transporting the following large deviation results onto the target sequence of random measures, which was originally given in [59, Theorem D]. It provides an LDP for random projections of product measures, which we will use in conjunction with the representation from Proposition 2.4.4 to prove this chapter's central theorems. In what follows we shall write

$$\mathcal{R}_2^{k\times\infty} := \left\{ A = (A_{ij})_{i,j=1}^{k,\infty} : \, (A_{ij})_{j\in\mathbb{N}} \in \ell_2, \, i = 1, \ldots, k \right\}$$

for the set of all matrices $A \in \mathbb{R}^{k\times\infty}$ with square-summable rows. For $A \in \mathbb{R}^{k\times\infty}$ we denote by $\|AA^T\|_{\mathrm{op}}$ the operator norm of the matrix $AA^T \in \mathbb{R}^{k\times k}$ (that is, the spectral norm given by the square root of the biggest eigenvalue of $AA^T$, see [88, Section 5.2]), where the condition $A \in \mathcal{R}_2^{k\times\infty}$ guarantees that $\|AA^T\|_{\mathrm{op}}$ is well-defined.

**Remark 4.1.1** Let us briefly address how to sample a uniform random $k$-frame from $\mathbb{V}_{n,k}$. As every $k$-frame in $\mathbb{V}_{n,k}$ is orthonormal, its $k$ row vectors are orthogonal and of length one, i.e., spherical. Hence, such a $k$-frame can be seen as an element of $\left(\mathbb{S}_2^{n-1}\right)^k$ with orthogonal coordinates. Thus, to uniformly sample a $k$-tuple of orthogonal spherical vectors, we begin by sampling a uniform vector $x_1$ from $\mathbb{S}_2^{n-1}$ by using Proposition 2.4.2, that is, sampling a standard Gaussian vector in $\mathbb{R}^n$ and normalizing. As for $p = 2$ the cone measure $\mathbf{C}_{n,p}$ and the surface measure $\sigma_p^n$ coincide (see Remark 2.4.1), this yields a uniform random spherical vector. Second, we sample another uniform random spherical vector $x_2$ from $\mathbb{S}_2^{n-1} \cap x_1^\perp$ in the same fashion, with $x_1^\perp$ denoting the orthogonal complement of $x_1$. We repeat this procedure successively until we have $k$ orthogonal vectors on $\mathbb{S}_2^{n-1}$, which, by construction, row-wise make up a uniform random $k$-frame.

We shall now present the results from [59, Theorem D].

**Proposition 4.1.2** *Fix $k \in \mathbb{N}$. For each $n \in \mathbb{N}$ let $Z^{(n)} = (Z_1, \dots, Z_n)$ be a random vector, where the $Z_i$ are i.i.d. non-Gaussian random variables with symmetric distribution, finite moments of all orders, and variance $\sigma^2 > 0$. Then the sequence of random probability measures $(\mu_{V_{n,k}Z^{(n)}})_{n\in\mathbb{N}}$, $n \geq k$, as in (4.1) satisfies an LDP on $\mathcal{M}_1(\mathbb{R}^k)$ with speed $n$ and good rate function $\mathcal{I}_{\text{proj}} \colon \mathcal{M}_1(\mathbb{R}^k) \to [0, \infty]$ given by*

$$\mathcal{I}_{\text{proj}}(\nu) = -\frac{1}{2} \log \det \left(I_k - AA^T\right),$$

*if $\nu$ admits a representation of the form*

$$\nu = \mathcal{D}\left(\sum_{j=1}^{\infty} A_{\bullet,j} Z_j + \sigma(I_k - AA^T)^{1/2} N_k\right)$$

*for some matrix $A \in \mathcal{R}_2^{k\times\infty}$ with columns $(A_{\bullet j})_{j\in\mathbb{N}}$ such that $\|AA^T\|_{\text{op}} < 1$, where $N_k$ is a $k$-dimensional standard Gaussian random vector independent of all $Z_i$. If $\nu$ does not admit a representation of this form, we set $\mathcal{I}_{\text{proj}}(\nu) = \infty$.*

Note that the specific distribution of the $Z_i$ has a rather subtle influence on the rate function of the LDP via the matrix $A$ used in the representation of a given measure $\nu \in \mathcal{M}_1(\mathbb{R}^k)$. As a side remark, note that in [59, Theorem D], the case of $\sigma^2 = 0$ actually has to be excluded. The result was amended accordingly.

## 4.2 LDPs for random projections of $p$-radial distributions on $\ell_p^n$-balls

We are now in the position to present the first of this chapter's main results for random projections of $p$-radial distributions $\mathbf{P}_{n,p,\mathbf{W}_n}$ on $\ell_p^n$-balls.

**Theorem 4.2.1** *Fix $p \in [1,\infty)$, $p \neq 2$, and $k \in \mathbb{N}$. Moreover, let $(\mathbf{W}_n)_{n\in\mathbb{N}}$ be a sequence of Borel probability measures on $[0,\infty)$ and $(W_n)_{n\in\mathbb{N}}$ a sequence of random variables with $W_n \sim \mathbf{W}_n$, such that $W_n/n \to \alpha \in [0,\infty)$ in probability. Finally, let $X^{(n)}$, $Y^{(n)}$ be random vectors in $\mathbb{B}_p^n$ with $Y^{(n)} \sim \mathbf{P}_{n,p,\mathbf{W}_n}$ and $X^{(n)} \stackrel{\mathcal{D}}{=} n^{1/p} Y^{(n)}$. Then the sequence of random probability measures $(\mu_{V_{n,k}X^{(n)}})_{n\in\mathbb{N}}$, $n \geq k$, as in (4.1) satisfies an LDP on $\mathcal{M}_1(\mathbb{R}^k)$ with speed $n$ and good rate function $\mathcal{I}_{\mathrm{proj}} \colon \mathcal{M}_1(\mathbb{R}^k) \to [0,\infty]$ given by*

$$\mathcal{I}_{\mathrm{proj}}(\nu) = -\frac{1}{2}\log\det\left(I_k - AA^T\right),$$

*if $\nu$ admits a representation of the form*

$$\nu = \mathcal{D}\left(\left(\frac{1}{1+\alpha}\right)^{1/p}\sum_{j=1}^{\infty} A_{\bullet,j}Z_j + \sigma_{p,\alpha}\left(I_k - AA^T\right)^{1/2} N_k\right)$$

*for some matrix $A \in \mathcal{R}_2^{k\times\infty}$ with columns $(A_{\bullet j})_{j\in\mathbb{N}}$ such that $\|AA^T\|_{\mathrm{op}} < 1$, where $Z_j \sim \mathbf{N}_p$ i.i.d.,*

$$\sigma_{p,\alpha}^2 := \left(\frac{p}{1+\alpha}\right)^{2/p}\frac{\Gamma\left(\frac{3}{p}\right)}{\Gamma\left(\frac{1}{p}\right)},$$

*and $N_k$ is an independent $k$-dimensional standard Gaussian random vector. If $\nu$ does not admit a representation of this form, we set $\mathcal{I}_{\mathrm{proj}}(\nu) = \infty$.*

As discussed in Remark 2.4.3, choosing $\mathbf{W}_n \equiv \delta_0$ gives $\mathbf{P}_{n,p,\mathbf{W}_n} = \mathbf{C}_{n,p}$ and $\mathbf{W}_n \equiv \mathbf{E}(1)$ yields $\mathbf{P}_{n,p,\mathbf{W}_n} = \mathbf{U}_{n,p}$. In both cases it holds for $W_n \sim \mathbf{W}_n$ that $W_n/n \to \alpha = 0$ in probability and Theorem 4.2.1 reduces to the results from [59, Theorem C]. Hence, we can see that in this setting both $\mathbf{C}_{n,p}$ and $\mathbf{U}_{n,p}$ share the same large deviation behaviour in high dimensions, which is in line with similar observations made for other functionals (see, e.g., [4, 62, 72]). Moreover, the result even implies a certain universality of the rate function, since despite the expected sensitivity of LDPs to the underlying distributions, the rate function is the same for all sequences $(\mathbf{W}_n)_{n\in\mathbb{N}}$ that share the same limiting behaviour on a scale of order $n$.

Given the setting of Theorem 4.2.1, if we consider the case $W_n/n \to \infty$ in probability (formally corresponding to the choice $\alpha = \infty$), by the representation result in Proposition 2.4.4 one can see that this corresponds to each component of $X^{(n)}$ converging to zero in probability, that is, we arrive at a trivial limit. To avoid this, we may choose a different scaling, as carried out in the following theorem.

**Theorem 4.2.2** *Fix $p \in [1,\infty)$, $p \neq 2$, and $k \in \mathbb{N}$. Moreover, let $(\mathbf{W}_n)_{n\in\mathbb{N}}$ be a sequence of Borel probability measures on $[0,\infty)$ and $(W_n)_{n\in\mathbb{N}}$ a sequence of random variables with $W_n \sim \mathbf{W}_n$ and $W_n/n^\kappa \to \beta \in (0,\infty)$ in probability for some $\kappa > 1$, and assume that the sequence of random variables $(W_n/n^\kappa)^{-2/p}$ is uniformly integrable. Finally, let $X^{(n)}$ and $Y^{(n)}$ be random vectors in $\mathbb{B}_p^n$ such that $Y^{(n)} \sim \mathbf{P}_{n,p,\mathbf{W}_n}$ and $X^{(n)} \overset{\mathcal{D}}{=} n^{\kappa/p} Y^{(n)}$. Then the sequence of random probability measures $(\mu_{V_{n,k}X^{(n)}})_{n\in\mathbb{N}}$, $n \geq k$, as in (4.1) satisfies an LDP on $\mathcal{M}_1(\mathbb{R}^k)$ with speed $n$ and good rate function $\mathcal{I}_{\mathrm{proj}} \colon \mathcal{M}_1(\mathbb{R}^k) \to [0,\infty]$ given by*

$$\mathcal{I}_{\mathrm{proj}}(\nu) = -\frac{1}{2}\log\det\left(I_k - AA^T\right)$$

*if $\nu$ admits a representation of the form*

$$\nu = \mathcal{D}\left(\left(\frac{1}{\beta}\right)^{1/p}\sum_{j=1}^{\infty} A_{\bullet,j}Z_j + \sigma_{p,\beta}\left(I_k - AA^T\right)^{1/2} N_k\right)$$

*for some matrix $A \in \mathcal{R}_2^{k\times\infty}$ with columns $(A_{\bullet j})_{j\in\mathbb{N}}$ such that $\|AA^T\|_{\mathrm{op}} < 1$, where $Z_j \sim \mathbf{N}_p$,*

$$\sigma_{p,\beta}^2 := \left(\frac{p}{\beta}\right)^{2/p}\frac{\Gamma\left(\frac{3}{p}\right)}{\Gamma\left(\frac{1}{p}\right)},$$

*and $N_k$ is an independent $k$-dimensional standard Gaussian random vector. If $\nu$ does not admit a representation of this form, we set $\mathcal{I}_{\mathrm{proj}}(\nu) = \infty$.*

Theorem 4.2.2 includes an additional integrability condition for $W_n$, due to the fact that $W_n$ is no longer of the same order as the $\ell_p^n$-norm of $Z^{(n)}$. Why this is needed specifically can be seen in its proof. Note that a helpful sufficient condition for the uniform integrability of $(W_n/n^\kappa)^{-2/p}$ is given by

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left[\left(\frac{n^\kappa}{W_n}\right)^{4/p}\right] \leq C$$

for some absolute constant $C > 0$ (since a sequence of random variables being bounded in $L_p$ for some $p > 1$ – with $p = 2$ in this case – implies its uniform integrability).

In particular, it can be applied to verify the uniform integrability for certain gamma distributions. In such settings the following lemma will also be useful.

**Lemma 4.2.3** *For each $n \in \mathbb{N}$ consider a random variable $W_n \sim \mathbf{G}(a_n, b)$, where $(a_n)_{n\in\mathbb{N}}$ is a positive increasing sequence and $b > 0$. Assume further that $a_n$ satisfies $\inf_{n\in\mathbb{N}} a_n =: m > \frac{4}{p}$ and $\lim_{n\to\infty} \frac{a_n}{n^\kappa} = \lambda \in (0,\infty)$ for some $\kappa \in (0,\infty)$. Then*

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left[\left(\frac{n^\kappa}{W_n}\right)^{4/p}\right] \leq b^{4/p} M_p (\lambda m)^{-4/p} < \infty,$$

*where $M_p^{-1} := \prod_{i=0}^4 \left(1 - \frac{4}{p(m+i)}\right)$.*

*Proof.* We start by observing that

$$\mathbb{E}\left[\left(\frac{n^\kappa}{W_n}\right)^{4/p}\right] = n^{4\kappa/p} \frac{b^{a_n}}{\Gamma(a_n)} \int_0^\infty x^{a_n - 1 - 4/p} e^{-bx}\,\mathrm{d}x = n^{4\kappa/p} \frac{b^{4/p}}{\Gamma(a_n)} \Gamma\left(a_n - \frac{4}{p}\right).$$

According to the inequality [55, Equation (12)] for quotients of gamma functions (applied with $x = a_n + 5 - \frac{4}{p}$ and $y = \frac{4}{p}$) one has that

$$\begin{aligned}
\frac{\Gamma\left(a_n - \frac{4}{p}\right)}{\Gamma(a_n)} &= \left(\prod_{i=0}^4 \frac{(a_n + i)}{\left(a_n - \frac{4}{p} + i\right)}\right) \frac{\Gamma\left(a_n - \frac{4}{p} + 5\right)}{\Gamma(a_n + 5)} \\
&= \left(\prod_{i=0}^4 \left(1 - \frac{4}{p(a_n + i)}\right)\right)^{-1} \frac{1}{\left(\frac{\Gamma\left(a_n - \frac{4}{p} + 5\right)}{\Gamma(a_n + 5)}\right)} \\
&\leq \left(\prod_{i=0}^4 \left(1 - \frac{4}{p(m + i)}\right)\right)^{-1} \frac{1}{\left(a_n + 4 - \frac{4}{p}\right)^{4/p}} \\
&\leq \left(\prod_{i=0}^4 \left(1 - \frac{4}{p(m + i)}\right)\right)^{-1} \frac{1}{a_n^{4/p}} \\
&= M_p\, a_n^{-4/p},
\end{aligned}$$

where we also used that $p \geq 1$. By our assumption on the growth of $a_n$ and since $a_n$ is increasing it follows that

$$\mathbb{E}\left[\left(\frac{n^\kappa}{W_n}\right)^{4/p}\right] \leq n^{4\kappa/p} b^{4/p} M_p a_n^{-4/p} \leq n^{4\kappa/p} b^{4/p} M_p (\lambda m n^\kappa)^{-4/p} = b^{4/p} M_p (\lambda m)^{-4/p}$$

for all $n \in \mathbb{N}$. This completes the proof. □

**Remark 4.2.4** A concrete and geometrically motivated example where Theorem 4.2.1 and Theorem 4.2.2 can be applied is given by the distribution on $\mathbb{B}_p^n$ arising as the projection to the first $n$ coordinates of the cone probability measure $\mathbf{C}_{n+m_n,p}$ on $\mathbb{B}_p^{n+m_n}$, where $m_n$ is an element of an increasing sequence $(m_n)_{n\in\mathbb{N}}$ satisfying $\inf_{n\in\mathbb{N}} m_n = m > 4$ and $\lim_{n\to\infty} \frac{m_n}{n^\kappa} = \lambda$ for some $\kappa \geq 1$ and $\lambda \in (0,\infty)$. As discussed in Remark 2.4.3, this case corresponds to $\mathbf{P}_{n,p,\mathbf{W}_n}$ with $\mathbf{W}_n = \mathbf{G}\big(\frac{m_n}{p}, \frac{1}{p}\big)$ and fits the assumptions of Lemma 4.2.3, and thereby of Theorem 4.2.1 or Theorem 4.2.2, depending on the value of $\kappa \geq 1$. The same holds for the projection of the uniform distribution $\mathbf{U}_{n+m_n,p}$ corresponding to $\mathbf{P}_{n,p,\mathbf{W}_n}$ with $\mathbf{W}_n = \mathbf{G}\big(1 + \frac{m_n}{p}, \frac{1}{p}\big)$. Thus, the LDPs from the main results of this section do hold for distributions of particular geometric interest.

## 4.3 Proof of the LDPs for projections of $p$-radial distributions on $\ell_p^n$-balls

This section shall prove Theorem 4.2.1 and Theorem 4.2.2. The proofs will follow in the footsteps of the proof of [59, Theorem C], adapting and generalizing the arguments therein where necessary. We start off by formulating some probabilistic representations of the target quantities and show some auxiliary results for the proofs.

Assume the set-up of Theorem 4.2.1 and for a fixed Stiefel matrix $V \in \mathbb{V}_{n,k}$ denote by $V_{\bullet,j}$, $j = 1,\dots,n$ its columns. Then, by (4.1) and the representation results from Proposition 2.4.4, it follows that for any Borel set $A \in \mathcal{B}(\mathbb{R}^k)$,

$$\mu_{VX^{(n)}}(A) = \mathbb{P}(VX^{(n)} \in A) = \mathbb{P}\left(\sum_{j=1}^n n^{1/p} \frac{Z_j}{(\|Z^{(n)}\|_p^p + W_n)^{1/p}} V_{\bullet,j} \in A\right), \tag{4.2}$$

where $Z^{(n)} = (Z_1,\dots,Z_n)$ with $Z_j \sim \mathbf{N}_p$ i.i.d. and $W_n \sim \mathbf{W}_n$ independent of $Z^{(n)}$. Moreover, let

$$\tilde{\mu}_{VX^{(n)}}(A) := \mathbb{P}\left(\left(\frac{1}{1+\alpha}\right)^{1/p} \sum_{j=1}^n Z_j V_{\bullet,j} \in A\right), \tag{4.3}$$

again with i.i.d. $Z_j \sim \mathbf{N}_p$. We shall see that we can confine our analysis to $\tilde{\mu}_{VX^{(n)}}$ instead of $\mu_{VX^{(n)}}$, since they are arbitrarily close to each other in $n \in \mathbb{N}$ with respect to the Lévy-Prokhorov metric. On the space $\mathcal{M}_1(\mathbb{R}^k)$ of probability measures on $\mathbb{R}^k$, the Lévy–Prokhorov metric $\rho_{\mathrm{LP}}$ is defined by

$$\rho_{\mathrm{LP}}(\mu,\nu) := \inf\left\{\varepsilon > 0\colon \mu(A) \leq \nu(A_\varepsilon) + \varepsilon,\, \nu(A) \leq \mu(A_\varepsilon) + \varepsilon \;\forall\, A \in \mathcal{B}\big(\mathbb{R}^k\big)\right\}, \tag{4.4}$$

where $A_\varepsilon$ denotes the $\varepsilon$-neighborhood of $A \in \mathcal{B}(\mathbb{R}^k)$, defined as

$$A_\varepsilon := \{x \in \mathbb{R}^k \colon \|a - x\|_2 < \varepsilon \text{ for some } a \in A\}, \qquad \varepsilon > 0.$$

We shall now prove that the distance of $\mu_{VX^{(n)}}$ and $\tilde{\mu}_{VX^{(n)}}$ in $\mathcal{M}_1(\mathbb{R}^k)$ with respect to the Lévy–Prokhorov metric, i.e., $\rho_{\mathrm{LP}}(\mu_V, \tilde{\mu}_V)$, converges to 0 uniformly over all $V \in \mathbb{V}_{n,k}$ as $n$ tends to infinity.

**Lemma 4.3.1** *For $p \in [1,\infty)$ and $n \in \mathbb{N}$, $n > 4$, set $X^{(n)}$ as in Theorem 4.2.1. Then, for $k \le n$, we have*

$$\lim_{n\to\infty} \sup_{V \in \mathbb{V}_{n,k}} \rho_{\mathrm{LP}}\left(\mu_{VX^{(n)}}, \tilde{\mu}_{VX^{(n)}}\right) = 0.$$

*Proof.* Let $A \in \mathcal{B}(\mathbb{R}^k)$, $V \in \mathbb{V}_{n,k}$ fixed, and $\varepsilon > 0$. Then,

$$\begin{aligned}
&\tilde{\mu}_{VX^{(n)}}(A) \\
=\ & \mathbb{P}\left(\left(\frac{1}{1+\alpha}\right)^{1/p} \sum_{j=1}^n Z_j V_{\bullet,j} \in A\right) \\
\le\ & \mathbb{P}\left(\sum_{j=1}^n n^{1/p} \frac{Z_j}{(\|Z^{(n)}\|_p^p + W_n)^{1/p}} V_{\bullet,j} \in A_\varepsilon\right) \\
& + \mathbb{P}\left(\left\| \left(\frac{1}{1+\alpha}\right)^{1/p} \sum_{j=1}^n Z_j V_{\bullet,j} - \sum_{j=1}^n n^{1/p} \frac{Z_j}{(\|Z^{(n)}\|_p^p + W_n)^{1/p}} V_{\bullet,j} \right\|_2 \ge \varepsilon\right) \\
=\ & \mu_{VX^{(n)}}(A_\varepsilon) \\
& + \mathbb{P}\left(\left\| \left(\frac{1}{1+\alpha}\right)^{1/p} \sum_{j=1}^n Z_j V_{\bullet,j} - \sum_{j=1}^n n^{1/p} \frac{Z_j}{(\|Z^{(n)}\|_p^p + W_n)^{1/p}} V_{\bullet,j} \right\|_2 \ge \varepsilon\right).
\end{aligned} \tag{4.5}$$

Let us prove that the second summand on the right-hand side converges to 0 as $n \to \infty$. By Markov's inequality,

$$\begin{aligned}
&\mathbb{P}\left(\left\| \left(\frac{1}{1+\alpha}\right)^{1/p} \sum_{j=1}^n Z_j V_{\bullet,j} - \sum_{j=1}^n n^{1/p} \frac{Z_j}{(\|Z^{(n)}\|_p^p + W_n)^{1/p}} V_{\bullet,j} \right\|_2 \ge \varepsilon\right) \\
&\qquad \le \varepsilon^{-1}\, \mathbb{E} \left\| \left(\frac{1}{1+\alpha}\right)^{1/p} \sum_{j=1}^n Z_j V_{\bullet,j} - \sum_{j=1}^n n^{1/p} \frac{Z_j}{(\|Z^{(n)}\|_p^p + W_n)^{1/p}} V_{\bullet,j} \right\|_2,
\end{aligned}$$

and by the Cauchy–Schwarz inequality,

$$
\begin{aligned}
&\mathbb{E}\left\| \left(\frac{1}{1+\alpha}\right)^{1/p} \sum_{j=1}^{n} Z_j V_{\bullet,j} - \sum_{j=1}^{n} n^{1/p} \frac{Z_j}{(\|Z^{(n)}\|_p^p + W_n)^{1/p}} V_{\bullet,j} \right\|_2 \\
&\quad = \mathbb{E}\left( \left\| \sum_{j=1}^{n} Z_j V_{\bullet,j} \right\|_2 \left| \left(\frac{1}{1+\alpha}\right)^{1/p} - \frac{n^{1/p}}{(\|Z^{(n)}\|_p^p + W_n)^{1/p}} \right| \right) \\
&\quad \leq \sqrt{\mathbb{E}\left[ \left\| \sum_{j=1}^{n} Z_j V_{\bullet,j} \right\|_2^2 \right]} \sqrt{\mathbb{E}\left[ \left| \left(\frac{1}{1+\alpha}\right)^{1/p} - \frac{n^{1/p}}{(\|Z^{(n)}\|_p^p + W_n)^{1/p}} \right|^2 \right]}. \qquad (4.6)
\end{aligned}
$$

As $Z_1, \dots, Z_n$ are i.i.d. with mean zero and $V_{\bullet,1}, \dots, V_{\bullet,n}$ are orthonormal vectors, the first factor in (4.6) reads

$$
\mathbb{E}\left[ \left\| \sum_{j=1}^{n} Z_j V_{\bullet,j} \right\|_2^2 \right] = \mathbb{E}\left[ \sum_{i,j=1}^{n} Z_i Z_j \langle V_{\bullet,i}, V_{\bullet,j} \rangle_2 \right] = \mathbb{E}\big[Z_1^2\big] \sum_{j=1}^{n} \langle V_{\bullet,i}, V_{\bullet,j} \rangle_2 = k\mathbb{E}\big[Z_1^2\big]. \quad (4.7)
$$

To address the second factor in (4.6), let us first argue that

$$
\xi_n := \left( \left(\frac{1}{1+\alpha}\right)^{1/p} - \frac{n^{1/p}}{(\|Z^{(n)}\|_p^p + W_n)^{1/p}} \right)^2 \longrightarrow 0 \qquad (4.8)
$$

in probability as $n \to \infty$. Indeed, by the continuous mapping theorem, it suffices to show that

$$
\frac{\|Z^{(n)}\|_p^p}{n} + \frac{W_n}{n} \longrightarrow 1 + \alpha
$$

in probability. This follows from the fact that $Z_1, \dots, Z_n$ are i.i.d. $p$-generalized Gaussian random variables, which means that $\mathbb{E}[|Z_i|^p] = 1$ (see (2.7)), and moreover that by assumption, $W_n \, n^{-1} \to \alpha$ in probability. In fact, we even have $\xi_n \to 0$ in $L_1$. To see this, by the Vitali convergence theorem (see [16, Theorem 4.5.4]), it is sufficient to show that $(\xi_n)_{n\in\mathbb{N}}$ is uniformly integrable, which in combination with (4.8) yields convergence in $L_1$. Clearly, $(\xi_n)_{n\in\mathbb{N}}$ is uniformly integrable if the sequence

$$
\left( \frac{n}{\|Z^{(n)}\|_p^p + W_n} \right)^{2/p} \leq \left( \frac{n}{\|Z\|_p^p} \right)^{2/p}
$$

is uniformly integrable, where we have used that $W_n \geq 0$. This, in turn, follows from the fact that $\|Z^{(n)}\|_p^p \sim \mathbf{G}\big(\frac{n}{p}, \frac{1}{p}\big)$ together with Lemma 4.2.3 for rate $a_n = \frac{n}{p} > \frac{1}{p}$, shape $b = \frac{1}{p}$, $\kappa = 1$, and $\lambda = \frac{1}{p}$, which yields

$$\mathbb{E}\left(\frac{n}{\|Z^{(n)}\|_p^p}\right)^{4/p} \leq p^{4/p} M_p\, m^{-4/p} \in (0,\infty)$$

for all $n > 4$. Hence, $\xi_n \longrightarrow 0$ in $L_1$, and as a consequence, the second factor in (4.6) converges to zero. This implies that the second summand in (4.5) converges to zero uniformly in $n \in \mathbb{N}$. Altogether, we have proven that for any $\varepsilon > 0$,

$$\tilde{\mu}_{VX^{(n)}}(A) \leq \mu_{VX^{(n)}}(A_\varepsilon) + \varepsilon$$

for $n$ sufficiently large. In the same way, we may also prove that

$$\mu_{VX^{(n)}}(A) \leq \tilde{\mu}_{VX^{(n)}}(A_\varepsilon) + \varepsilon$$

for $n$ sufficiently large, which finishes the proof. □

Next, we replace the fixed element $V \in \mathbb{V}_{n,k}$ by a random variable $V_{n,k} \sim \mathbf{U}_{n,k,\mathbb{V}}$ on the Stiefel manifold. Based on Lemma 4.3.1, we may prove that a weak LDP for the modified sequence $\tilde{\mu}_{V_{n,k}X^{(n)}}$ implies a weak LDP for $\mu_{V_{n,k}X^{(n)}}$, both respectively defined as in (4.2) and (4.3) with respect to $V_{n,k}$.

**Lemma 4.3.2** *Assume the set-up of Theorem 4.2.1 and recall the notation* (4.2) *and* (4.3). *If the sequence $\tilde{\mu}_{V_{n,k}X^{(n)}}$ satisfies a weak LDP on $\mathcal{M}_1(\mathbb{R}^k)$ at speed $n$ and rate function $\mathcal{I}_{\mathrm{proj}}$, then the sequence $\mu_{V_{n,k}X^{(n)}}$ satisfies the same weak LDP.*

*Proof.* It suffices to check the weak LDP on a basis of the topology of $\mathcal{M}_1(\mathbb{R}^k)$, e. g., the balls

$$B_r(\nu) := \left\{\mu \in \mathcal{M}_1\left(\mathbb{R}^k\right) \colon \rho_{\mathrm{LP}}(\mu,\nu) < r\right\}$$

for any radius $r \in (0,\infty)$. By Lemma 4.3.1, for $n$ sufficiently large we have

$$\rho_{\mathrm{LP}}\left(\tilde{\mu}_{V_{n,k}X^{(n)}}, \mu_{V_{n,k}X^{(n)}}\right) < r/2$$

uniformly over all realizations of $V_{n,k} \in \mathbb{V}_{n,k}$. Therefore, by the triangle inequality for the Lévy–Prokhorov metric $\rho_{\mathrm{LP}}$, it follows that

$$\begin{aligned}\frac{1}{n}\log\mathbb{P}\left(\tilde{\mu}_{V_{n,k}X^{(n)}} \in B_{r/2}(\nu)\right) &\leq \frac{1}{n}\log\mathbb{P}\left(\mu_{V_{n,k}X^{(n)}} \in B_r(\nu)\right)\\ &\leq \frac{1}{n}\log\mathbb{P}\left(\tilde{\mu}_{V_{n,k}X^{(n)}} \in B_{3r/2}(\nu)\right),\end{aligned}$$

and hence,

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\tilde{\mu}_{V_{n,k}X^{(n)}}\in B_{r/2}(\nu)\right) &\le \liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\mu_{V_{n,k}X^{(n)}}\in B_r(\nu)\right)\\ &\le \limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\mu_{V_{n,k}X^{(n)}}\in B_r(\nu)\right)\\ &\le \liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\tilde{\mu}_{V_{n,k}X^{(n)}}\in B_{3r/2}(\nu)\right).\end{aligned}$$

Thus, by monotonicity in $r$, taking the infimum over $r\in(0,\infty)$, the LDP for $\tilde{\mu}_{V_{n,k}X^{(n)}}$ yields

$$\begin{aligned}-\mathcal{I}_{\mathrm{proj}}(\nu) &\le \inf_{r\in(0,\infty)}\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\mu_{V_{n,k}X^{(n)}}\in B_r(\nu)\right)\\ &\le \inf_{r\in(0,\infty)}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\mu_{V_{n,k}X^{(n)}}\in B_r(\nu)\right)\\ &\le -\mathcal{I}_{\mathrm{proj}}(\nu).\end{aligned}$$

From here the claim follows directly from Proposition 2.3.8. □

On a compact space, weak and full LDPs coincide. Thus, to "lift" a weak LDP to a full LDP in our setting, we provide compactness by the following lemma.

**Lemma 4.3.3** *There is a constant $C\in(0,\infty)$ such that for all $n\ge k$ and all $V\in\mathbb{V}_{n,k}$,*

$$\mu_{VX^{(n)}}\in M_C:=\left\{\mu\in\mathcal{M}_1(\mathbb{R}^k):\int_{\mathbb{R}^k}\|x\|_2\,\mu(\mathrm{d}x)\le C\right\},$$

*where the set $M_C$ is compact for any choice of $C\in(0,\infty)$.*

*Proof.* The compactness of the set $M_C$ in the topology of weak convergence on $\mathcal{M}_1(\mathbb{R}^k)$ has been shown in [59, Proof of Lemma 5.3], so it remains to prove the first assertion. To this end, recalling the representation of the distribution $\mu_{VX^{(n)}}$ given in (4.2), it suffices to prove that

$$\limsup_{n\to\infty}\sup_{V\in\mathbb{V}_{n,k}}\mathbb{E}\left\|\sum_{j=1}^{n}n^{1/p}\frac{Z_j}{(\|Z^{(n)}\|_p^p+W_n)^{1/p}}V_{\bullet,j}\right\|_2<\infty,$$

for i.i.d. $Z_j\sim\mathbf{N}_p$ and $W_n\sim\mathbf{W}_n$ as in Theorem 4.2.1. By the triangle inequality it then follows that

$$\begin{aligned}
& \mathbb{E}\left\| \sum_{j=1}^{n} n^{1/p} \frac{Z_j}{(\|Z^{(n)}\|_p^p + W_n)^{1/p}} V_{\bullet,j} \right\|_2 \\
\leq \quad & \mathbb{E}\left\| \left(\frac{1}{1+\alpha}\right)^{1/p} \sum_{j=1}^{n} Z_j V_{\bullet,j} - \sum_{j=1}^{n} n^{1/p} \frac{Z_j}{(\|Z^{(n)}\|_p^p + W_n)^{1/p}} V_{\bullet,j} \right\|_2 \\
+ \; & \mathbb{E}\left\| \left(\frac{1}{1+\alpha}\right)^{1/p} \sum_{j=1}^{n} Z_j V_{\bullet,j} \right\|_2 .
\end{aligned}$$

The first summand on the right-hand side converges to zero uniformly in $n \in \mathbb{N}$, as was shown after (4.6). Moreover, by Hölder's inequality and (4.7), the second summand is uniformly bounded by $\sqrt{k\mathbb{E}[Z_1^2]}/(1+\alpha)^{1/p}$, and thus the claim follows. □

Combining the accumulated auxiliary results, we now have the sufficient tools to prove Theorem 4.2.1.

*Proof of Theorem 4.2.1.* We apply Proposition 4.1.2 to a random vector with coordinates being the symmetric non-Gaussian random variables $Z_j/(1+\alpha)^{1/p}$, $Z_j \sim \mathbf{N}_p$, which, by (2.7), have finite moments of all orders and, in particular, variance

$$\sigma_{p,\alpha}^2 = \left(\frac{p}{1+\alpha}\right)^{2/p} \frac{\Gamma\left(\frac{3}{p}\right)}{\Gamma\left(\frac{1}{p}\right)}.$$

Hence, the sequence $(\tilde{\mu}_{V_{n,k}X^{(n)}})_{n\in\mathbb{N}}$ as in (4.3) satisfies an LDP on $\mathcal{M}_1(\mathbb{R}^k)$ with speed $n$ and rate function $\mathcal{I}_{\text{proj}}$ as stated in Theorem 4.2.1. Therefore, by Lemma 4.3.2, $\mu_{V_{n,k}X^{(n)}}$ satisfies the same weak LDP, which extends to a full LDP by the compactness arguments given in Lemma 4.3.3, thus finishing the proof. □

The proof of Theorem 4.2.2 works in a very similar way to that of Theorem 4.2.1, hence we will only point out the steps where it differs from the previous proof. Given the different scaling of the $p$-radial component of $X^{(n)}$, it follows via the same probabilistic representation arguments as previously that for a Stiefel matrix $V \in \mathbb{V}_{n,k}$,

$$\mu_{VX^{(n)}}(A) := \mathbb{P}\left(VX^{(n)} \in A\right) = \mathbb{P}\left(\sum_{j=1}^{n} n^{\kappa/p} \frac{Z_j}{(\|Z^{(n)}\|_p^p + W_n)^{1/p}} V_{\bullet,j} \in A\right)$$

for any $A \in \mathcal{B}(\mathbb{R}^k)$, and set

$$\tilde{\mu}_{VX^{(n)}}(A) := \mathbb{P}\left(\left(\frac{1}{\beta}\right)^{1/p}\sum_{j=1}^{n} Z_j V_{\bullet,j} \in A\right),$$

using the same notation as in Theorem 4.2.1 and its proof. The only argument that needs to be adapted is the proof of Lemma 4.3.1, which will be replaced by the following Lemma.

**Lemma 4.3.4** *For $p \in [1,\infty)$ and any $n \in \mathbb{N}$, $n > 4$, set $X^{(n)}$ as in Theorem 4.2.2. Then, for $k \le n$, we have*

$$\lim_{n\to\infty} \sup_{V\in\mathbb{V}_{n,k}} \rho_{\mathrm{LP}}\left(\mu_{VX^{(n)}}, \tilde{\mu}_{VX^{(n)}}\right) = 0.$$

*Proof.* Let $A \in \mathcal{B}(\mathbb{R}^k)$, $V \in \mathbb{V}_{n,k}$ fixed, and $\varepsilon > 0$. Then, by analogue expansion as in (4.5), we have that

$$\begin{aligned}
&\tilde{\mu}_{VX^{(n)}}(A)\\
\le\ & \mu_{VX^{(n)}}(A_\varepsilon)\\
&+\mathbb{P}\left(\left\|\left(\frac{1}{\beta}\right)^{1/p}\sum_{j=1}^{n} Z_j V_{\bullet,j} - \sum_{j=1}^{n} n^{\kappa/p}\frac{Z_j}{(\|Z^{(n)}\|_p^p + W_n)^{1/p}} V_{\bullet,j}\right\|_2 \ge \varepsilon\right). \qquad (4.9)
\end{aligned}$$

Again, we need to show that the second summand on the right-hand side in the above converges to zero as $n$ tends to infinity. By Markov's inequality it holds that

$$\begin{aligned}
&\mathbb{P}\left(\left\|\left(\frac{1}{\beta}\right)^{1/p}\sum_{j=1}^{n} Z_j V_{\bullet,j} - \sum_{j=1}^{n} n^{\kappa/p}\frac{Z_j}{(\|Z^{(n)}\|_p^p + W_n)^{1/p}} V_{\bullet,j}\right\|_2 \ge \varepsilon\right)\\
&\qquad \le \varepsilon^{-1}\mathbb{E}\left\|\left(\frac{1}{\beta}\right)^{1/p}\sum_{j=1}^{n} Z_j V_{\bullet,j} - \sum_{j=1}^{n} n^{\kappa/p}\frac{Z_j}{(\|Z^{(n)}\|_p^p + W_n)^{1/p}} V_{\bullet,j}\right\|_2,
\end{aligned}$$

and a further application of the Cauchy–Schwarz inequality as in (4.6) yields

$$\begin{aligned}
&\mathbb{E}\left\|\left(\frac{1}{\beta}\right)^{1/p}\sum_{j=1}^{n} Z_j V_{\bullet,j} - \sum_{j=1}^{n} n^{\kappa/p}\frac{Z_j}{(\|Z^{(n)}\|_p^p + W_n)^{1/p}} V_{\bullet,j}\right\|_2\\
&\quad \le \sqrt{\mathbb{E}\left[\left\|\sum_{j=1}^{n} Z_j V_{\bullet,j}\right\|_2^2\right]}\sqrt{\mathbb{E}\left[\left|\left(\frac{1}{\beta}\right)^{1/p} - \frac{n^{\kappa/p}}{(\|Z^{(n)}\|_p^p + W_n)^{1/p}}\right|^2\right]}, \qquad (4.10)
\end{aligned}$$

with the first factor simplifying to $k\mathbb{E}[Z_1^2]$ as in (4.7). It remains to show that

$$\xi_n := \left( \left(\frac{1}{\beta}\right)^{1/p} - \frac{n^{\kappa/p}}{(\|Z^{(n)}\|_p^p + W_n)^{1/p}} \right)^2 \longrightarrow 0 \tag{4.11}$$

in probability as $n \to \infty$ to address the second factor. We again do so by showing

$$\frac{\|Z^{(n)}\|_p^p}{n^\kappa} + \frac{W_n}{n^\kappa} \longrightarrow \beta$$

in probability due to the continuous mapping theorem. Since $\kappa > 1$, by the same arguments as in the proof of Theorem 4.2.1, it follows that $\|Z^{(n)}\|_p^p/n^\kappa \to 0$ and the behaviour of $W_n$ dominates. By assumption, $W_n/n^\kappa \longrightarrow \beta$ in probability. In fact, we even have $\xi_n \longrightarrow 0$ in $L_1$. Indeed, since the sequence of random variables $(W_n/n^\kappa)^{-2/p}$ is uniformly integrable by assumption, it follows that the sequence of random variables $\xi_n$ is uniformly integrable as well, which in combination with (4.11) yields convergence in $L_1$. As a consequence, the second factor in (4.10) converges to zero. This implies that the second summand in (4.9) converges to zero uniformly in $n \in \mathbb{N}$. Altogether, by these and analogous arguments for the reverse case of (4.9), it follows that for $\varepsilon > 0$ and $n$ sufficiently large

$$\tilde{\mu}_{VX^{(n)}}(A) \le \mu_{VX^{(n)}}(A_\varepsilon) + \varepsilon \quad \text{and} \quad \mu_{VX^{(n)}}(A) \le \tilde{\mu}_{VX^{(n)}}(A_\varepsilon) + \varepsilon,$$

thus finishing the proof. □

Since the rest of the proof of Theorem 4.2.1 does not depend on the specific choice of $\alpha$ or the scaling of the $X^{(n)}$, the remainder of the proof of Theorem 4.2.2 can proceed in the very same way.

# Chapter 5

# Sharp large deviations on $\ell_p^n$-balls

Within this chapter we will move away from showing large deviation results in high-dimensional convex geometry and begin doing the same for *sharp* large deviation (SLD) results in the spirit of Bahadur and Ranga Rao [10]. The differences between these two closely related areas of study, especially regarding their goals and methodology, will be outlined in Section 5.1.

We then turn to showing concrete sharp large deviation results within $\ell_p^n$-balls. Section 5.2 provides such results for the $q$-norm of random vectors distributed in the $\ell_p^n$-ball $\mathbb{B}_p^n$ according to the cone probability measure $\mathbf{C}_{n,p}$ or the uniform distribution $\mathbf{U}_{n,p}$ for $1 \leq q < p < \infty$. As will be explained therein, the regime $1 \leq p < q < \infty$ cannot be handeled by our approach, as certain exponential moment conditions are no longer met in this case. In Section 5.3 the $p$-generalized arithmetic-geometric mean ($p$-AGM) inequality for vectors chosen randomly from the $\ell_p^n$-ball in $\mathbb{R}^n$ is considered and sharpened by showing a Bahadur-Ranga Rao-type result for the underlying probabilistic representation of the random vectors from $\mathbb{B}_p^n$.

## 5.1 Preliminaries

We have seen in Section 2.3.1 that classic LDPs give an idea of the asymptotic deviation behaviour of a sequence of probability distributions on a logarithmic scale, that is, for a sequence of probability measures $(\mathbf{P}_n)_{n\in\mathbb{N}}$ on a Polish space $\mathbb{X}$ an LDP allows to describe the limiting behaviour of

$$\frac{1}{s(n)} \log \mathbf{P}_n(\cdot)$$

via the rate function $\mathcal{I}(\cdot)$, with $s(\cdot)$ being the speed. Hence, we can write

$$\mathbf{P}_n(G) = e^{-s(n)\left[\inf_{x\in G} \mathcal{I}(x) + o(1)\right]},$$

where, for simplicity, we assumed $G \in \mathcal{B}(\mathbb{R}^n)$ to be a set with

$$\inf_{x\in G^\circ} \mathcal{I}(x) = \inf_{x\in \overline{G}} \mathcal{I}(x).$$

By considering these *logarithmic* probabilities, however, a lot of subtleties of the underlying distributions can be drowned out. Many small- and medium-scale properties of a given sequence of distributions can be missed in the asymptotic analysis of LDPs, since they either disappear for very large $n \in \mathbb{N}$ or are overshadowed by other, more significant phenomena of the distribution. This is clear when considering the error of concrete probability estimates based on large deviation results, given by $e^{-s(n)o(1)}$, for which we cannot even say whether or not it goes to zero in $n \in \mathbb{N}$, as we generally do not know the relative behaviour of $s(n)$ and $o(1)$. Thus, one is also interested in considering large deviations on a *non-logarithmic* scale, which we refer to as "sharp" large deviations (also called "precise" or "strong" large deviations in the literature).

One of the first results in this regard was given by Bahadur and Ranga Rao in [10]. They showed that for a sequence $(X^{(n)})_{n\in\mathbb{N}}$ of non-lattice i.i.d. random variables and any $z > \mathbb{E}[X^{(n)}]$ with $\Lambda_X^*(z) < \infty$ it holds for the sequence of empirical averages $\left(\frac{1}{n}S^{(n)}\right)_{n\in\mathbb{N}}$ that

$$\mathbb{P}\left(\frac{1}{n}\, S^{(n)} > z\right) = \frac{1}{\sqrt{2\pi n}\,\kappa(z)\xi(z)}\, e^{-n\Lambda_X^*(z)}\,(1 + o(1)), \tag{5.1}$$

where $\kappa(z)$ and $\xi(z)$ are prefactor functions which are only dependent on the distribution of the $X^{(n)}$ and the deviation size $z$. Specifically, they are given by functions of the first and second derivative of $\Lambda_X^*$, which, as the Legendre-Fenchel transform of the cumulant generating function $\Lambda_X$, is very dependent on the distribution of the $X^{(n)}$.

This result is a significant improvement on the theorem of Cramér, which in comparison states that in the same setting as previously

$$\mathbb{P}\left(\frac{1}{n}\, S^{(n)} > z\right) = e^{-s(n)[\Lambda_X^*(z)+o(1)]}.$$

However, showing sharp large deviation results has proven substantially more difficult, even for functionals of empirical averages, which for LDPs can often be handled in straight-forward fashion via the theorem of Cramér and the contraction principle. Let us take a closer look at how the methodology of the proofs need to be adapted.

The result of Bahadur and Ranga Rao is proven via a (somewhat implicit) application of the the so-called *saddle point method* (also called method of steepest descent or method of stationary phase), which was established by Debye [28], and brought to the realm of probability by Esscher [33] and Daniels [26]. The saddle point method generalizes Laplace's method for integral approximation to the complex plane, and is therefore highly useful when dealing with integrals over characteristic functions.

In general, for analytic functions $f, g$ and $n \in \mathbb{N}$ large, the saddle point method gives a way to approximate Laplace-type integrals

$$\int_P g(z)\, e^{nf(z)}\, \mathrm{d}z$$

along complex paths $P$ by deforming the path of integration using Cauchy's theorem (in the homotopic interpretation as in [84, Theorem 5.1]), into some $\tilde{P}$ that passes through a critical point of $f$, around which the mass of the reformulated integral then heavily concentrates. This is done such that standard integral approximation methods can be used to great effect. Thus, the path of integration should be chosen such that the integrand is consistently small along most of the path and then take the steepest route through said critical point. It follows from the Cauchy-Riemann equations that this point is always a saddle point, since the second derivatives of an analytical function in any critical point have opposite signs, hence the name of the method. Like in the classical Laplace-type approach, one would expect such a critical point to be attained where the magnitude $\mathrm{Re}(f(z))$ maximizes, but also where the phase $\mathrm{Im}(f(z))$ is constant, such that the oscillating components of the integral do not cancel out. And indeed, due to the Cauchy-Riemann equations, it again follows that the path of stationary phase and the path of steepest descent coincide. For large $n \in \mathbb{N}$ the integral can then be approximated well by only considering it locally around the saddle point. Also, since the phase $\mathrm{Im}(f(z))$ along the deformed path of integration is constant, the remaining integral can be approached using the standard Laplace principle.

In the realm of probability, this has been used for both tail probabilities (e.g., Esscher [33], Cramér [25]) and densities of random variables (e.g., Daniels [26], Richter [101, 102]), by writing them as an integral over their characteristic functions, using the Fourier inversion formula (see [110, Theorem 1.9]), and then approximating those integrals via the use of a complex saddle point.

We say that this was used "somewhat implicitly" in certain results, such as those of Esscher [33], Cramér [25], and Bahadur and Ranga Rao [10], since the technique used therein, which is an Edgeworth expansion in conjunction with a certain change of measure, often called *exponential tilting* or Esscher/Cramér transform, under the surface employs saddle points as well. Edgeworth expansion approximates an unknown density via a series expansion of its (unknown) characteristic function by using known (Gaussian) characteristic functions and then applying the Fourier inversion theorem, so we see how the saddle point method is employed here as well.

As pointed out by Richter in [101], the condition of finite exponential moments in a neighbourhood of the origin in all results that implicitly or explicitly use the saddle point method just amounts to assuming that the moment and cumulant generating functions are analytical in an open strip $\{z \in \mathbb{C} : |\operatorname{Re}(z)| < r\}$ of the complex plane containing the origin, so the aforementioned arguments can be applied.

Let us briefly walk through the application of the saddle point method to derive an asymptotic density estimate for the classical example of the empirical average of i.i.d. random vectors: Let $X_1, ..., X_n \in \mathbb{R}^d$ be i.i.d. random vectors with moment generating function $\varphi_X$, cumulant generating function $\Lambda_X$, and denote by $f_{S^{(n)}}$ the unknown density of their empirical average $\frac{1}{n} S^{(n)}$. Assume its Fourier transform $\mathcal{F}(f_{S^{(n)}})$ to be known and sufficiently integrable, i.e., $\mathcal{F}(f_{S^{(n)}}) \in L_1(\mathbb{R}^d)$. Then the Fourier inversion theorem lets us write the density $f_{S^{(n)}}$ of $S^{(n)}$ for some $x \in \operatorname{Dom}(\Lambda_X^*)^\circ$ as

$$
\begin{aligned}
f_{S^{(n)}}(x) &= \left(\frac{n}{2\pi}\right)^d \int_{-\infty}^{+\infty} \mathcal{F}(f_{S^{(n)}})(t) e^{-\langle it, nx\rangle} \,\mathrm{d}t \\
&= \left(\frac{n}{2\pi}\right)^d \int_{-\infty}^{+\infty} \varphi_X(it)^n e^{-\langle it, nx\rangle} \,\mathrm{d}t \\
&= \left(\frac{n}{2\pi}\right)^d \int_{-\infty}^{+\infty} e^{n[\Lambda_X(it) - \langle it, x\rangle]} \,\mathrm{d}t.
\end{aligned}
$$

where we rewrote the Fourier transform via the moment generating function $\varphi_X$, as outlined in Remark 2.2.2. As discussed above, we assume that $\varphi_X$ is finite around the origin in order for the moment and cumulant generating functions to be analytical in an open strip $\{z \in \mathbb{C} : |\operatorname{Re}(z)| < r\}$ of the complex plane containing the origin. For the same reason we restrict ourselves to arguments $x \in \operatorname{Dom}(\Lambda_X^*)^\circ$. Thus, Cauchys integral theorem allows for a change of the path of integration, such that

$$f_{S^{(n)}}(x) = \Big(\frac{n}{2\pi}\Big)^d \int_{-\infty}^{+\infty} e^{n[\Lambda_X(\tau+it)-\langle(\tau+it),x\rangle]}\,\mathrm{d}t,$$

for $\tau \in \mathbb{R}^d$ such that $\Lambda_X(\tau+it) < \infty$. By some standard arguments for the cumulant generating function and complex integrals (see, e.g., Daniels [26, Section 2]), one can show that $[\Lambda_X(\tau+it) - \langle(\tau+it),x\rangle]$ is a convex function in $\tau \in \mathrm{Dom}(\Lambda_X)$, attaining its minimum in $\tau$ at some $\tau(x) \in \mathrm{Dom}(\Lambda_X)$, and a concave function in $t \in \mathbb{R}$, attaining its maximum in $t$ at $t = 0$, with the modulus of the integrand itself also attaining its maximum there. Note that this $\tau(x)$ minimizing $[\Lambda_X(\tau+it) - \langle(\tau+it),x\rangle]$ is the same as that from Lemma 2.2.1 (4) for which $\Lambda_X^*(x) = \langle(\tau(x)),x\rangle - \Lambda_X(\tau(x))$. Thus, the point $(\tau(x)+i0) \in \mathrm{Dom}(\Lambda_X) \times i\mathbb{R}$ is the saddle point of the exponent in the integrand and we can write

$$f_{S^{(n)}}(x) = \Big(\frac{n}{2\pi}\Big)^d \int_{-\infty}^{+\infty} e^{-n[\langle(\tau(x)+it),x\rangle-\Lambda_X(\tau(x)+it)]}\,\mathrm{d}t. \tag{5.2}$$

Thus, for large $n \in \mathbb{N}$, most of the mass in the integral in (5.2) concentrates in a neighbourhood around $\tau(x)$, and hence, appropriate expansions of the integrand around $\tau(x)$ yield good approximations to the integral (5.2) and thereby $f_{S^{(n)}}$. By local approximation as in [7, Theorem 3.1] it then follows that

$$\begin{aligned} f_{S^{(n)}}(x) &= \Big(\frac{n}{2\pi}\Big)^d \big[\det \mathcal{H}_\tau(\Lambda_X(\tau(x))\big]^{-1/2} e^{n[\Lambda_X(\tau(x))-\langle\tau(x),x\rangle]}\,(1+o(1)) \\ &= \Big(\frac{n}{2\pi}\Big)^d \big[\det \mathcal{H}_\tau(\Lambda_X(\tau(x)))\big]^{-1/2} e^{-n\Lambda_X^*(x)}\,(1+o(1)). \end{aligned}$$

Versions of the saddle point method were used for density approximations of empirical averages of both sequences of i.i.d. random variables (e.g., Daniels [26], Richter [101, 102], Borovkov and Rogozin [17] etc.) and arbitrary sequences of random variables (e.g., Chaganty and Sethuraman [21, 22, 23, 24], Joutard [57, 58] etc.), so this procedure is not limited to the case of considering i.i.d. random variables.

We previously mentioned how many results from (sharp) large deviations theory implicitly employ the saddle point method in the form of so-called exponential tilting. Let us consider how this works in more detail as well. The accuracy of classic results for the approximation of probabilities and densities severly detereorate in the tails of a distribution, as can be seen when trying to estimate large deviation probabilities using, e.g., CLT results. Such results are accurate around the expectation, but worsen considerably in the tails. Hence, the idea of exponential tilting is to construct a new distribution for a given large deviation event, under which this event becomes typical and classic approximation techniques can again be applied.

We again consider the previous example of trying to give a good estimate for the density $f_{S^{(n)}}$ of the empirical average $\frac{1}{n}S^{(n)}$ of i.i.d. random vectors $X_1, \dots, X_n$ in $\mathbb{R}^d$. For a given $x \in \mathrm{Dom}(\Lambda_X^*)^\circ$ we define a family of exponentially tilted densities (often also called "conjugate densities") as

$$f_{S^{(n)},\tau}(x) := e^{n[\langle \tau, x\rangle - \Lambda_X(\tau)]} f_{S^{(n)}}(y),$$

where $\tau$ is in the effective domain $\mathrm{Dom}(\Lambda_X)$ of $\Lambda_X$. Hence, we can write $f_{S^{(n)}}$ via this exponentially tilted density as

$$f_{S^{(n)}}(x) := e^{-n[\langle \tau, x\rangle - \Lambda_X(\tau)]} f_{S^{(n)},\tau}(x).$$

Choosing $\tau = \tau(x) \in \mathrm{Dom}(\Lambda_X)^\circ$ from Lemma 2.2.1 (4) yields a distribution centered around $x \in \mathrm{Dom}(\Lambda_X^*)^\circ$, and the approximation of $f_{S^{(n)}}(x)$ may now be obtained efficiently by approximating $f_{S^{(n)},\tau(x)}(x)$ via various techniques such as Edgeworth expansion, which, as we mentioned, employ the saddle point method as well.

## 5.2 Sharp large deviations for $q$-norms of $\ell_p^n$-balls

In this section sharp large deviation results in the spirit of Bahadur and Ranga Rao as in (5.1) are provided for the $q$-norm of random vectors distributed in the $\ell_p^n$-ball $\mathbb{B}_p^n$ according to $\mathbf{C}_{n,p}$ or $\mathbf{U}_{n,p}$ for $1 \le q < p < \infty$. As stated in Section 1.2, there is a close connection between the behaviour of $q$-norms of random vectors in $\ell_p^n$-balls and the intersection volumes of $t$-multiples of volume-normalized $\ell_p^n$-balls $\mathrm{vol}_n(\mathbb{D}_p^n \cap t\mathbb{D}_q^n)$ with $t \in [0, \infty)$. We will use this connection to derive sharp asymptotics for said intersection volumes at an improved rate compared to those provided by previous results. Lastly, these sharp large deviation results will be applied to retain sharp asymptotics for the length of the projection of an $\ell_p^n$-ball onto a line with uniform random direction.

The proof of the main results is separated into three steps. Using the results of Schechtman and Zinn in Proposition 2.4.2, we first rewrite the target deviation probability of the $q$-norm of a random vector from $\mathbb{B}_p^n$ as the probability of the empirical average of some vector of $p$-generalized Gaussians lying in some domain, which we will refer to as the *deviation area*. The geometric shape of this deviation area will turn out to have a major influence on the main result. Hence, we can write the target deviation probability as an integral of the (unknown) density of this empirical average over the deviation area. The second step then consists of deriving this density explicitly using the saddle point method (in the implicit form of exponential tilting). Thus, the deviation probability can be written as the integral of the density approximation over the deviation area. In the third and final step we then calculate this integral concretely for $\ell_p^n$-spheres and $\ell_p^n$-balls. To do so, we will utilize some geometric results for asymptotic expansions of Laplace integrals from Andriani and Baldi [7] and Breitung and Hohenbichler [19].

### 5.2.1 LDPs for $q$-norms of $\ell_p^n$-balls

We start by laying out the target variables and presenting the relevant results that were established for them previously. Throughout Section 5.2 we assume $1 \leq q < p < \infty$, even if the results we present therein hold for $q > p$ as well, since this section's main results will only hold in this particular case. The main variables of interest will be the $q$-norms of the random vectors

$$Z^{(n)} \sim \mathbf{C}_{n,p} \qquad \text{and} \qquad \mathscr{Z}^{(n)} \sim \mathbf{U}_{n,p}.$$

Note, that within this section we will denote quantities related to $\mathscr{Z}^{(n)} \sim \mathbf{U}_{n,p}$ cursively. To get non-trivial results, our target variables also need to be appropriately rescaled. Thus, for random vectors $Z^{(n)}, \mathscr{Z}^{(n)} \in \mathbb{B}_p^n$ with $Z^{(n)} \sim \mathbf{C}_{n,p}$ and $\mathscr{Z}^{(n)} \sim \mathbf{U}_{n,p}$, our target variables will be

$$n^{1/p\,-\,1/q}\,\|Z^{(n)}\|_q \qquad \text{and} \qquad n^{1/p\,-\,1/q}\,\|\mathscr{Z}^{(n)}\|_q,$$

and we set

$$\|Z\| := \left(n^{1/p\,-\,1/q}\,\|Z^{(n)}\|_q\right)_{n\in\mathbb{N}} \qquad \text{and} \qquad \|\mathscr{Z}\| := \left(n^{1/p\,-\,1/q}\,\|\mathscr{Z}^{(n)}\|_q\right)_{n\in\mathbb{N}}.$$

Applying the result of Schechtman and Zinn in Proposition 2.4.2, we get the following probabilistic representation for this section's target random variables: Let $(Y^{(n)})_{n\in\mathbb{N}}$ be a sequence of i.i.d. random vectors $Y^{(n)} := \left(Y_1^{(n)}, \ldots, Y_n^{(n)}\right)$ with $Y_i^{(n)} \sim \mathbf{N}_p$, and $U$ a random variable uniformly distributed on $[0,1]$ independent of the $Y_i^{(n)}$. Then

$$n^{1/p\,-\,1/q}\,\|Z^{(n)}\|_q \overset{\mathcal{D}}{=} n^{1/p\,-\,1/q}\,\frac{\left\|Y^{(n)}\right\|_q}{\left\|Y^{(n)}\right\|_p} = \frac{\left(\frac{1}{n}\sum\limits_{i=1}^n \left|Y_i^{(n)}\right|^q\right)^{1/q}}{\left(\frac{1}{n}\sum\limits_{i=1}^n \left|Y_i^{(n)}\right|^p\right)^{1/p}}, \tag{5.3}$$

and

$$n^{1/p\,-\,1/q}\,\|\mathscr{Z}^{(n)}\|_q \overset{\mathcal{D}}{=} n^{1/p\,-\,1/q}\,U^{1/n}\,\frac{\|Y^{(n)}\|_q}{\|Y^{(n)}\|_p} = U^{1/n}\,\frac{\left(\frac{1}{n}\sum\limits_{i=1}^n \left|Y_i^{(n)}\right|^q\right)^{1/q}}{\left(\frac{1}{n}\sum\limits_{i=1}^n \left|Y_i^{(n)}\right|^p\right)^{1/p}}. \tag{5.4}$$

Since the $p$-th absolute moment of a $p$-generalized Gaussian is one (see (2.7)), it follows via the strong law of large numbers and the continuous mapping theorem applied to the probabilistic representations in (5.3) and (5.4) that the expectations of $\|Z\|$ and $\|\mathscr{Z}\|$ converge in $n \in \mathbb{N}$ to

$$m_{p,q} := \mathbb{E}\left[\left|Y_1^{(n)}\right|^q\right]^{1/q} = \left[\frac{p^{q/p}}{q+1}\,\frac{\Gamma\left(1+\frac{q+1}{p}\right)}{\Gamma\left(1+\frac{1}{p}\right)}\right]^{1/q}. \tag{5.5}$$

For fixed $n \in \mathbb{N}$ we will denote

$$\mathbb{E}\left[n^{1/p-1/q}\left\|Z^{(n)}\right\|_q\right] =: m_{n,p,q} \qquad \text{and} \qquad \mathbb{E}\left[n^{1/p-1/q}\left\|\mathscr{Z}^{(n)}\right\|_q\right] =: \mathscr{m}_{n,p,q}.$$

For $\|Z\|$ and $\|\mathscr{Z}\|$ LDPs have been given in previous works, which we will present here explicitly. But first, let us consider some auxiliary probabilistic representations used both in the proofs of those LDPs and our sharp large deviations results. Define

$$V^{(n)} := \left(V_1^{(n)},\dots,V_n^{(n)}\right) \in \mathbb{R}^{2n} \qquad \text{with} \qquad V_i^{(n)} := \left(\left|Y_i^{(n)}\right|^q, \left|Y_i^{(n)}\right|^p\right), \tag{5.6}$$

and

$$\mathscr{V}^{(n)} := \left(\mathscr{V}_1^{(n)},\dots,\mathscr{V}_n^{(n)}\right) \in \mathbb{R}^{3n} \quad \text{with} \quad \mathscr{V}_i^{(n)} := \left(\left|Y_i^{(n)}\right|^q, \left|Y_i^{(n)}\right|^p, U^{1/n}\right). \tag{5.7}$$

We denote the moment and cumulant generating function of the $V_i^{(n)}$ respectively as

$$\varphi_p(\tau) := \int_{\mathbb{R}} e^{\tau_1|y|^q+\tau_2|y|^p} f_{\mathbf{N}_p}(y)\,\mathrm{d}y \quad \text{and} \quad \Lambda_p(\tau) := \log\int_{\mathbb{R}} e^{\tau_1|y|^q+\tau_2|y|^p} f_{\mathbf{N}_p}(y)\,\mathrm{d}y, \tag{5.8}$$

for $\tau = (\tau_1,\tau_2) \in \mathbb{R}^2$ and by $\Lambda_p^*$ the Legendre-Fenchel transform of $\Lambda_p$ as in (2.2). Since $q < p$, for the integral in both $\varphi_p$ and $\Lambda_p$ to be finite, the sign of the dominant term $|y|^p$ in the exponent must be negative. Recalling the definition of $f_{\mathbf{N}_p}$, one can see that this is given for $\tau_2 < \frac{1}{p}$, thus $\mathrm{Dom}(\Lambda_p) = \mathbb{R} \times \big(-\infty, \frac{1}{p}\big)$. Note that for $q > p$ we no longer have finite exponential moments in a neighbourhood of the origin, which is why we set the condition $1 \le q < p < \infty$. By Lemma 2.2.1 we have $\mathrm{Dom}(\Lambda_p^*)^\circ = \nabla_\tau\Lambda_p\big(\mathrm{Dom}(\Lambda_p)^\circ\big)$ and for every $x \in \mathrm{Dom}(\Lambda_p^*)^\circ$ there exists a unique $\tau(x) \in \mathrm{Dom}(\Lambda_p)^\circ$ such that $\Lambda_p^*(x) = \langle x, \tau(x)\rangle - \Lambda_p(\tau(x))$. Since $\mathrm{Dom}(\Lambda_p) = \mathbb{R} \times \big(-\infty, \frac{1}{p}\big)$ is open and, by Lemma 2.2.1 (2), the derivative $\nabla_\tau\Lambda_p$ is continuous, we even have that

$$\mathrm{Dom}(\Lambda_p^*) = \mathrm{Dom}(\Lambda_p)^\circ = \nabla_\tau\Lambda_p\big(\mathrm{Dom}(\Lambda_p)^\circ\big) = \nabla_\tau\Lambda_p\big(\mathrm{Dom}(\Lambda_p)\big). \tag{5.9}$$

Furthermore, for $x \in \mathrm{Dom}(\Lambda_p^*)$, set

$$\mathfrak{H}_x := \mathcal{H}_\tau\Lambda_p(\tau(x)) \tag{5.10}$$

to be the Hessian of $\Lambda_p(\tau)$ in $\tau \in \mathbb{R}^2$, evaluated at $\tau(x)$ as in Lemma 2.2.1 (4).

**Lemma 5.2.1** *It holds that*

*i)* $\nabla_x \Lambda_p^*(x) = \tau(x)$,

*ii)* $\mathcal{H}_x \Lambda_p^*(x) = {\mathfrak{H}_x}^{-1}$.

*Proof.* By the definition of $\tau(x)$ in Lemma 2.2.1 (4) it follows that $x - \nabla_\tau \Lambda_p(\tau(x)) = 0$. Hence,

$$\begin{aligned}
\nabla_x \Lambda_p^*(x) &= \nabla_x \big[\langle x, \tau(x)\rangle - \Lambda_p(\tau(x))\big] \\
&= \tau(x) + J_x \tau(x)\, x - J_x \tau(x) \nabla_\tau \Lambda_p(\tau(x)) \\
&= \tau(x) + J_x \tau(x)\, \big[x - \nabla_\tau \Lambda_p(\tau(x))\big] \\
&= \tau(x).
\end{aligned}$$

Let us now prove that $\mathcal{H}_x \Lambda_p^*(x) = {\mathfrak{H}_x}^{-1}$. On the one hand, it follows from the above

$$\mathcal{H}_x \Lambda_p^*(x) = J_x \tau(x), \tag{5.11}$$

while on the other hand, it holds that

$$\begin{aligned}
\mathcal{H}_x \Lambda_p^*(x) &= \mathcal{H}_x \big[\langle x, \tau(x)\rangle - \Lambda_p(\tau(x))\big] \\
&= \mathcal{H}_x \big[\langle x, \tau(x)\rangle\big] - \mathcal{H}_x \big[\Lambda_p(\tau(x))\big] \\
&= J_x \big[\nabla_x \langle x, \tau(x)\rangle\big] - J_x \big[\nabla_x \Lambda_p(\tau(x))\big] \\
&= J_x \big[\tau(x) + J_x \tau(x) x\big] - J_x \big[J_x \tau(x)\, \nabla_\tau \Lambda_p(\tau(x))\big] \\
&= J_x \tau(x) + J_x \big[J_x \tau(x) x\big] - \mathcal{H}_x \tau(x)\, \nabla_\tau \Lambda_p(\tau(x)) - J_x \tau(x) J_x \big[\nabla_\tau \Lambda_p(\tau(x))\big] \\
&= 2 J_x \tau(x) + \mathcal{H}_x \tau(x) \big[x - \nabla_\tau \Lambda_p(\tau(x))\big] - J_x \tau(x)\, J_x \tau(x) \mathcal{H}_\tau \Lambda_p(\tau(x)) \\
&= 2 J_x \tau(x) - J_x \tau(x)\, J_x \tau(x)\, \mathcal{H}_\tau \Lambda_p(\tau(x)).
\end{aligned} \tag{5.12}$$

Equating the terms (5.11) and (5.12) yields

$$\begin{aligned}
J_x \tau(x) &= 2 J_x \tau(x) - J_x \tau(x)\, J_x \tau(x)\, \mathcal{H}_\tau \Lambda_p(\tau(x)) \\
\Leftrightarrow \quad 0 &= J_x \tau(x) - J_x \tau(x)\, J_x \tau(x)\, \mathcal{H}_\tau \Lambda_p(\tau(x)) \\
\Leftrightarrow \quad 0 &= I_2 - J_x \tau(x)\, \mathcal{H}_\tau \Lambda_p(\tau(x)) \\
\Leftrightarrow \quad J_x \tau(x) &= \mathcal{H}_\tau \Lambda_p(\tau(x))^{-1},
\end{aligned}$$

where $I_2$ again denotes the identity matrix in $\mathbb{R}^2$. Again applying (5.11) then gives

$$\mathcal{H}_x \Lambda_p^*(x) = J_x \tau(x) = \mathcal{H}_\tau \Lambda_p(\tau(x))^{-1} = \mathfrak{H}_x^{-1},$$

and thereby finishes the proof. $\square$

For the sequence $\|Z\|$ the following LDP has already been shown by Kabluchko, Prochno, and Thäle [61, Section 5.1]:

**Proposition 5.2.2** *Let $1 \leq q < p < \infty$ and $Z^{(n)} \sim \mathbf{C}_{n,p}$ be a random vector in $\mathbb{S}_p^{n-1}$. Then the sequence $\left(n^{1/p-1/q}\,\|Z^{(n)}\|_q\right)_{n\in\mathbb{N}}$ satisfies an LDP with speed $n$ and good rate function*

$$\mathcal{I}_{\|Z\|}(z) := \begin{cases} \inf\limits_{\substack{t_1,\,t_2\,>\,0\\ t_1^{1/q}t_2^{-1/p}\,=\,z}} \Lambda_p^*(t_1,t_2) & : z > 0 \\ +\infty & : z \leq 0. \end{cases}$$

**Remark 5.2.3** In the case $1 \leq q < p < \infty$ this is proven in [61] in the classical way established by Gantert, Kim and Ramanan [36]: First, give a probabilistic representation for the target random variables, which works out to be a function of empirical averages of the coordinates of the $V_i^{(n)}$ as in (5.3). Second, use the theorem of Cramér from Proposition 2.3.3 to establish an auxiliary LDP for those empirical averages of the $V_i^{(n)}$ with speed $n$ and rate function $\Lambda_p^*$. Third, map this auxiliary LDP to the target sequence using the contraction principle from 2.3.5 for the function $F(t_1,t_2) = t_1^{1/q}t_2^{-1/p}$, yielding the infimum operator in the above rate function. So we see that the functional structure of the representation in (5.3) has a direct influence on the LDP in the form of the infimum of the rate function over the level sets of the "transporting function" $F$. Since for sharp large deviations we do not have access to the contraction principle, this influence will be quite a bit more subtle, as we will see.

In [85, Lemma 2.1, Appendix A] Liao and Ramanan established a simplification of a similar rate function in a different setting by calculating the infimum in the rate function explicitly. Their arguments can be analogously applied in our setting to derive the following result:

**Lemma 5.2.4** *Let $z > m_{p,q}$ such that $z^* := (z^q, 1) \in \mathrm{Dom}(\Lambda_p^*)$. Then*

$$\mathcal{I}_{\|Z\|}(z) = \inf_{\substack{y_1,\,y_2\,>\,0\\ y_1^{1/q}y_2^{-1/p}\,=\,z}} \Lambda_p^*(y_1,y_2) = \Lambda_p^*(z^*),$$

*with $z^*$ being the unique point at which $\Lambda_p^*$ attains its infimum under the above conditions.*

To keep this section self-contained, we will present the analogous proof of this as well.

*Proof.* Let $z > m_{p,q}$ such that $z^* = (z^q, 1) \in \mathrm{Dom}(\Lambda_p^*)$. Then it holds that

$$\mathcal{I}_{\|Z\|}(z) = \inf_{\substack{y_1,\, y_2 \,>\, 0 \\ y_1^{1/q} y_2^{-1/p} \,=\, z}} \Lambda_p^*(y_1, y_2) = \inf_{x_1, x_2 > 0 \colon x_1 = z\, x_2} \Lambda_p^*(x_1^q, x_2^p) = \inf_{x_2 > 0} \Lambda_p^*(z^q {x_2}^q, {x_2}^p).$$

We set $x_z := (z^q x_2^q, x_2^p)$, then with Lemma 2.2.1 it follows that

$$\mathcal{I}_{\|Z\|}(z) = \inf_{x_2 > 0} \sup_{\tau \in \mathbb{R}^2} [\langle \tau, x_z \rangle - \Lambda_p(\tau)] = \inf_{x_2 > 0} \Big[ \langle \tau(x_z), x_z \rangle - \Lambda_p(\tau(x_z)) \Big].$$

Our goal is to show that the infimum is attained at $x_z^* := z^*$, i.e., at $x_2 = 1$. For the function

$$g_x(\tau) := \langle \tau, x \rangle - \Lambda_p(\tau), \qquad \tau \in \mathbb{R}^2,$$

with $x \in \mathrm{Dom}(\Lambda_p^*)$, Lemma 2.2.1 implies it attains its supremum at $\tau(x)$. Hence, it holds for $x = x_z$ that

$$\nabla_\tau\, g_{x_z}(\tau(x_z)) = x_z - \nabla_\tau \Lambda_p(\tau(x_z)) = 0,$$

which gives

$$x_z = (z^q x_2^q, x_2^p) = \left( \frac{\partial}{\partial \tau_1} \Lambda_p(\tau(x_z)), \frac{\partial}{\partial \tau_2} \Lambda_p(\tau(x_z)) \right). \tag{5.13}$$

We now aim to write $\frac{\partial}{\partial \tau_2} \Lambda_p(\tau)$ with respect to $\frac{\partial}{\partial \tau_1} \Lambda_p(\tau)$ and then use the above equation. To do so, we first want to reformulate $\Lambda_p$ along the lines of [36, Lemma 5.7]. It holds that

$$\begin{aligned} \Lambda_p(\tau) &= \log \int_{\mathbb{R}} e^{\tau_1 |y|^q + \tau_2 |y|^p} f_{\mathbf{N}_p}(y)\, \mathrm{d}y \\ &= \log \left( \frac{1}{2\, p^{1/p} \Gamma\Big(1 + \frac{1}{p}\Big)} \int_{\mathbb{R}} e^{\tau_1 |y|^q - \frac{1}{p}(1 - p\tau_2)|y|^p}\, \mathrm{d}y \right). \end{aligned}$$

The change of variable $\tilde{y} = (1 - p\tau_2)^{1/p} y$ then gives

$$\begin{aligned} \Lambda_p(\tau) &= \log \left( (1 - p\tau_2)^{-1/p} \int_{\mathbb{R}} e^{\frac{\tau_1}{(1 - p\tau_2)^{q/p}} |\tilde{y}|^q} f_{\mathbf{N}_p}(\tilde{y})\, \mathrm{d}\tilde{y} \right) \\ &= -\frac{1}{p} \log(1 - p\tau_2) + \log \varphi_{|Y|^q} \left( \frac{\tau_1}{(1 - p\tau_2)^{q/p}} \right), \end{aligned}$$

where $\varphi_{|Y|^q}$ is the moment generating function of a random variable $|Y|^q$ with $Y \sim \mathbf{N}_p$.

Hence,

$$\begin{aligned}
\frac{\partial}{\partial\tau_1}\Lambda_p(\tau) &= \frac{\partial}{\partial\tau_1}\left[\log\varphi_{|Y|^q}\left(\frac{\tau_1}{(1-p\tau_2)^{q/p}}\right)\right] \\
&= \varphi_{|Y|^q}\left(\frac{\tau_1}{(1-p\tau_2)^{q/p}}\right)^{-1}\frac{\partial}{\partial\tau_1}\left[\varphi_{|Y|^q}\left(\frac{\tau_1}{(1-p\tau_2)^{q/p}}\right)\right] \\
&= \varphi_{|Y|^q}\left(\frac{\tau_1}{(1-p\tau_2)^{q/p}}\right)^{-1}\int_{\mathbb{R}}(1-p\tau_2)^{-q/p}\,|y|^q\,e^{\frac{\tau_1}{(1-p\tau_2)^{q/p}}|y|^q}f_{\mathbf{N}_p}(y)\,\mathrm{d}y \\
&= (1-p\tau_2)^{-q/p}\varphi_{|Y|^q}\left(\frac{\tau_1}{(1-p\tau_2)^{q/p}}\right)^{-1}\varphi'_{|Y|^q}\left(\frac{\tau_1}{(1-p\tau_2)^{q/p}}\right),
\end{aligned}$$

where $\varphi'_{|Y|^q}\left(\frac{\tau_1}{(1-p\tau_2)^{q/p}}\right)=\frac{\partial}{\partial\tau_1}\,\varphi_{|Y|^q}\left(\frac{\tau_1}{(1-p\tau_2)^{q/p}}\right)$. Moreover, with the above we get that

$$\begin{aligned}
&\frac{\partial}{\partial\tau_2}\Lambda_p(\tau) \\
&= (1-p\tau_2)^{-1}+\frac{\partial}{\partial\tau_2}\left[\log\varphi_{|Y|^q}\left(\frac{\tau_1}{(1-p\tau_2)^{q/p}}\right)\right] \\
&= (1-p\tau_2)^{-1}+\varphi_{|Y|^q}\left(\frac{\tau_1}{(1-p\tau_2)^{q/p}}\right)^{-1}\frac{\partial}{\partial\tau_2}\left[\varphi_{|Y|^q}\left(\frac{\tau_1}{(1-p\tau_2)^{q/p}}\right)\right] \\
&= (1-p\tau_2)^{-1}+\varphi_{|Y|^q}\left(\frac{\tau_1}{(1-p\tau_2)^{q/p}}\right)^{-1}\int_{\mathbb{R}}\frac{q\tau_1}{(1-p\tau_2)^{(q+p)/p}}\,|y|^q\,e^{\frac{\tau_1}{(1-p\tau_2)^{q/p}}|y|^q}f_{\mathbf{N}_p}(y)\,\mathrm{d}y \\
&= (1-p\tau_2)^{-1}+\frac{q\tau_1}{(1-p\tau_2)^{(q+p)/p}}\,\varphi_{|Y|^q}\left(\frac{\tau_1}{(1-p\tau_2)^{q/p}}\right)^{-1}\varphi'_{|Y|^q}\left(\frac{\tau_1}{(1-p\tau_2)^{q/p}}\right) \\
&= (1-p\tau_2)^{-1}+\frac{q\tau_1}{1-p\tau_2}\,(1-p\tau_2)^{-q/p}\,\varphi_{|Y|^q}\left(\frac{\tau_1}{(1-p\tau_2)^{q/p}}\right)^{-1}\varphi'_{|Y|^q}\left(\frac{\tau_1}{(1-p\tau_2)^{q/p}}\right) \\
&= (1-p\tau_2)^{-1}+q\tau_1(1-p\tau_2)^{-1}\frac{\partial}{\partial\tau_1}\Lambda_p(\tau). \qquad (5.14)
\end{aligned}$$

Plugging in the identities from (5.13) into (5.14) it follows for $(\tau_1,\tau_2)=\big(\tau(x_z)_1,\tau(x_z)_2\big)$:

$$x_2^p=(1-p\tau(x_z)_2)^{-1}+q\tau(x_z)_1(1-p\tau(x_z)_2)^{-1}\,z^q x_2^q. \qquad (5.15)$$

Using this, we can calculate the derivative of $\Lambda_p^*(x_z)$ in $x$ (we write $x$ instead of $x_2$ for notational brevity), where $\tau(x_z)$ is considered as a function in $x$ as well. It holds that

$$\begin{aligned}
\frac{\partial}{\partial x}\Lambda_p^*(x_z) &= \frac{\partial}{\partial x}\Lambda_p^*(z^q x^q, x^p) \\
&= \frac{\partial}{\partial x}\left[\langle x_z, \tau(x_z)\rangle - \Lambda_p(\tau(x_z))\right] \\
&= \frac{\partial}{\partial x}\left[z^q x^q \tau(x_z)_1 + x^p \tau(x_z)_2 - \Lambda_p(\tau(x_z))\right] \\
&= z^q q x^{q-1}\tau(x_z)_1 + z^q x^q \frac{\partial}{\partial x}\tau(x_z)_1 + p x^{p-1}\tau(x_z)_2 + x^p \frac{\partial}{\partial x}\tau(x_z)_2 \\
&\quad -\frac{\partial}{\partial x}\Lambda_p(\tau(x_z)) \\
&= z^q q x^{q-1}\tau(x_z)_1 + z^q x^q \frac{\partial}{\partial x}\tau(x_z)_1 + p x^{p-1}\tau(x_z)_2 + x^p \frac{\partial}{\partial x}\tau(x_z)_2 \\
&\quad -J_x(\tau(x_z))\nabla_\tau \Lambda_p(\tau(x_z)) \\
&= z^q q x^{q-1}\tau(x_z)_1 + z^q x^q \frac{\partial}{\partial x}\tau(x_z)_1 + p x^{p-1}\tau(x_z)_2 + x^p \frac{\partial}{\partial x}\tau(x_z)_2 \\
&\quad -\frac{\partial}{\partial x}\tau(x_z)_1 \frac{\partial}{\partial \tau_1}\Lambda_p(\tau(x_z)) - \frac{\partial}{\partial x}\tau(x_z)_2 \frac{\partial}{\partial \tau_2}\Lambda_p(\tau(x_z)).
\end{aligned}$$

We now use the identity from (5.13), which yields

$$\begin{aligned}
\frac{\partial}{\partial x}\Lambda_p^*(x_z) &= z^q q x^{q-1}\tau(x_z)_1 + z^q x^q \frac{\partial}{\partial x}\tau(x_z)_1 + p x^{p-1}\tau(x_z)_2 \\
&\quad + x^p \frac{\partial}{\partial x}\tau(x_z)_2 - \frac{\partial}{\partial x}\tau(x_z)_1 z^q x^q - \frac{\partial}{\partial x}\tau(x_z)_2 x^p \\
&= z^q q x^{q-1}\tau(x_z)_1 + p x^{p-1}\tau(x_z)_2. \qquad (5.16)
\end{aligned}$$

Reformulating the identity in (5.15) yields

$$\begin{aligned}
& x^p = (1-p\tau(x_z)_2)^{-1} + q\tau(x_z)_1 (1-p\tau(x_z)_2)^{-1} z^q x^q \\
\Leftrightarrow\quad & (1-p\tau(x_z)_2)x^{p-1} - x^{-1} = z^q x^{q-1} q \tau(x_z)_1. \qquad (5.17)
\end{aligned}$$

Thus, if we set $\frac{\partial}{\partial x}\Lambda_p^*(x_z) = 0$, we get from (5.16) and (5.17) that

$$\begin{aligned}
\frac{\partial}{\partial x}\Lambda_p^*(x_z) = 0 \quad &\Leftrightarrow \quad 0 = z^q q x^{q-1}\tau(x_z)_1 + p x^{p-1}\tau(x_z)_2 \\
&\Leftrightarrow \quad 0 = (1-p\tau(x_z)_2)x^{p-1} - x^{-1} + p x^{p-1}\tau(x_z)_2 \\
&\Leftrightarrow \quad x = 1.
\end{aligned}$$

Hence, the infimum of $\Lambda_p^*$ under the conditions in the rate function from Proposition 5.2.2 is attained at $x_z^* = (z^q, 1) = z^*$. Since by Lemma 2.2.1 we know that $\Lambda_p^*$ is strictly convex on $\mathrm{Dom}(\Lambda_p^*)^\circ$ and $\mathrm{Dom}(\Lambda_p^*)^\circ = \mathrm{Dom}(\Lambda_p^*)$, this minimum is unique. Thereby, our claim is proven. □

For the sequence $\|\mathscr{Z}\|$ the following LDP was also provided by Kabluchko, Prochno, and Thäle in [61, Theorem 1.2]:

**Proposition 5.2.5** *Let $1 \le q < p < \infty$ and $\mathscr{Z}^{(n)} \sim \mathbf{U}_{n,p}$ be a random vector in $\mathbb{B}_p^n$. Then the sequence $\big(n^{1/p-1/q}\,\|\mathscr{Z}^{(n)}\|_q\big)_{n\in\mathbb{N}}$ satisfies an LDP with speed $n$ and good rate function*

$$\mathcal{I}_{\|\mathscr{Z}\|}(z) := \begin{cases} \inf\limits_{\substack{z = z_1 z_2 \\ z_1, z_2 > 0}} \big[\mathcal{I}_{\|Z\|}(z_1) + \mathcal{I}_U(z_2)\big] & : z > 0 \\ +\infty & : z \le 0, \end{cases}$$

*with $\mathcal{I}_{\|Z\|}$ as in Proposition 5.2.2 and*

$$\mathcal{I}_U(z_2) := \begin{cases} -\log(z_2) & : z_2 \in (0,1] \\ +\infty & : \textit{otherwise}. \end{cases}$$

We again show that the infimum in the rate function is attained at a unique point.

**Lemma 5.2.6** *Assume the same setting as in Proposition 5.2.5. For $z > m_{p,q}$, we can simplify the rate function by combining the two infimum operations to get*

$$\mathcal{I}_{\|\mathscr{Z}\|}(z) = \inf_{\substack{z = x_1^{1/q} x_2^{-1/p} x_3 \\ x_1, x_2 > 0,\, x_3 \in (0,1]}} \big[\Lambda_p^*(x_1, x_2) - \log(x_3)\big].$$

*We define*

$$\mathcal{I}_{\mathscr{S}}(x) := \Lambda_p^*(x_1, x_2) - \log(x_3), \qquad x_1, x_2 \in \mathbb{R},\ x_3 \in (0,1],$$

*and set $z^* := (z^q, 1) \in \mathbb{R}^2$, $z^{**} := (z^q, 1, 1) \in \mathbb{R}^3$. It then holds for $z > m_{p,q}$ with $z^* \in \mathrm{Dom}(\Lambda_p^*)$ that*

$$\mathcal{I}_{\|\mathscr{Z}\|}(z) = \mathcal{I}_{\mathscr{S}}(z^{**}) = \Lambda_p^*(z^*),$$

*with $z^{**}$ being the unique point at which $\mathcal{I}_{\mathscr{S}}$ attains its infimum under the above conditions.*

Thus, for $z > m_{p,q}$ with $z^* \in \mathrm{Dom}(\Lambda_p^*)$ both $\|Z\|$ and $\|\mathscr{Z}\|$ satisfy LDPs with the same speed and rate function.

*Proof.* Let $z > m_{p,q}$ such that $z^* = (z^q, 1) \in \mathrm{Dom}(\Lambda_p^*)$. Furthermore, set $z^{**} := (z^q, 1, 1)$ and $\mathcal{I}_{\mathcal{S}}(x) := \Lambda_p^*(x_1, x_2) - \log(x_3)$, $x \in \mathbb{R}^3$. We use the definitions of $\mathcal{I}_{\|Z\|}$ and $\mathcal{I}_U$, together with Lemma 5.2.4, to get that

$$\begin{aligned}
\mathcal{I}_{\|\mathscr{Z}\|}(z) &= \inf_{\substack{z = x_1^{1/q} x_2^{-1/p} x_3 \\ x_1, x_2 > 0,\, x_3 \in (0,1]}} \mathcal{I}_{\mathcal{S}}(x) \\
&= \inf_{\substack{z = z_1 z_2 \\ z_1 > 0,\, z_2 \in (0,1]}} \left[ \inf_{\substack{x_1, x_2 > 0 \\ x_1^{1/q} x_2^{-1/p} = z_1}} \Lambda_p^*(x_1, x_2) + \mathcal{I}_U(z_2) \right] \\
&= \inf_{\substack{z = z_1 z_2 \\ z_1 > 0,\, z_2 \in (0,1]}} \left[ \Lambda_p^*(z_1^q, 1) - \log(z_2) \right].
\end{aligned}$$

It follows by Lemma 2.2.1 (3) that $\mathcal{I}_{\|Z\|}(z) = \Lambda_p^*(z^q, 1)$ is strictly convex in $z$ on $\mathrm{Dom}(\Lambda_p^*)^\circ = \mathrm{Dom}(\Lambda_p^*)$. As $\mathcal{I}_{\|Z\|}$ is a rate function, it has a root in the (limit) expectation $z = m_{p,q}$ of the underlying sequence $\|Z\|$ (see, e.g., [29, Lemma 2.2.5]). Hence, it follows that for $z > m_{p,q}$ with $z \in \mathrm{Dom}(\Lambda_p^*)$ it holds that $\mathcal{I}_{\|Z\|}(z) = \Lambda_p^*(z^q, 1)$ is strictly increasing in $z$. Since $z_2 \leq 1$, $z = z_1 z_2$, and $1 \leq q$, we have $z_1^q \geq z > m_{p,q}$, meaning that $\Lambda_p^*(z_1^q, 1)$ is strictly increasing in $z_1$. Furthermore, we can see that $(-\log(z_2))$ is strictly decreasing in $z_2$. Hence, rewriting $z_1$ with respect to $z_2$ then gives

$$\mathcal{I}_{\|\mathscr{Z}\|}(z) = \inf_{\substack{z_1 = z/z_2 \\ z_2 \in (0,1]}} \left[ \Lambda_p^*\Big( \Big(\frac{z}{z_2}\Big)^q, 1 \Big) - \log(z_2) \right],$$

which is strictly decreasing in $z_2$. Thus, choosing $z_2 = 1$ gives $z_1 = z$ and

$$\mathcal{I}_{\|\mathscr{Z}\|}(z) = \mathcal{I}_{\mathcal{S}}(z^{**}) = \Lambda_p^*(z^*),$$

finishing the proof. $\square$

**Remark 5.2.7** As mentioned in Section 1.2, Schmuckenschläger [108] gave a central limit theorem for $q$-norms of random vectors with either distribution $\mathbf{C}_{n,p}$ or $\mathbf{U}_{n,p}$. He showed that for $p, q \in [1, \infty)$ with $q \neq p$, and a random vector $Z^{(n)} \in \mathbb{B}_p^n$ with $Z^{(n)} \sim \mathbf{C}_{n,p}$ or $Z^{(n)} \sim \mathbf{U}_{n,p}$ it holds that

$$\sqrt{n} \left( n^{1/p - 1/q} \, \frac{\|Z^{(n)}\|_q}{m_{p,q}} - 1 \right) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \sigma^2),$$

with $\mathcal{N}(0, \sigma^2)$ denoting a centered normal distribution, where $\sigma^2$ is also given explicitly in terms of $p$ and $q$ and moments of $\mathbf{N}_p$. This result shows that on the Gaussian scale the $q$-norm behaviours of both $\mathbf{C}_{n,p}$ and $\mathbf{U}_{n,p}$ coincide. It is thus a natural question whether one can tell the distributions apart in terms of their their $q$-norm behaviours by considering it beyond the Gaussian scale, e.g., via an LDP. And while the original LDPs in Proposition 5.2.2 and Proposition 5.2.5 given by Kabluchko, Prochno, and Thäle [61] at first seem to answer this in the positive by giving different rate functions for the LDPs, the reformulations of Lemma 5.2.4 and Lemma 5.2.6 show that one still observes the same $q$-norm behaviour of cone measure and uniform distribution when employing the tool of LDPs. Since the sharp large deviation results in this section actually look different for $\mathbf{C}_{n,p}$ and $\mathbf{U}_{n,p}$, they provide the first concentration result in which one can actually tell the two distributions apart. This again goes to show the high sensitivity of sharp large deviation results towards the underlying distributions.

**Remark 5.2.8** Note that in the results within this section, deviations from the "limit expectation" $m_{p,q}$ are considered, even though the elements of the sequences $\|Z\|$ and $\|\mathscr{Z}\|$ have respective expectations $m_{n,p,q}$ and $\mathscr{m}_{n,p,q}$, that only converge to $m_{p,q}$ in $n \in \mathbb{N}$. This, however, is not an issue for our results. As shown in (5.3) and (5.4), the sequences are represented via the empirical averages of probabilistic representations seen in (5.6) and (5.7). The expectations of these representations only ever play a role in our proofs regarding the behaviour of the corresponding cumulant generating functions, specifically only in the case of $\|Z\|$ (e.g., in the proofs of Lemma 5.2.6 and Lemma 5.2.18 or implicitly in the proof of the density approximations in Section 5.2.7). As the $V_i^{(n)}$ in (5.6) are i.i.d., they all share the same cumulant generating function as given in (5.8) and the same expectation

$$\mathbb{E}[V_i^{(n)}] = \left(\mathbb{E}\left[|Y_1^{(n)}|^q\right], \mathbb{E}\left[|Y_1^{(n)}|^p\right]\right) = \big((m_{p,q})^q, 1\big).$$

Hence, the fact that the expectation $m_{n,p,q}$ only converges to $m_{p,q}$ does not affect our proofs. This is in keeping with classical results from large deviations theory like the Theorem of Gärtner-Ellis (see Proposition 2.3.7), where an arbitrary (i.e., not necessarily i.i.d.) sequence of random variables is not required to have a shared expectation, but rather that the sequence of the (appropriately rescaled) cumulant generating functions of the individual random variables in the sequence converge to a fixed function with the origin in the interior of its effective domain. The resulting LDP then considers deviation probabilities from the limit expectation as well. In the case of $\|\mathscr{Z}\|$ the cumulant generating functions of the $\mathscr{V}_i^{(n)}$ are not employed at all (neither themselves nor their limit in $n$). Since our main results assume $n \in \mathbb{N}$ to be sufficiently large (that is, large enough for the local density approximations in Section 5.2.7 to hold), this effectively means that for $n \in \mathbb{N}$ sufficiently large, the difference of $m_{p,q}$ and $m_{n,p,q}, \mathscr{m}_{n,p,q}$ is of order at most $o(1)$ and therefore does not affect our sharp large deviation estimates.

### 5.2.2 Weingarten maps and curvature

The proof of the main result for $\ell_p^n$-spheres will proceed by integrating over a previously established density estimate via a result of Andriani and Baldi [7] for Laplace-type integrals. This result has a heavily geometric flavour and relies on the Weingarten maps of certain hypersurfaces, which in our setting will simply be curves in $\mathbb{R}^2$. We will therefore just give a brief reminder of the Weingarten map in this setting, recall some of its properties, and refer to the relevant literature (e.g., [48, 79]) or Andriani and Baldi [7] for a more in-depth discussion of the topic.

In general, the Weingarten map of a smooth hypersurface $M \subset \mathbb{R}^d$ at a point $p \in M$ is an endomorphism of the tangent space $T_pM$ at $p$, mapping any $y \in T_pM$ to the directional derivative of a normal field of $M$ in $p$ in the direction of $y$. However, as remarked in [7, Example 4.3], for $d = 2$, hypersurfaces simplify to planar curves and the Weingarten map at a point $p$ simplifies to the absolute value of the curvature $K(p)$ of the curve at $p$. For implicit curves, i.e., curves given as the zero set of a function, we have the following formula for its curvature from [39, Proposition 3.1]:

**Lemma 5.2.9** *Let $F : \mathbb{R}^2 \to \mathbb{R}$ be a twice differentiable function. Further, let $\mathscr{C} := \{x \in \mathbb{R}^2 : F(x) = 0\}$ be a curve given as the zero set of $F$, and $p \in \mathscr{C}$ be a point where $\nabla_x F(p) \neq 0$. Using the derivative notation $F_{[i,j]} = F_{[i,j]}(p)$ as in (2.1), it holds that*

$$K(p) = \frac{\left(-F_{[0,1]}, F_{[1,0]}\right) \begin{pmatrix} F_{[2,0]} & F_{[1,1]} \\ F_{[1,1]} & F_{[0,2]} \end{pmatrix} \left(-F_{[0,1]}, F_{[1,0]}\right)}{\left({F_{[1,0]}}^2 + {F_{[0,1]}}^2\right)^{3/2}} .$$

**Corollary 5.2.10**

*i) Given the set-up of the previous lemma, straightforward calculation of the above fraction gives that*

$$K(p) = \frac{{F_{[0,1]}}^2 F_{[2,0]} - 2F_{[0,1]}F_{[1,0]}F_{[1,1]} + {F_{[1,0]}}^2 F_{[0,2]}}{\left({F_{[1,0]}}^2 + {F_{[0,1]}}^2\right)^{3/2}} .$$

*ii) In case that $\mathscr{C}$ is the graph of a twice differentiable function $f : \mathbb{R} \to \mathbb{R}$, that is, $\mathscr{C} = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = f(x_1)\}$, and $p = (x, f(x))$, the above reduces to*

$$K(p) = \frac{|f''(x)|}{\left(1 + f'(x)^2\right)^{3/2}} .$$

### 5.2.3 SLD results for $q$-norms of $\ell_p^n$-balls and -spheres

Using the concepts and notation previously established, we now proceed to present the main results of Section 5.2:

For $1 \le q < p < \infty$ and $Z^{(n)} \sim \mathbf{C}_{n,p}$ we want to give sharp asymptotics for the probability

$$\mathbb{P}\Big(n^{1/p-1/q} \left\|Z^{(n)}\right\|_q > z\Big)$$

for $z > m_{p,q}$ such that $z^* \in \mathrm{Dom}(\Lambda_p^*)$, with $z^* := (z^q, 1)$ as in Lemma 5.2.4. Let us define prefactor functions $\xi(z)$ and $\kappa(z)$, as mentioned also in the sharp large deviation results of Bahadur and Ranga Rao in (5.1). For $z > m_{p,q}$ such that $z^* \in \mathrm{Dom}(\Lambda_p^*)$, we set

$$\xi(z)^2 := \langle \mathfrak{H}_{z^*}\, \tau(z^*), \tau(z^*)\rangle \, \det \mathfrak{H}_{z^*}, \tag{5.18}$$

with $\mathfrak{H}_{z^*} = \mathcal{H}_\tau \Lambda_p(\tau(z^*))$ as in (5.10), and

$$\kappa(z)^2 := 1 - c_\kappa(z), \tag{5.19}$$

with

$$c_\kappa(z) := \frac{\left(\tau(z^*)_1^2 + \tau(z^*)_2^2\right)^{3/2} |pq(p-q)z^q|}{\left|\tau(z^*)_2^2 \left(\mathfrak{H}_{z^*}^{-1}\right)_{11} - 2\tau(z^*)_1\tau(z^*)_2 \left(\mathfrak{H}_{z^*}^{-1}\right)_{12} + \tau(z^*)_1^2 \left(\mathfrak{H}_{z^*}^{-1}\right)_{22}\right| \left(z^{2q} + p^2q^{-2}\right)^{3/2}}.$$

We refer to Lemma 5.2.1 to see why the inverse matrix $\mathfrak{H}_{z^*}^{-1}$ in the above is in fact well-defined for $z^* \in \mathrm{Dom}(\Lambda_p^*)$. Let us now present the main result of Section 5.2 for $\ell_p^n$-spheres.

**Theorem 5.2.11** *Let $1 \le q < p < \infty$, $n \in \mathbb{N}$, and $Z^{(n)}$ be a random vector in $\mathbb{B}_p^n$ with $Z^{(n)} \sim \mathbf{C}_{n,p}$. Then, for $n$ sufficiently large and any $z > m_{p,q}$ such that $z^* \in \mathrm{Dom}(\Lambda_p^*)$, it holds that*

$$\mathbb{P}\Big(n^{1/p-1/q} \left\|Z^{(n)}\right\|_q > z\Big) = \frac{1}{\sqrt{2\pi n}\,\kappa(z)\xi(z)}\, e^{-n\,\Lambda_p^*(z^*)}\,(1+o(1)).$$

We shall now turn from $\ell_p^n$-spheres to $\ell_p^n$-balls. For $1 \le q < p < \infty$ and $\mathscr{Z}^{(n)} \sim \mathbf{U}_{n,p}$ we want to provide similar sharp asymptotics for

$$\mathbb{P}\Big(n^{1/p-1/q} \left\|\mathscr{Z}^{(n)}\right\|_q > z\Big).$$

for $z > m_{p,q}$ such that $z^* \in \mathrm{Dom}(\Lambda_p^*)$.

Again, we start by defining a prefactor function for $z > m_{p,q}$ with $z^* \in \mathrm{Dom}(\Lambda_p^*)$ as

$$\begin{aligned}\gamma(z)^2 \;:=\; & \det \mathfrak{H}_{z^*}\, \tau(z^*)_1^2 \left(q z^q \tau(z^*)_1 + 1\right)^2 \\ & \times \left[\frac{z^{2q}q^2}{p^2}\left(\mathfrak{H}_{z^*}^{-1}\right)_{11} + \frac{2z^q q}{p}\left(\mathfrak{H}_{z^*}^{-1}\right)_{12} + \left(\mathfrak{H}_{z^*}^{-1}\right)_{22} + \tau(z^*)_1 \frac{z^q q(q-p)}{p^2}\right], \quad (5.20)\end{aligned}$$

with which we can now give this section's main result for $\ell_p^n$-balls.

**Theorem 5.2.12** *Let $1 \le q < p < \infty$, $n \in \mathbb{N}$, and $\mathscr{Z}^{(n)}$ be a random vector in $\mathbb{B}_p^n$ with $\mathscr{Z}^{(n)} \sim \mathbf{U}_{n,p}$. Then, for $n$ sufficiently large and any $z > m_{p,q}$ such that $z^* \in \mathrm{Dom}(\Lambda_p^*)$, it holds that*

$$\mathbb{P}\Big(n^{1/p\,-\,1/q}\,\big\|\mathscr{Z}^{(n)}\big\|_q > z\Big) = \frac{1}{\sqrt{2\pi n}\,\gamma(z)}\, e^{-n\,\Lambda_p^*(z^*)}\,(1 + o(1)).$$

We have seen in Section 5.2.1 that $\|Z\|$ and $\|\mathscr{Z}\|$ both satisfy LDPs with the same speed and rate function for $z > m_{p,q}$ such that $z^* \in \mathrm{Dom}(\Lambda_p^*)$, despite the underlying distributions being different (see Remark 5.2.7). Comparing Theorem 5.2.11 and Theorem 5.2.12 now paints a different picture, with the sharp asymptotics for $\|Z\|$ and $\|\mathscr{Z}\|$ being noticeably different. As mentioned in the introduction and Section 5.1, idiosyncratic phenomena of underlying distributions, which can be drowned out on the LDP scale, are often still visible on the scale of sharp large deviations. This is in keeping with what was shown in [85, Theorem 2.4, Theorem 2.6] for one-dimensional projections of $\ell_p^n$-spheres and $\ell_p^n$-balls.

**Remark 5.2.13** Let us draw a brief comparison between our results and the concentration inequality that follows by the Gromov-Milman Theorem as discussed in [92, Remark, p. 1062]. Therein, it is shown that the Gromov-Milman theorem from [44] implies that for $1 < q \le p < \infty$ and a random vector $Z^{(n)} \sim \mathbf{C}_{n,p}$, it holds that

$$\mathbb{P}\Big(\big|n^{1/p\,-\,1/q}\,\big\|Z^{(n)}\big\|_q - m_{n,p,q}\big| \ge z\Big) \le C\,\exp\Big(-\,c\,n\,z^{\max\{2,p\}}\Big),$$

where $C > 0$ and $c > 0$ are constants. If we consider the set-up of Theorem 5.2.11, i.e., $1 \le q < p < \infty$ and $z > m_{n,p,q}$, and only consider deviations without the absolute value, we can derive from the above that

$$\mathbb{P}\Big(n^{1/p\,-\,1/q}\,\big\|Z^{(n)}\big\|_q > z\Big) \le C\,\exp\Big(-\,c\,n\,z^{\max\{2,p\}}\Big).$$

Comparing this with our sharp large deviation results from Theorem 5.2.11 for $z > m_{p,q}$ such that $z^* \in \mathrm{Dom}(\Lambda_p^*)$,

$$\mathbb{P}\Big(n^{1/p\,-\,1/q}\big\|Z^{(n)}\big\|_q > z\Big) = \frac{1}{\sqrt{2\pi n}\,\kappa(z)\xi(z)}\, e^{-n\,\Lambda_p^*(z^*)}\,(1+o(1)),$$

we can see that our results improve on the estimate in terms of $n \in \mathbb{N}$ by a factor of $n^{-1/2}$ and give explicit and deviation-dependent prefactor functions $\kappa(z)$ and $\xi(z)$ instead of fixed constants for all deviations $z$.

**Remark 5.2.14** When comparing the sharp large deviation results in Theorem 5.2.11 and Theorem 5.2.12 to those of Liao and Ramanan [85, Theorem 2.4] and [85, v2, Theorem 2.6], one directly notices the core difference in the settings. Liao and Ramanan examine projections of random vectors on $\mathbb{S}_p^{n-1}$ and $\mathbb{B}_p^n$ with respective distributions $\mathbf{C}_{n,p}$ and $\mathbf{U}_{n,p}$ onto fixed one-dimensional subspaces, and therefore have to consider weighted sums of *dependent* random vectors as probabilistic representations. Thus, all their results have to be conditioned on the projection space and include additional terms accounting for the specifics of the subspace. In our case, however, the probabilistic representations are given as sums of *i.i.d.* random variables (see Section 5.2.6), which does not necessitate these additional factors. Therefore, when using results from Liao and Ramanan [85], we adapt their usage accordingly to the given probabilistic representations in our setting. Beyond that, however, the sharp large deviation results share several similarities, especially when comparing the prefactor functions $\kappa$, $\xi$ and $\gamma$, which for $q = 1$ are almost equal.

### 5.2.4 Application 1: Intersection volumes of $\ell_p^n$-balls

We want to use our sharp large deviation results to further the findings of Schechtman and Schmuckenschläger [105] and Schmuckenschläger [108] for intersection volumes of $t$-multiples of different volume-normalized $\ell_p^n$-balls. We will first give a brief overview of the original results along the lines of [61, Section 2.1]. For $p \in [1, \infty)$, we define

$$\mathbb{D}_p^n := \mathrm{vol}_n(\mathbb{B}_p^n)^{-1/n}\,\mathbb{B}_p^n$$

to be the volume-normalized $\ell_p^n$-ball and recall that

$$\mathrm{vol}_n(\mathbb{B}_p^n) = \frac{\Big(2\,\Gamma\Big(1+\frac{1}{p}\Big)\Big)^n}{\Gamma\Big(1+\frac{n}{p}\Big)}.$$

We furthermore set

$$c_{n,p} := n^{1/p}\,\mathrm{vol}_n\left(\mathbb{B}_p^n\right)^{1/n} \qquad \text{and} \qquad c_p := 2\,e^{1/p}\,p^{1/p}\,\Gamma\left(1+\frac{1}{p}\right),$$

and recall that it was shown in [105] that

$$\lim_{n\to\infty} c_{n,p} = c_p.$$

Moreover, for $p,q \in [1,\infty), p \neq q$, we set

$$c_{n,p,q} := \frac{c_{n,p}}{c_{n,q}}, \qquad A_{n,p,q} := \frac{c_{n,p}}{m_{p,q}\,c_{n,q}}, \qquad \text{and} \qquad A_{p,q} := \lim_{n\to\infty} A_{n,p,q}.$$

Hence, it follows that

$$A_{p,q} = \frac{c_p}{m_{p,q}\,c_q} = \frac{\Gamma\left(1+\frac{1}{p}\right)^{1+(1/q)}}{\Gamma\left(1+\frac{1}{q}\right)\Gamma\left(\frac{q+1}{p}\right)^{1/q}}\; e^{1/p\,-\,1/q}.$$

Lastly, for $t \geq 0$ and $n \in \mathbb{N}$, we define $t_n \geq 0$ such that

$$t_n \frac{A_{p,q}}{A_{n,p,q}} = t.$$

We shall now recall the result of Schmuckenschläger [108, Theorem 3.3]. Therein, it was shown that for $p,q \in [1,\infty), p \neq q$, and $t \geq 0$ it holds that

$$\mathrm{vol}_n\left(\mathbb{D}_p^n \cap t\mathbb{D}_q^n\right) \underset{n\to\infty}{\longrightarrow} \begin{cases} 1 & : A_{p,q}\,t > 1 \\ \frac{1}{2} & : A_{p,q}\,t = 1 \\ 0 & : A_{p,q}\,t < 1 \end{cases} \tag{5.21}$$

(where the cases $A_{p,q}\,t > 1$ and $A_{p,q}\,t < 1$ had already been established in [105]). To prove this, a central limit theorem for $n^{1/p\,-\,1/q}\|\mathscr{Z}^{(n)}\|_q$ with $\mathscr{Z}^{(n)} \sim \mathbf{U}_{n,p}$ and $p,q \in [1,\infty)$, $p \neq q$, is shown in [108, Proposition 2.4, Proof of Theorem 3.2], since $\mathrm{vol}_n(\mathbb{D}_p^n \cap t\mathbb{D}_q^n)$ can be written as

$$\begin{aligned} \mathrm{vol}_n(\mathbb{D}_p^n \cap t\mathbb{D}_q^n) &= \mathrm{vol}_n\left(\left\{z \in \mathbb{D}_p^n : z \in t_n\,\frac{A_{p,q}}{A_{n,p,q}}\,\mathbb{D}_q^n\right\}\right) \\ &= \mathrm{vol}_n\left(\left\{z \in \mathbb{D}_p^n : z \in t_n\,A_{p,q}\,m_{p,q}\,\frac{c_{n,q}}{c_{n,p}}\,\mathbb{D}_q^n\right\}\right) \end{aligned}$$

$$
\begin{aligned}
&= \mathrm{vol}_n\big(\big\{z \in \mathrm{vol}_n(\mathbb{B}_p^n)^{-1/n}\,\mathbb{B}_p^n : z \in t_n\,A_{p,q}\,m_{p,q}\,n^{1/q\,-\,1/p}\,\mathrm{vol}_n(\mathbb{B}_p^n)^{-1/n}\,\mathbb{B}_q^n\big\}\big) \\
&= \mathrm{vol}_n(\mathbb{B}_p^n)^{-1}\,\mathrm{vol}_n\big(\big\{z \in \mathbb{B}_p^n : z \in t_n\,A_{p,q}\,m_{p,q}\,n^{1/q\,-\,1/p}\,\mathbb{B}_q^n\big\}\big) \\
&= \mathbb{P}\Big(n^{1/p\,-\,1/q}\big\|\mathscr{Z}^{(n)}\big\|_q \le t_n\,A_{p,q}\,m_{p,q}\Big)\,. \qquad (5.22)
\end{aligned}
$$

However, we know from the Berry-Esseen Theorem (see [114, Theorem 2.1.3]) that the error of the Gaussian approximation given by a central limit theorem decreases with rate $n^{-1/2}$. Thus, using (5.22) and the central limit theorem from [108], we can only infer a rate of convergence of $n^{-1/2}$ in (5.21). The LDP for $q$-norms from Proposition 5.2.5 already takes this from a sublinear rate to an exponential rate. This can then be improved upon for $1 \le q < p < \infty$ by using Theorem 5.2.12, yielding a more precise result that also allows to derive concrete estimates of this intersection volume, due to its asymptotic sharpness.

**Proposition 5.2.15** *Let $1 \le q < p < \infty$ and $n \in \mathbb{N}$. Using the notation established above, for $t > m_{p,q}\,c_{n,p,q}^{-1}$ such that $(t\,c_{n,p,q})^* \in \mathrm{Dom}(\Lambda_p^*)$, and sufficiently large $n \in \mathbb{N}$ it then holds that*

$$\mathrm{vol}_n\big(\mathbb{D}_p^n \cap t\mathbb{D}_q^n\big) = 1 - \frac{1}{\sqrt{2\pi n}\,\gamma(t\,c_{n,p,q})}\,e^{-n\,\Lambda_p^*((t\,c_{n,p,q})^*)}\,(1+o(1)).$$

*Proof.* Let $1 \le q < p < \infty$ and $t > m_{p,q}\,c_{n,p,q}^{-1}$ such that $(t\,c_{n,p,q})^* \in \mathrm{Dom}(\Lambda_p^*)$. Further, assume $\mathscr{Z}^{(n)}$ is a random vector in $\mathbb{B}_p^n$ with $\mathscr{Z}^{(n)} \sim \mathbf{U}_{n,p}$. Using (5.22), we get that

$$
\begin{aligned}
\mathrm{vol}_n\big(\mathbb{D}_p^n \cap t\mathbb{D}_q^n\big) &= \mathbb{P}\Big(n^{1/p\,-\,1/q}\big\|\mathscr{Z}^{(n)}\big\|_q \le t_n A_{p,q}\,m_{p,q}\Big) \\
&= 1 - \mathbb{P}\Big(n^{1/p\,-\,1/q}\big\|\mathscr{Z}^{(n)}\big\|_q > t_n A_{p,q}\,m_{p,q}\Big).
\end{aligned}
$$

It now holds, by $t > m_{p,q}\,c_{n,p,q}^{-1}$, that we have $t\,m_{p,q}^{-1}\,c_{n,p,q} = t\,A_{n,p,q} = t_n\,A_{p,q} > 1$, and hence $t\,c_{n,p,q} = t_n\,A_{p,q}\,m_{p,q} > m_{p,q}$ with $(t\,c_{n,p,q})^* \in \mathrm{Dom}(\Lambda_p^*)$. Thus, by Theorem 5.2.12, it follows that

$$\mathrm{vol}_n\big(\mathbb{D}_p^n \cap t\mathbb{D}_q^n\big) = 1 - \frac{1}{\sqrt{2\pi n}\,\gamma(t\,c_{n,p,q})}\,e^{-n\,\Lambda_p^*((t\,c_{n,p,q})^*)}\,(1+o(1)),$$

which finishes the proof. $\square$

### 5.2.5 Application 2: One-dimensional projections of $\ell_q^n$-balls

In Remark 5.2.14 we have already discussed the differences between the setting of the results of Liao and Ramanan [85] and the setting of this section. However, a geometrically similar result to those in [85] follows from Theorem 5.2.11. In [61, Section 2.4] Kabluchko, Prochno, and Thäle derived a central limit theorem for the the length of the projection of an $\ell_p^n$-ball onto the line spanned by a random vector $\theta^{(n)} \in \mathbb{S}^{n-1}$ with $\theta^{(n)} \sim \mathbf{C}_{n,2}$ as a corollary of their main results. We will proceed similarly and derive sharp large deviation results in the same setting. To be specific, in [85] sharp asymptotics where provided for the *scalar product* of a random vector $Z^{(n)} \sim \mathbf{C}_{n,p}$ on $\mathbb{S}_p^{n-1}$ with a vector $\theta^{(n)} \in \mathbb{S}^{n-1}$, which can be negative. We, on the other hand, consider the *absolute value* of the scalar product of such random vectors. Additionally, in [85] the vector $\theta^{(n)} \in \mathbb{S}_p^{n-1}$ is fixed instead of chosen randomly.

In the following define for $q \in [1,\infty]$ its conjugate $q^*$ via $\frac{1}{q} + \frac{1}{q^*} = 1$, setting $\frac{1}{\infty} = 0$ by convention. Furthermore, for a direction $\theta^{(n)} \in \mathbb{S}^{n-1}$, we write $P_{\theta^{(n)}}\mathbb{B}_q^n$ for the projection of $\mathbb{B}_q^n$ onto the line spanned by $\theta^{(n)}$. Then, our quantity of interest is the projection length $\mathrm{vol}_1\big(P_{\theta^{(n)}}\mathbb{B}_q^n\big)$ for some random direction $\theta^{(n)} \sim \mathbf{C}_{n,2}$.

**Corollary 5.2.16** *Let $2 < q \le \infty$ and $\theta^{(n)} \in \mathbb{S}^{n-1}$ be a random vector with $\theta^{(n)} \sim \mathbf{C}_{n,2}$. Then, for any $z > 2\,m_{2,q^*}$ such that $\left(\frac{z}{2}\right)^* \in \mathrm{Dom}(\Lambda_p^*)$, and sufficiently large $n \in \mathbb{N}$, it holds that*

$$\mathbb{P}\Big(n^{1/2-1/q}\,\mathrm{vol}_1\big(P_{\theta^{(n)}}\mathbb{B}_q^n\big) > z\Big) = \frac{1}{\sqrt{2\pi n}\,\kappa\left(\frac{z}{2}\right)\xi\left(\frac{z}{2}\right)}\, e^{-n\,\Lambda_2^*((\frac{z}{2})^*)}\,(1+o(1)),$$

*with $\Lambda_2$ as in* (5.8) *and $\xi,\kappa$ as in* (5.18), (5.19), *respectively, defined for $q^*$ and $p = 2$.*

*Proof.* It holds that

$$\begin{aligned}\mathbb{P}\Big(n^{1/2-1/q}\,\mathrm{vol}_1\big(P_{\theta^{(n)}}\mathbb{B}_q^n\big) > z\Big) &= \mathbb{P}\left(n^{1/2-1/q}\,2\sup_{x\in\mathbb{B}_q^n}\big|\langle x,\theta^{(n)}\rangle\big| > z\right)\\ &= \mathbb{P}\left(n^{1/2-1/q}\,\big\|\theta^{(n)}\big\|_{q^*} > \frac{z}{2}\right).\end{aligned}$$

Since $2 < q \le \infty$, we have $1 \le q^* < 2 = p$, whereby we can apply Theorem 5.2.11 to get that

$$\mathbb{P}\Big(n^{1/2-1/q}\,\mathrm{vol}_1\big(P_{\theta^{(n)}}\mathbb{B}_q^n\big) > z\Big) = \frac{1}{\sqrt{2\pi n}\,\kappa\left(\frac{z}{2}\right)\xi\left(\frac{z}{2}\right)}\, e^{-n\,\Lambda_2^*\left((\frac{z}{2})^*\right)}\,(1+o(1)),$$

with $\Lambda_2, \xi, \kappa$ as described above, which concludes the proof. □

### 5.2.6 Probabilistic representations for $q$-norms of $\ell_p^n$-balls

The first step in proving Theorem 5.2.11 and Theorem 5.2.12 will be rewriting the target probabilities in both theorems with respect to convenient probabilistic representations. Recalling the definitions of the random vectors $V^{(n)}$ and $\mathscr{V}^{(n)}$ from (5.6) and (5.7), we define

$$S^{(n)} := \frac{1}{n}\sum_{i=1}^{n} V_i^{(n)} \qquad \text{and} \qquad \mathscr{S}^{(n)} := \frac{1}{n}\sum_{i=1}^{n} \mathscr{V}_i^{(n)} \tag{5.23}$$

as the empirical averages of their respective coordinates. Furthermore, for $z > m_{p,q}$ we define the sets

$$D_z := \big\{(t_1,t_2)\in\mathbb{R}^2 : t_1,t_2>0,\ t_1^{1/q}\,t_2^{-1/p} > z\big\},$$

and

$$\mathscr{D}_z := \big\{(t_1,t_2,t_3)\in\mathbb{R}^3 : t_1,t_2>0,\ t_3\in(0,1],\ t_3\,t_1^{1/q}\,t_2^{-1/p} > z\big\}.$$

It then follows from the reformulations of $\|Z^{(n)}\|_q$ and $\|\mathscr{Z}^{(n)}\|_q$ in (5.3) and (5.4) that we can write the probabilities within Theorem 5.2.11 and Theorem 5.2.12 with respect to $S^{(n)}$ and $\mathscr{S}^{(n)}$, respectively, as

$$\begin{aligned}\mathbb{P}\Big(n^{1/p-1/q}\,\big\|Z^{(n)}\big\|_q > z\Big) &= \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\big|Y_i^{(n)}\big|^q > z^q\left(\frac{1}{n}\sum_{i=1}^{n}\big|Y_i^{(n)}\big|^p\right)^{q/p}\right)\\ &= \mathbb{P}\Big(S^{(n)}\in D_z\Big),\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}\Big(n^{1/p-1/q}\,\big\|\mathscr{Z}^{(n)}\big\|_q > z\Big) &= \mathbb{P}\left(U^{q/n}\,\frac{1}{n}\sum_{i=1}^{n}\big|Y_i^{(n)}\big|^q > z^q\left(\frac{1}{n}\sum_{i=1}^{n}\big|Y_i^{(n)}\big|^p\right)^{q/p}\right)\\ &= \mathbb{P}\Big(\mathscr{S}^{(n)}\in\mathscr{D}_z\Big).\end{aligned}$$

We refer to these sets $D_z$ and $\mathscr{D}_z$ as "deviation areas", since $S^{(n)}$ and $\mathscr{S}^{(n)}$ lying in $D_z$ and $\mathscr{D}_z$, respectively, represent a deviation of $\|Z^{(n)}\|_q$ and $\|\mathscr{Z}^{(n)}\|_q$ of size $z$. The idea will then be to write the target deviation probabilities as an integral of the densities of $S^{(n)}$ and $\mathscr{S}^{(n)}$ over the deviation areas $D_z$ and $\mathscr{D}_z$.

**Remark 5.2.17** Note that the boundaries of the deviation areas

$$\partial D_z = \{(t_1,t_2)\in\mathbb{R}^2 : t_1,t_2>0, t_1^{1/q}\,t_2^{-1/p} = z\} \tag{5.24}$$

and

$$\partial\mathscr{D}_z = \{(t_1,t_2,t_3)\in\mathbb{R}^3 : t_1,t_2>0,\ t_3\in(0,1],\ t_3\,t_1^{1/q}\,t_2^{-1/p} = z\}$$

are the same sets given by the infimum conditions in the respective LDPs for $\|Z\|$ and $\|\mathscr{Z}\|$ in Proposition 5.2.2 and Proposition 5.2.5. As mentioned in Remark 5.2.3, the structure of the functional one considers – the $q$-norm of a random vector from $\mathbb{B}_p^n$ in our setting – has a direct influence on an LDP via the infimum operator from the contraction principle. In the realm of sharp large deviations this influence is still there, however, it is quite a bit more subtle, as it originates from the geometric shape of the deviation area. Indeed, as we will see in the proofs of Theorem 5.2.11 and Theorem 5.2.12, these shapes actually do have an influence on the sharp large deviation behaviours, specifically via the prefactor functions.

The fact that for $z > m_{p,q}$ the rate functions of the LDPs for $\|Z\|$ and $\|\mathscr{Z}\|$ in Proposition 5.2.2 and Proposition 5.2.5 both assume a unique minimum on $\partial D_z$ and $\partial \mathscr{D}_z$, respectively, as was shown in Lemma 5.2.4 and Lemma 5.2.6, will be essential to the proof of Theorem 5.2.11 and Theorem 5.2.12 in Section 5.2.8 and Section 5.2.9. We can expand this unique infimum property onto the entirety of $\overline{D_z}$ and $\overline{\mathscr{D}_z}$, as the following lemma will show:

**Lemma 5.2.18** *Assume the same set-up as in Lemma 5.2.4 and Lemma 5.2.6. Let $z > m_{p,q}$ such that $z^* \in \mathrm{Dom}(\Lambda_p^*)$. Then*

*i) $z^* = (z^q, 1)$ is the unique point at which $\Lambda_p^*$ attains its infimum on $\overline{D_z}$,*

*ii) $z^{**} = (z^q, 1, 1)$ is the unique point at which $\mathcal{I}_{\mathscr{S}}$ attains its infimum on $\overline{\mathscr{D}_z}$.*

*Proof.* We start off by showing *i)*. Let $t \in \mathbb{R}^2$ such that $t \in D_z^\circ$, which means that $t_1^{1/q} t_2^{-1/p} > z$. We assume $t \in \mathrm{Dom}(\Lambda_p^*)$, as otherwise it trivially holds that $\Lambda_p^*(z^q, 1) < \Lambda_p^*(t_1, t_2) = \infty$. We then have that $t \in \partial D_{\tilde{z}}$ for $\tilde{z} := t_1^{1/q} t_2^{-1/p}$, and thus, by Lemma 5.2.4, $\Lambda_p^*(t_1, t_2) > \Lambda_p^*(\tilde{z}^q, 1) = \mathcal{I}_{\|Z\|}(\tilde{z})$. We know by Lemma 2.2.1 that $\Lambda_p^*$ is strictly convex on $\mathrm{Dom}(\Lambda_p^*)$ with a unique root in the limit expectation $(m_{p,q}^q, 1)$ of the $V_i^{(n)}$. Thus follows the strict convexity of $\mathcal{I}_{\|Z\|}(z) = \Lambda_p^*(z^*)$ in $z$ on $\mathrm{Dom}(\Lambda_p^*)$ with a unique root in $m_{p,q}$. Hence we know that $\mathcal{I}_{\|Z\|}(z)$ is strictly increasing in $z$ for $z > m_{p,q}$, and as $\tilde{z} > z > m_{p,q}$, it follows that

$$\Lambda_p^*(t_1, t_2) > \Lambda_p(\tilde{z}^q, 1) = \mathcal{I}_{\|Z\|}(\tilde{z}) > \mathcal{I}_{\|Z\|}(z) = \Lambda_p^*(z^q, 1) = \Lambda_p^*(z^*),$$

showing that $z^* = (z^q, 1)$ minimizes $\Lambda_p^*$ over $\overline{D_z}$. The proof of *ii)* is analogous, also using the strict monotonicity of the rate function. □

Suppose that the distributions of $S^{(n)}$ and $\mathscr{S}^{(n)}$ have respective densities $h^{(n)}$ and $\hbar^{(n)}$. Then we can formulate our probabilities of interest as

$$\mathbb{P}\Big(n^{1/p\,-\,1/q}\,\big\|Z^{(n)}\big\|_q > z\Big) = \mathbb{P}\Big(S^{(n)} \in D_z\Big) = \int\limits_{D_z} h^{(n)}(x)\,\mathrm{d}x, \tag{5.25}$$

and

$$\mathbb{P}\Big(n^{1/p\,-\,1/q}\,\big\|\mathscr{Z}^{(n)}\big\|_q > z\Big) = \mathbb{P}\Big(\mathscr{S}^{(n)} \in \mathscr{D}_z\Big) = \int\limits_{\mathscr{D}_z} \hbar^{(n)}(x)\,\mathrm{d}x. \tag{5.26}$$

In the following section we will show the existence of these densities $h^{(n)}$ and $\hbar^{(n)}$ and present asymptotic estimates for them, while Section 5.2.8 and Section 5.2.9 will then approximate the above integrals over the respective deviation areas $D_z$ and $\mathscr{D}_z$.

### 5.2.7 Asymptotic density estimate for $q$-norms of $\ell_p^n$-balls

The second step in proving Theorem 5.2.11 and Theorem 5.2.12 is giving local density approximations for the probabilistic representations $S^{(n)}$ and $\mathscr{S}^{(n)}$. Recalling the notation and definitions established in Section 5.2.1, we assume the same set-up as in Section 5.2.6 and formulate the following local limit theorems for the densities $h^{(n)}$ and $\hbar^{(n)}$ of our probabilistic representations:

**Proposition 5.2.19** *For $S^{(n)} = \frac{1}{n}\sum_{i=1}^n V_i^{(n)}$ with $V_i^{(n)} = \big(\big|Y_i^{(n)}\big|^q, \big|Y_i^{(n)}\big|^p\big)$, $Y_i^{(n)} \sim \mathbf{N}_p$ i.i.d., and $x \in \mathrm{Dom}(\Lambda_p^*)$, it holds that for sufficiently large $n \in \mathbb{N}$ the distribution of $S^{(n)}$ has Lebesgue density*

$$h^{(n)}(x) = \frac{n}{2\pi}\,(\det \mathfrak{H}_x)^{-1/2}\, e^{-n\,\Lambda_p^*(x)}\,(1 + o(1)),$$

*where $\mathfrak{H}_x := \mathcal{H}_\tau\Lambda_p(\tau(x))$ as in (5.10).*

For the proof of this we proceed along the lines of Borovkov and Rogozin [17] – or rather their convenient reformulation in [7, Theorem 3.1] and subsequent proof. Therein, a local density estimate is derived for a sum of i.i.d. random vectors in $\mathbb{R}^d$ via the saddle point method. As discussed in Section 5.1, this means one writes the density via the Fourier inversion theorem as a complex path integral of its Fourier transform and then uses Cauchy's theorem to deform the path of integration, such that it passes through a complex saddle point. For sufficiently large $n \in \mathbb{N}$, the mass of the integral then heavily concentrates around that saddle point and standard integral expansion methods can be used to great effect.

Naturally, this requires the conditions of the Fourier inversion theorem to be met, that is, the Fourier transform of the density has to be integrable. In [7, Theorem 3.1] this follows from the assumption that all the i.i.d. random vectors have a common bounded density, though it is noted in [7, Remark 3.2], that this can be replaced by any argument ensuring that the Fourier inversion theorem can be applied.

In the setting of this chapter (and the case of $\ell_p^n$-spheres) the i.i.d. vectors are given by the

$$V_i^{(n)} := \left( \left| Y_i^{(n)} \right|^q, \left| Y_i^{(n)} \right|^p \right) \in \mathbb{R}^2,$$

whose coordinates are highly dependent and who lie on the curve

$$\varpi_{p,q} := \{ y \in \mathbb{R}^2 : |y_2| = |y_1|^{p/q} \}.$$

Thus, such a non-degenerate density of the $V_i^{(n)}$ in $\mathbb{R}^2$ is not available. However, their Fourier transforms can be given explicitly. Additionally, while the coordinates of the $V_i^{(n)}$ are highly dependent, the different $V_i^{(n)}$ are i.i.d. and their *sums* do not lie on said curve, as one can see in Figure 5.1.

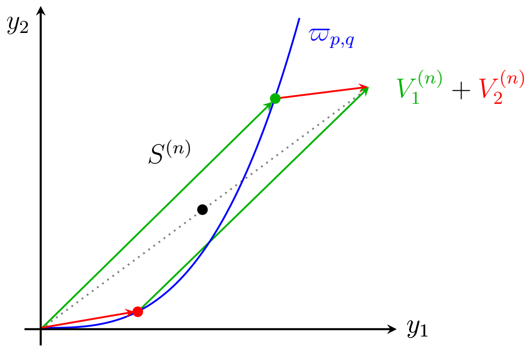


FIGURE 5.1: Empirical average $S^{(n)}$ (black) of $V_1^{(n)}$ (green), $V_2^{(n)}$ (red) for $n = 2, q = 1, p = 3$ lying outside of the curve $\varpi_{p,q}$ (blue)

We will show that the empirical average of the $V_i^{(n)}$ does in fact have an integrable Fourier transform and thus its density can be written as an integral of this Fourier transform. We will show this using the underlying distribution $\mathbf{N}_p$ of the $Y_i^{(n)}$ and the Hausdorff-Young inequality [43, Proposition 2.2.16], as was done by Liao and Ramanan in [85, Lemma 6.1]. Heuristically speaking, this means that while the individual $V_i^{(n)}$ do not possess densities in $\mathbb{R}^2$, for $n \in \mathbb{N}$ sufficiently large their empirical average $S^{(n)}$ asymptotically does. Beyond that, the proof will follow along the lines of Borovkov and Rogozin [17, Theorem 2] as presented in [7], using the saddle point method (implicitly via an exponential tilting argument) to approximate the integral of the Fourier transform.

The first step is to calculate the density $g^{(n)}$ of

$$G^{(n)} := \sum_{i=1}^{n} V_i^{(n)},$$

for $n > 1$. As previously mentioned, due to the high dependence (even comonotonicity) of the coordinates of the $V_i^{(n)}$, the individual $V_i^{(n)}$ do not have a density in $\mathbb{R}^2$. But since the $Y_i^{(n)} \sim \mathbf{N}_p$ have sufficiently finite exponential moments for $1 \le q < p < \infty$, both the moment and cumulant generating functions $\varphi_p$ and $\Lambda_p$ of the $V_i^{(n)}$ from (5.8) are finite in a neighbourhood of the origin. This can be used to show integrability of the Fourier transform of $G^{(n)}$. We will begin by showing this for two summands, that is, for $V_j^{(n)} + V_k^{(n)}$, $j \neq k$, in the following lemma in the spirit of [85, p.20, Claim].

**Lemma 5.2.20** *Let $x \in \mathrm{Dom}(\Lambda_p^*)$ and denote by $g_{j,k}^{(n)}$ the density of $V_j^{(n)} + V_k^{(n)}$, $j,k \in \{1,\dots n\}, j \neq k$, as in (5.6). For general $\tau \in \mathrm{Dom}(\Lambda_p)$ consider the function $f_\tau : \mathbb{R}^2 \to \mathbb{R}$ with $f_\tau(x) := e^{\langle \tau, x\rangle} g_{j,k}^{(n)}(x)$, where we suppress its dependence on other parameters such as $j,k$ in the notation. Then there exists an $s > 1$ such that the Fourier transform of $f_{\tau(x)}$ is $L_s$-integrable, i.e., $\mathcal{F}(f_{\tau(x)}) \in L_s(\mathbb{R}^2)$.*

*Proof.* Without loss of generality we chose $j = 1, k = 2$ for ease of notation. For $x \in \mathrm{Dom}(\Lambda_p^*)$ and general $\tau \in \mathrm{Dom}(\Lambda_p)$ consider the Fourier transform at some $t \in \mathbb{R}^2$:

$$\mathcal{F}(f_\tau)(t) = \int_{\mathbb{R}^2} e^{\langle \tau, x\rangle} g_{1,2}^{(n)}(x)\, e^{\langle it, x\rangle}\, \mathrm{d}x = \int_{\mathbb{R}^2} e^{\langle \tau + it, x\rangle} g_{1,2}^{(n)}(x)\, \mathrm{d}x = \mathbb{E}\left( e^{\left\langle \tau + it, V_1^{(n)} + V_2^{(n)} \right\rangle} \right).$$

By the definition and the independence of the $V_i^{(n)}$ and the properties of the moment generating function we have

$$\begin{aligned}
\mathcal{F}(f_\tau)(t) &= \mathbb{E}\left( e^{\left\langle \tau + it, V_1^{(n)} \right\rangle} \right)^2 \\
&= \mathbb{E}\left( e^{\left\langle \tau + it, \left( |Y_1^{(n)}|^q, |Y_1^{(n)}|^p \right) \right\rangle} \right)^2 \\
&= \left( \int_{\mathbb{R}} e^{\langle \tau + it, (|y|^q, |y|^p)\rangle} f_{\mathbf{N}_p}(y)\, \mathrm{d}y \right)^2,
\end{aligned}$$

where the last equality yields that $\mathcal{F}(f_\tau)(t) = \varphi_p(\tau + it)^2$ with $\varphi_p$ being the moment generating function of the $V_i^{(n)}$ as in (5.8) at the complex argument $\tau + it$. Since this holds for any $\tau \in \mathrm{Dom}(\Lambda_p)$, for $\tau(x) \in \mathrm{Dom}(\Lambda_p)$ we have that

$$\mathcal{F}(f_{\tau(x)})(t) = \varphi_p(\tau(x) + it)^2. \tag{5.27}$$

By the Hausdorff-Young inequality [43, Proposition 2.2.16] we can show that there exists an $s > 0$ such that $\mathcal{F}(f_\tau) \in L_s(\mathbb{R}^2)$ by showing instead that there exists an $r \in (0,1)$ such that $f_\tau \in L_{1+r}(\mathbb{R}^2)$ for general $\tau \in \mathrm{Dom}(\Lambda_p)$. We thus want to prove that for some $r \in (0,1)$

$$\int_{\mathbb{R}^2} \left| e^{\langle \tau, x \rangle} g_{1,2}^{(n)}(x) \right|^{1+r} \mathrm{d}x < \infty. \tag{5.28}$$

Since the density $g_{1,2}^{(n)}$ itself is unknown, we want to rewrite it as a transformation of the density of the $Y_i^{(n)}$ in the coordinates of $V_1^{(n)}$ and $V_2^{(n)}$, whose density $f_{\mathbf{N}_p}$ is known. It holds that

$$V_1^{(n)} + V_2^{(n)} = \Big( \big|Y_1^{(n)}\big|^q + \big|Y_2^{(n)}\big|^q, \big|Y_1^{(n)}\big|^p + \big|Y_2^{(n)}\big|^p \Big) =: T\Big(Y_1^{(n)}, Y_2^{(n)}\Big). \tag{5.29}$$

In order to make a transformation of densities argument, for a given $x \in \mathbb{R}_+^2$ we need to solve $x = T(y)$ for $y \in \mathbb{R}^2$. This means we are interested in the set

$$T^{-1}(x) = \big\{ y \in \mathbb{R}^2 : |y_1|^q + |y_2|^q = x_1,\ |y_1|^p + |y_2|^p = x_2 \big\}. \tag{5.30}$$

If either of the $x_1, x_2$ is zero, either both are zero and $T^{-1}(x) = (0,0)$, or $T^{-1}(x) = \emptyset$ otherwise. As this only holds for a zero set of $x \in \mathbb{R}_+^2$, we assume $x_1 > 0, x_2 > 0$. It follows from (5.30) that $T^{-1}(x)$ is the intersection of two $\ell_p^2$-spheres of radius $x_1^{1/q}$ and $x_2^{1/p}$, respectively, which we will denote as $\mathbb{S}_q^1\big(x_1^{1/q}\big)$ and $\mathbb{S}_p^1\big(x_2^{1/p}\big)$. Therefore, we have

$$T^{-1}(x) = \mathbb{S}_q^1\big(x_1^{1/q}\big) \cap \mathbb{S}_p^1\big(x_2^{1/p}\big).$$

As one can see in Figure 5.2, the number of intersection points depends on the relative size of their radii $x_1^{1/q}$ and $x_2^{1/p}$. Since $q < p$ and $\mathbb{B}_q^n \subseteq \mathbb{B}_p^n$, there are no intersection points if $x_1^{1/q} < x_2^{1/p}$. The same holds if $x_1^{1/q} > 2^{1/q - 1/p} x_2^{1/p}$, yielding that $T^{-1}(x)$ is empty, hence we disregard both cases. If $x_1^{1/q} = x_2^{1/p}$ we have the canonical four intersection points of $\ell_p^n$-spheres on the coordinate axes scaled by their common radius, and if $x_1^{1/q} = 2^{1/q - 1/p} x_2^{1/p}$ we again have exactly four intersections at the points $2^{-1/p} x_2^{1/p}(\pm 1, \pm 1) \in \mathbb{R}^2$. However, the set of $x \in \mathbb{R}_+^2$ for which these equalities hold is merely a zero set with respect to the distribution of $V_j^{(n)} + V_k^{(n)}$, hence we disregard these cases as well. Hence, we only concentrate on $x_1^{1/q} \in \big(x_2^{1/p}, 2^{1/q - 1/p} x_2^{1/p}\big)$, which yields eight intersection points in $\mathbb{R}^2$, which we denote by

$$T^{-1}(x)^{(1)} = y^{(1)}, \ \ldots\ , T^{-1}(x)^{(8)} = y^{(8)}. \tag{5.31}$$

Further, we write

$$H_{p,q} := \Big\{ x \in \mathbb{R}_+^2 : x_1, x_2 > 0,\ x_1^{1/q} \in \big(x_2^{1/p}, 2^{1/q - 1/p} x_2^{1/p}\big) \Big\}.$$

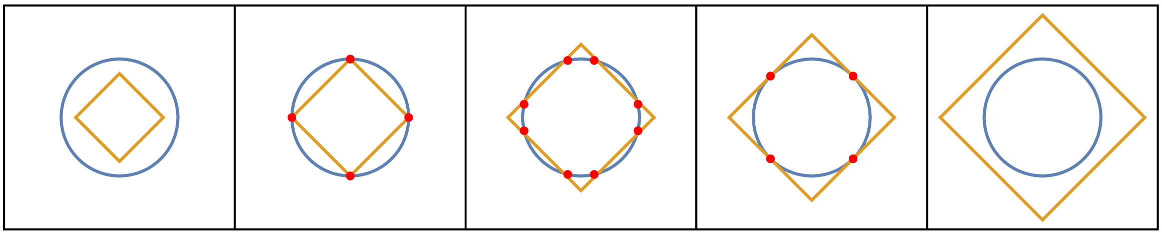

FIGURE 5.2: Intersection points (red) of $\mathbb{S}_1^1(r)$ (orange) and $\mathbb{S}_2^1(1)$ (blue) for different radii of $\mathbb{S}_1^1(r)$ with $r \in \{0.75,\, 1,\, 1.25,\, \sqrt{2},\, 1.75\}$ (from left to right).

We use a version of the well-known change of variable argument for "many-to-one" transformation functions, i.e., locally bijective functions, from [54, Section 4.5, p.151 ff.] which states the following: Let $Y_1, \dots, Y_n$ be continuous real-valued random variables with joint density $\varphi(y_1, \dots, y_n)$, and $\mathscr{A} := \{y \in \mathbb{R}^n : \varphi(y) > 0\}$, denoting $Y = (Y_1, \dots, Y_n)$. Further, let $X_1 = T_1(Y_1, \dots, Y_n), \dots, X_n = T_n(Y_1, \dots, Y_n)$ be real-valued transformations for which one can partition $\mathscr{A}$ (up to zero sets) into disjoint $A_1, \dots, A_k$ such that $X = T(Y) = (T_1(Y), \dots, T_n(Y))$ is bijective on these sets. We denote these bijective components as $T_{\bullet,j}^{-1} : \mathbb{R}^n \to A_j$ with $T_{\bullet,j}^{-1} := (T_{1,j}^{-1}, \dots, T_{n,j}^{-1})$, where $T_{i,j}^{-1}$ maps to the preimage of $T_i$ on $A_j$. Then the joint density of the $X_1, \dots, X_n$ at some $x = (x_1, \dots, x_n)$ is given by

$$g(x) = \sum_{y \in \mathbb{R}^n :\, T(y) = x} \left|\det\left[J_x T^{-1}(x)\right]\right| \varphi(y) = \sum_{j=1}^{k} \left|\det\left[J_x T_{\bullet,j}^{-1}(x)\right]\right| \varphi\big(T_{\bullet,j}^{-1}(x)\big).$$

We apply this to the transformation $T$ of $\big(Y_1^{(n)}, Y_2^{(n)}\big)$ from (5.29) with $\mathscr{A} = \mathbb{R}^2$ and $A_1, \dots, A_8$ chosen to be a partition of $\mathbb{R}^2$ up to zero sets such that $y^{(j)} \in A_j$ for all $j \in \{1, \dots, 8\}$ with $y^{(j)}$ as in (5.31). This yields that the density $g_{1,2}^{(n)}$ of $V_1^{(n)} + V_2^{(n)}$ at some $x \in \mathbb{R}_+^2$ is given by

$$g_{1,2}^{(n)}(x) = \left( \sum_{j=1}^{8} \left|\det\left[J_x T_{\bullet,j}^{-1}(x)\right]\right| f_{\mathbf{N}_p}\Big(T_{1,j}^{-1}(x)\Big)\, f_{\mathbf{N}_p}\Big(T_{2,j}^{-1}(x)\Big) \right) \mathbf{1}_{H_{p,q}}(x), \tag{5.32}$$

with $T_{\bullet,j}^{-1} = \Big(T_{1,j}^{-1}(x), T_{2,j}^{-1}(x)\Big)$. Let us calculate the determinant of $J_x T_{\bullet,j}^{-1}(x)$ explicitly. First off, it holds that $T(y_1, y_2) = (|y_1|^q + |y_2|^q, |y_1|^p + |y_2|^p)$ is continuously differentiable for any $1 \le q < p < \infty$ outside of $(0,0)$ with

$$J_y T(y) = \begin{pmatrix} \operatorname{sgn}(y_1) q |y_1|^{q-1} & \operatorname{sgn}(y_2) q |y_2|^{q-1} \\ \operatorname{sgn}(y_1) p |y_1|^{p-1} & \operatorname{sgn}(y_2) p |y_2|^{p-1} \end{pmatrix},$$

with $\operatorname{sgn}(t)$ denoting the sign of $t \in \mathbb{R}$, and

$$\det\big[J_yT(y)\big] = \operatorname{sgn}(y_1)\operatorname{sgn}(y_2)\, qp\,\Big(|y_1|^{q-1}|y_2|^{p-1} - |y_1|^{p-1}|y_2|^{q-1}\Big), \tag{5.33}$$

which is only zero for $|y_1| = |y_2|$. This case can be disregarded, however, since geometrically this would mean that $\mathbb{S}_q^1(x_1^{1/q})$ and $\mathbb{S}_p^1(x_2^{1/p})$ intersect at $2^{-1/p}x_2^{1/p}(\pm1,\pm1)$ and hence, $x_1^{1/q} = 2^{1/q\,-\,1/p}\,x_2^{1/p}$, which we excluded, as we only consider $T(y) = x \in H_{p,q}$. Since $T$ is continuously differentiable with non-zero functional determinant at the points $y^{(1)},\dots,y^{(8)}$, the inverse function theorem yields that for all $j \in \{1,\dots,8\}$

$$J_xT^{-1}_{\bullet,j}(x) = \Big[J_yT\big(y^{(j)}\big)\Big]^{-1},$$

which, together with standard rules for determinants of square matrices, yields that

$$\det\Big[J_xT^{-1}_{\bullet,j}(x)\Big] = \det\bigg[\Big[J_yT\big(y^{(j)}\big)\Big]^{-1}\bigg] = \det\Big[J_yT\big(y^{(j)}\big)\Big]^{-1}. \tag{5.34}$$

Further, it holds that

$$\begin{aligned} f_{\mathbf{N}_p}\Big(T^{-1}_{1,j}(x)\Big)\, f_{\mathbf{N}_p}\Big(T^{-1}_{2,j}(x)\Big) &= \bigg(2\,p^{1/p}\,\Gamma\Big(1+\frac{1}{p}\Big)\bigg)^{-2} e^{-\frac{1}{p}\left(T^{-1}_{1,j}(x)^p + T^{-1}_{2,j}(x)^p\right)} \\ &= \bigg(2\,p^{1/p}\,\Gamma\Big(1+\frac{1}{p}\Big)\bigg)^{-2} e^{-\frac{1}{p}\left(T_2\left(T^{-1}_{1,j}(x),\,T^{-1}_{2,j}(x)\right)\right)} \\ &= \bigg(2\,p^{1/p}\,\Gamma\Big(1+\frac{1}{p}\Big)\bigg)^{-2} e^{-\frac{1}{p}x_2}. \end{aligned}$$

Setting $\eta_p := 2\,p^{1/p}\,\Gamma\Big(1+\frac{1}{p}\Big)$ and using the above, (5.32) yields

$$g^{(n)}_{1,2}(x) = {\eta_p}^{-2}\left(\sum_{j=1}^{8}\left|\det\Big[J_xT^{-1}_{\bullet,j}(x)\Big]\right| e^{-\frac{1}{p}x_2}\right)\mathbf{1}_{H_{p,q}}(x).$$

To finish the proof of Lemma 5.2.20, as stated in (5.28), we want to show that $g^{(n)}_{1,2}(x) \in L_{1+r}(\mathbb{R}^2)$ for some $r > 0$. We thus consider

$$\begin{aligned} &\int_{\mathbb{R}^2}\Big|e^{\langle\tau,x\rangle}g^{(n)}_{1,2}(x)\Big|^{1+r}\,\mathrm{d}x \\ &= \int_{\mathbb{R}^2}\left|e^{\langle\tau,x\rangle}{\eta_p}^{-2}\left(\sum_{j=1}^{8}\left|\det\Big[J_xT^{-1}_{\bullet,j}(x)\Big]\right| e^{-\frac{1}{p}x_2}\right)\mathbf{1}_{H_{p,q}}(x)\right|^{1+r}\mathrm{d}x \\ &= {\eta_p}^{-2(1+r)}\int_{\mathbb{R}^2}\Big|e^{\tau_1x_1+\frac{1}{p}(p\tau_2-1)x_2}\Big|^{1+r}\left|\sum_{j=1}^{8}\det\Big[J_xT^{-1}_{\bullet,j}(x)\Big]\right|^{1+r}\mathbf{1}_{H_{p,q}}(x)\,\mathrm{d}x. \end{aligned}$$

Now it generally holds for $x_1, x_2 > 0$ and $r > 0$ that $(x_1+x_2)^{1+r} \le 2^r(x_1^{1+r}+x_2^{1+r})$. Successively applying this for $x_1,\dots,x_k > 0$ yields

$$\left(\sum_{i=1}^{k} x_i\right)^{1+r} \leq 2^{kr} \sum_{i=1}^{n} x_i^{1+r}.$$

Since we know from (5.33) and (5.34) that $\left|\det\left[J_x T_{\bullet,j}^{-1}(x)\right]\right| > 0$ for $x \in H_{p,q}$ and all $j \in \{1, \ldots, 8\}$, it follows from the above that

$$\begin{aligned}
&\int\limits_{\mathbb{R}^2} \left|e^{\langle\tau,x\rangle} g_{1,2}^{(n)}(x)\right|^{1+r} \mathrm{d}x \\
\leq\ & 2^{8r}\, \eta_p^{-2(1+r)} \int\limits_{H_{p,q}} \left|e^{\tau_1 x_1 + \frac{1}{p}(p\tau_2 - 1)x_2}\right|^{1+r} \sum_{j=1}^{8} \left|\det\left[J_x T_{\bullet,j}^{-1}(x)\right]\right|^{1+r} \mathrm{d}x \\
=\ & 2^{8r}\, \eta_p^{-2(1+r)} \sum_{j=1}^{8} \int\limits_{H_{p,q}} \left|\det\left[J_x T_{\bullet,j}^{-1}(x)\right]\right|^{1+r} e^{(1+r)(\tau_1 x_1 + \frac{1}{p}(p\tau_2 - 1)x_2)}\, \mathrm{d}x.
\end{aligned}$$

Moreover, using the change of variable argument in the other direction for each $T_{\bullet,j}^{-1}$, together with the fact that $T_{\bullet,j}^{-1}(H_{p,q}) = A_j$ by construction, and that for $y \in A_j$ with $T(y) = x$ it holds by (5.34) that

$$\det\left[J_x T_{\bullet,j}^{-1}(x)\right] \varphi\big(T_{\bullet,j}^{-1}(x)\big) = \det\left[J_y T(y)\right]^{-1} \varphi(y),$$

one can conclude by the partition property of the $A_1, \ldots, A_8$ that

$$\begin{aligned}
&\sum_{j=1}^{8} \int\limits_{H_{p,q}} \left|\det\left[J_x T_{\bullet,j}^{-1}(x)\right]\right|^{1+r} e^{(1+r)(\tau_1 x_1 + \frac{1}{p}(p\tau_2 - 1)x_2)}\, \mathrm{d}x \\
=\ & \sum_{j=1}^{8} \int\limits_{A_j} \left|\det\left[J_y T(y)\right]\right|^{-r} e^{(1+r)(\tau_1(|y_1|^q + |y_2|^q) + \frac{1}{p}(p\tau_2 - 1)(|y_1|^p + |y_2|^p))}\, \mathrm{d}y \\
=\ & \int\limits_{\mathbb{R}^2} \left|\det\left[J_y T(y)\right]\right|^{-r} e^{(1+r)(\tau_1(|y_1|^q + |y_2|^q) + \frac{1}{p}(p\tau_2 - 1)(|y_1|^p + |y_2|^p))}\, \mathrm{d}y \\
=\ & \int\limits_{\mathbb{R}^2} \left(qp\big|\big(|y_1|^{q-1}|y_2|^{p-1} - |y_1|^{p-1}|y_2|^{q-1}\big)\big|\right)^{-r} e^{(1+r)(\tau_1(|y_1|^q + |y_2|^q) + \frac{1}{p}(p\tau_2 - 1)(|y_1|^p + |y_2|^p))}\, \mathrm{d}y,
\end{aligned}$$

where the last equality is due to (5.33). For a neighbourhood of the origin $B \subset \mathbb{R}^2$ we can split up the above integral on $B$ and its complement $B^c$ as

$$\begin{aligned}
&\int\limits_{B} \left(qp\big|\big(|y_1|^{q-1}|y_2|^{p-1} - |y_1|^{p-1}|y_2|^{q-1}\big)\big|\right)^{-r} e^{(1+r)(\tau_1(|y_1|^q + |y_2|^q) + \frac{1}{p}(p\tau_2 - 1)(|y_1|^p + |y_2|^p))}\, \mathrm{d}y \\
+\ & \int\limits_{B^c} \left(qp\big|\big(|y_1|^{q-1}|y_2|^{p-1} - |y_1|^{p-1}|y_2|^{q-1}\big)\big|\right)^{-r} e^{(1+r)(\tau_1(|y_1|^q + |y_2|^q) + \frac{1}{p}(p\tau_2 - 1)(|y_1|^p + |y_2|^p))}\, \mathrm{d}y.
\end{aligned} \tag{5.35}$$

Since $\tau \in \mathrm{Dom}(\Lambda_p) = \mathbb{R} \times \big(-\infty, \frac{1}{p}\big)$, it holds for any $r > 0$ that

$$e^{(\tau_1(|y_1|^q+|y_2|^q)+\frac{1}{p}(p\tau_2-1)(|y_1|^p+|y_2|^p))} \in L_r(\mathbb{R}^2).$$

For the first integral in (5.35) we can find a sufficiently small $r_1 \in (0,1)$ such that

$$\big(qp\big|\big(|y_1|^{q-1}|y_2|^{p-1} - |y_1|^{p-1}|y_2|^{q-1}\big)\big|\big)^{-1} \in L_{r_1}(B).$$

Hence, for $r = r_1$ both factors in the first integral term in (5.35) are in $L_1(B)$, and since $B$ is bounded it follows via Hölders inequality that their product also lies in $L_1(B)$, i.e., the first integral is finite. For the second integral expression, we can find a sufficiently large $r_2 > 1$ such that

$$\big(qp\big(|y_1|^{q-1}|y_2|^{p-1} - |y_1|^{p-1}|y_2|^{q-1}\big)\big)^{-1} \in L_{r_2}(B^c).$$

Hence, for $r_3 := \frac{r_2}{r_1} > 1$ it follows that

$$\big(qp\big(|y_1|^{q-1}|y_2|^{p-1} - |y_1|^{p-1}|y_2|^{q-1}\big)\big)^{-r_1} \in L_{r_3}(B^c).$$

Lastly, for $r_3^* > 1$ such that $\frac{1}{r_3} + \frac{1}{r_3^*} = 1$ we know that

$$e^{(1+r_1)(\tau_1(|y_1|^q+|y_2|^q)+\frac{1}{p}(p\tau_2-1)(|y_1|^p+|y_2|^p))} \in L_{r_3^*}(B^c),$$

thus, by again applying Hölders inequality, we get that the product of the two functions lies in $L_1(B^c)$, and thereby the second integral term in (5.35) is also finite. Overall, we have shown that for sufficiently small $r = r_1 \in (0,1)$

$$\begin{aligned}
\int\limits_{\mathbb{R}^2} \big|f_\tau(x)\big|^{1+r}\,\mathrm{d}x &= \int\limits_{\mathbb{R}^2} \big|e^{\langle\tau,x\rangle} g_{1,2}^{(n)}(x)\big|^{1+r}\,\mathrm{d}x \\
&\le 2^{8r}\,\eta_p^{-2(1+r)} \sum_{j=1}^{8} \int\limits_{H_{p,q}} \det\Big[J_x T_{\bullet,j}^{-1}(x)\Big]^{1+r} e^{(1+r)(\tau_1 x_1+\frac{1}{p}(p\tau_2-1)x_2)}\,\mathrm{d}x \\
&= 2^{8r}\,\eta_p^{-2(1+r)} \int\limits_{\mathbb{R}^2} \big(qp\big|\big(|y_1|^{q-1}|y_2|^{p-1} - |y_1|^{p-1}|y_2|^{q-1}\big)\big|\big)^{-r} \\
&\qquad\qquad \times\, e^{(1+r)(\tau_1(|y_1|^q+|y_2|^q)+\frac{1}{p}(p\tau_2-1)(|y_1|^p+|y_2|^p))}\,\mathrm{d}y \\
&< \infty.
\end{aligned}$$

By the same arguments, one can also infer $f_\tau(x) \in L_{1+\tilde{r}}$ for any $\tilde{r} \in (0, r_1)$. As stated previously, by the Hausdorff-Young inequality [43, Proposition 2.2.16] we can hence conclude for $\tau = \tau(x) \in \mathrm{Dom}(\Lambda_p)$ that there exists an $s \in \big(\frac{1+r}{r}, \infty\big)$ such that $\mathcal{F}(f_{\tau(x)}) \in L_s(\mathbb{R}^2)$, which proves Lemma 5.2.20. $\square$

By (5.27) and Lemma 5.2.20 it follows that there exists an $s > 1$ such that

$$\varphi_p(\tau(x) + it) \in L_{s/2}(\mathbb{R}^2). \tag{5.36}$$

This will now be used to write the density $g^{(n)}$ of $\sum_{i=1}^n V_i^{(n)}$ via is Fourier transform, that is, to prove the following lemma.

**Lemma 5.2.21** *For $G^{(n)} = \sum_{i=1}^n V_i^{(n)}$ with $V_i^{(n)} := \left(\left|Y_i^{(n)}\right|^q, \left|Y_i^{(n)}\right|^p\right)$, $Y_i^{(n)} \sim \mathbf{N}_p$ i.i.d., $x \in \mathrm{Dom}(\Lambda_p^*)$, and sufficiently large $n \in \mathbb{N}$, it holds that the distribution of $\sum_{i=1}^n V_i^{(n)}$ has Lebesgue density*

$$g^{(n)}(x) = \left(\frac{1}{2\pi}\right)^2 e^{-\langle \tau(x), x\rangle} \int\limits_{\mathbb{R}^2} e^{-\langle it, x\rangle} (\varphi_p(\tau(x) + it))^n \mathrm{d}t.$$

*Proof.* For $x \in \mathrm{Dom}(\Lambda_p^*)$ and general $\tau \in \mathrm{Dom}(\Lambda_p)$ define $f_\tau : \mathbb{R}^2 \to \mathbb{R}$ with $f_\tau(x) := e^{\langle \tau, x\rangle} g^{(n)}(x)$ and proceed similar to the proof of Lemma 5.2.20 by considering its Fourier transform at some $t \in \mathbb{R}^2$. The first goal is to show that $\mathcal{F}(f_{\tau(x)}) \in L_1(\mathbb{R}^2)$ to apply the Fourier inversion theorem. It holds that

$$\mathcal{F}(f_\tau)(t) = \int\limits_{\mathbb{R}^2} e^{\langle \tau, x\rangle} g^{(n)}(x)\, e^{\langle it, x\rangle}\, \mathrm{d}x = \int\limits_{\mathbb{R}^2} e^{\langle \tau + it, x\rangle} g^{(n)}(x)\, \mathrm{d}x = \mathbb{E}\left(e^{\langle \tau + it, \sum_{i=1}^n V_i^{(n)}\rangle}\right).$$

Again, the independence of the $V_i^{(n)}$ and the properties of the moment generating function yield that

$$\mathcal{F}(f_\tau)(t) = \mathbb{E}\left(e^{\langle \tau + it, (|Y_1|^q, |Y_1|^p)\rangle}\right)^n = \left(\int\limits_{\mathbb{R}} e^{\langle \tau + it, (|y|^q, |y|^p)\rangle} f_{\mathbf{N}_p}(y) \mathrm{d}y\right)^n = \varphi_p(\tau + it)^n.$$

As seen previously, it follows from Lemma 5.2.20 that there exists an $s > 1$ such that, as in (5.36), we have

$$\varphi_p(\tau(x) + it) \in L_{s/2}(\mathbb{R}^2).$$

For $n \in \mathbb{N}$ large enough such that $n > \frac{s}{2}$, it thus follows that

$$\mathcal{F}(f_{\tau(x)})(t) = \varphi_p(\tau(x) + it)^n \in L_1(\mathbb{R}^2).$$

Applying the Fourier inversion theorem (cf. [110, Theorem 1.9]) to $f_\tau(x) = e^{\langle \tau, x\rangle} g^{(n)}(x)$ then yields

$$e^{\langle \tau, x\rangle} g^{(n)}(x) = \left(\frac{1}{2\pi}\right)^2 \int\limits_{\mathbb{R}^2} e^{-\langle it, x\rangle} \mathcal{F}(f_\tau)(t) \mathrm{d}t = \left(\frac{1}{2\pi}\right)^2 \int\limits_{\mathbb{R}^2} e^{-\langle it, x\rangle} (\varphi_p(\tau + it))^n \mathrm{d}t,$$

thus,

$$g^{(n)}(x) = \left(\frac{1}{2\pi}\right)^2 e^{-\langle \tau, x\rangle} \int\limits_{\mathbb{R}^2} e^{-\langle it, x\rangle} (\varphi_p(\tau + it))^n \mathrm{d}t.$$

Since the above holds for arbitrary $\tau \in \mathrm{Dom}(\Lambda_p)$, it also follows for $\tau(x) \in \mathrm{Dom}(\Lambda_p)$ that

$$g^{(n)}(x) = \left(\frac{1}{2\pi}\right)^2 e^{-\langle \tau(x), x\rangle} \int\limits_{\mathbb{R}^2} e^{-\langle it, x\rangle} (\varphi_p(\tau(x) + it))^n \mathrm{d}t, \tag{5.37}$$

which proves the claim. □

The exponential term $e^{\langle \tau, x\rangle}$ in the definition of $f_\tau$ and the specific choice of $\tau(x) \in \mathrm{Dom}(\Lambda_p)$ in (5.37) were not necessary up to this point, but will be helpful to approximate the remaining integral term in the density. We will now expand the results of Lemma 5.2.21 to yield a similar integral expression for the density $h^{(n)}$ of $S^{(n)}$:

**Proposition 5.2.22** *For $S^{(n)} = \frac{1}{n}\sum_{i=1}^n V_i^{(n)}$ with $V_i^{(n)} := \left(\left|Y_i^{(n)}\right|^q, \left|Y_i^{(n)}\right|^p\right)$, $Y_i^{(n)} \sim \mathbf{N}_p$ i.i.d., $x \in \mathrm{Dom}(\Lambda_p^*)$, and for $n \in \mathbb{N}$ large enough the distribution of $S^{(n)}$ has Lebesgue density*

$$h^{(n)}(x) = \left(\frac{n}{2\pi}\right)^2 e^{-n\Lambda_p^*(x)} \int\limits_{\mathbb{R}^2} e^{-n\langle it, x\rangle} \left(\frac{\varphi_p(\tau(x) + it)}{\varphi_p(\tau(x))}\right)^n \mathrm{d}t.$$

*Proof.* Setting $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ with $\phi(x) = \frac{1}{n}x$ such that $S^{(n)} = \phi\left(\sum_{i=1}^n V_i^{(n)}\right)$, yields

$$h^{(n)}(x) = g^{(n)}\big(\phi^{-1}(x)\big) \left|\det \mathrm{J}_x(\phi^{-1})\right| = g^{(n)}(nx)\, n^2.$$

Hence, Lemma 5.2.21 gives

$$h^{(n)}(x) = \left(\frac{n}{2\pi}\right)^2 e^{-n\langle \tau(x), x\rangle} \int\limits_{\mathbb{R}^2} e^{-n\langle it, x\rangle} (\varphi_p(\tau(x) + it))^n \mathrm{d}t.$$

Furthermore, note that by Lemma 2.2.1 (4) we have $\langle \tau(x), x\rangle = \Lambda_p^*(x) + \Lambda_p(\tau(x))$, and by definition (5.8) it holds that $e^{-\Lambda_p(\tau(x))} = \varphi_p(\tau(x))^{-1}$. Thus,

$$\begin{aligned} h^{(n)}(x) &= \left(\frac{n}{2\pi}\right)^2 e^{-n(\Lambda_p^*(x) + \Lambda_p(\tau(x)))} \int\limits_{\mathbb{R}^2} e^{-n\langle it, x\rangle} (\varphi_p(\tau(x) + it))^n \mathrm{d}t \\ &= \left(\frac{n}{2\pi}\right)^2 e^{-n\Lambda_p^*(x)} \int\limits_{\mathbb{R}^2} e^{-n\langle it, x\rangle} \left(\frac{\varphi_p(\tau(x) + it)}{\varphi_p(\tau(x))}\right)^n \mathrm{d}t, \end{aligned}$$

finishing the proof. □

We denote the integral term in the above as

$$\mathcal{I}^{(n)}(x) := \int_{\mathbb{R}^2} e^{-n\langle it,x\rangle} \left( \frac{\varphi_p(\tau(x)+it)}{\varphi_p(\tau(x))} \right)^n \mathrm{d}t. \tag{5.38}$$

The approximation of $\mathcal{I}^{(n)}(x)$ will be the content of the next result and the final step in the proof of the asymptotic density estimate in Proposition 5.2.19.

**Lemma 5.2.23** *Let $x \in \mathrm{Dom}(\Lambda_p^*)$ and $h^{(n)}$ be the density from Proposition 5.2.22 with integral coefficient $\mathcal{I}^{(n)}(x)$ as in* (5.38)*. It then holds that*

$$\mathcal{I}^{(n)}(x) = \left(\frac{n}{2\pi}\right)^{-1} \left(\det \mathfrak{H}_x\right)^{-1/2} (1+o(1)),$$

*where $\mathfrak{H}_x := \mathcal{H}_\tau \Lambda_p(\tau(x))$.*

As one can see, Proposition 5.2.22 and Lemma 5.2.23 directly imply the asymptotic density estimate in Proposition 5.2.19.

We prove Lemma 5.2.23 via the saddle point method, whose basic idea was outlined in Section 5.1. Specifically, this will be done via what we have called an "implicit" application of the saddle point method in the form of exponential tilting. We will define a conveniently shifted and exponentially tilted distribution, such that the empirical average of random variables with that tilted distribution is centered, and has the integrand of $\mathcal{I}^{(n)}(x)$ as its Fourier transform. The integral over this distribution's Fourier transform then heavily concentrates around the origin and can therefore be efficiently approximated using standard techniques.

For general $\tau \in \mathrm{Dom}(\Lambda_p)$ and $n \in \mathbb{N}$, we define the probability measure on $\mathbb{R}$

$$\mathbf{N}_{p,\tau} := e^{\langle \tau, (|y|^q, |y|^p)\rangle - \Lambda_p(\tau)} \, \mathbf{N}_p, \tag{5.39}$$

which is $\mathbf{N}_p$ exponentially tilted with respect to $\tau \in \mathrm{Dom}(\Lambda_p)$. Accordingly, $\mathbf{N}_{p,\tau}$ has the density

$$f_{\mathbf{N}_{p,\tau}}(y) := e^{\langle \tau, (|y|^q, |y|^p)\rangle - \Lambda_p(\tau)} \, f_{\mathbf{N}_p}(y).$$

One can quickly check that this is indeed still a probability measure, since

$$\int_{\mathbb{R}} \mathbf{N}_{p,\tau}(\mathrm{d}y) = e^{-\Lambda_p(\tau)} \int_{\mathbb{R}} e^{\langle \tau, (|y|^q, |y|^p)\rangle} \, f_{\mathbf{N}_p}(y) \, \mathrm{d}y = \varphi_p(\tau)^{-1} \, \varphi_p(\tau) = 1.$$

For $x \in \mathrm{Dom}(\Lambda_p^*)$ we define the i.i.d. random vectors $\mathcal{V}_1^{(n)}, \dots, \mathcal{V}_n^{(n)}$, with

$$\mathcal{V}_i^{(n)} := \Big(|\mathcal{Y}_i^{(n)}|^q, |\mathcal{Y}_i^{(n)}|^p\Big), \qquad \text{with} \qquad \mathcal{Y}_i^{(n)} \sim \mathbf{N}_{p,\tau(x)},$$

resulting in an exponentially tilted version of $V_i^{(n)}$. We also define the sums

$$\mathcal{G}^{(n)} := \sum_{i=1}^n \Big(\mathcal{V}_i^{(n)} - x\Big) \quad \text{and} \quad \mathcal{S}^{(n)} := \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big(\mathcal{V}_i^{(n)} - x\Big) \tag{5.40}$$

of the $\mathcal{V}_i^{(n)}$ shifted by $x$, denoting by $g_x^{(n)}$ the density of $\mathcal{G}^{(n)}$. We show two useful results for these auxiliary quantities.

**Lemma 5.2.24** *For $x \in \mathrm{Dom}(\Lambda_p^*)$ and $\mathcal{G}^{(n)}$ as defined in* (5.40) *it holds that*

$$\mathcal{F}\big(g_x^{(n)}\big)(t) = e^{-n\langle it,x\rangle} \left(\frac{\varphi_p(\tau(x)+it)}{\varphi_p(\tau(x))}\right)^n.$$

*Proof.* For $t \in \mathbb{R}^2$ it holds that

$$\begin{aligned}
\mathcal{F}\big(g_x^{(n)}\big)(t) &= \int_{\mathbb{R}^2} e^{\langle it,y\rangle}\, g_x^{(n)}(y)\,\mathrm{d}y \\
&= \mathbb{E}\left[e^{\left\langle it,\mathcal{G}^{(n)}\right\rangle}\right] \\
&= \mathbb{E}\left[e^{\left\langle it,\sum_{i=1}^n\left(\mathcal{V}_i^{(n)}-x\right)\right\rangle}\right] \\
&= e^{-\langle it,nx\rangle}\,\mathbb{E}\left[e^{\left\langle it,\mathcal{V}_1^{(n)}\right\rangle}\right]^n \\
&= e^{-\langle it,nx\rangle}\left(\int_{\mathbb{R}} e^{\langle it,(|y|^q,|y|^p)\rangle}\,\mathbf{N}_{p,\tau(x)}(\mathrm{d}y)\right)^n \\
&= e^{-\langle it,nx\rangle}\left(\int_{\mathbb{R}} e^{\langle it,(|y|^q,|y|^p)\rangle}\, e^{\langle \tau(x),(|y|^q,|y|^p)\rangle-\Lambda_p(\tau(x))}\, f_{\mathbf{N}_p}(y)\,\mathrm{d}y\right)^n \\
&= e^{-\langle it,nx\rangle}\; e^{-n\Lambda_p(\tau(x))}\left(\int_{\mathbb{R}} e^{\langle \tau(x)+it,(|y|^q,|y|^p)\rangle}\, f_{\mathbf{N}_p}(y)\,\mathrm{d}y\right)^n \\
&= e^{-n\langle it,x\rangle}\left(\frac{\varphi_p(\tau(x)+it)}{\varphi_p(\tau(x))}\right)^n,
\end{aligned}$$

finishing the proof. $\square$

**Lemma 5.2.25** *For $\mathcal{S}^{(n)}$ as defined in* (5.40) *it holds that*

*i)* $\mathbb{E}\Big[\mathcal{S}^{(n)}\Big] = 0$,

*ii)* $\mathrm{Cov}\Big[\mathcal{S}^{(n)}\Big] = \mathfrak{H}_x$, *with* $\mathfrak{H}_x := \mathcal{H}_\tau \Lambda_p(\tau(x))$,

*with $\Lambda_p$ still being the cumulant generating function of $V_i^{(n)}$ as defined as in* (5.8) *with respect to $\mathbf{N}_p$.*

*Proof.* Since the $\mathcal{V}_i^{(n)}$ are i.i.d., it holds that $\mathbb{E}\Big[\mathcal{S}^{(n)}\Big] = \sqrt{n}\left(\mathbb{E}\Big[\mathcal{V}_1^{(n)}\Big] - x\right)$. We continue by showing

$$\mathbb{E}\Big[\mathcal{V}_1^{(n)}\Big] = \nabla_s \mathbb{E}\Big[e^{\left\langle s,\mathcal{V}_1^{(n)}\right\rangle}\Big]\Big|_{s=(0,0)} = \nabla_\tau \Lambda_p(\tau(x)).$$

It holds that

$$\begin{aligned}
\nabla_s \mathbb{E}\Big[e^{\left\langle s,\mathcal{V}_i^{(n)}\right\rangle}\Big] &= \left(\frac{\partial}{\partial s_1}\int_{\mathbb{R}} e^{\langle s,(|y|^q,|y|^p)\rangle}\,\mathbf{N}_{p,\tau(x)}(\mathrm{d}y),\ \frac{\partial}{\partial s_2}\int_{\mathbb{R}} e^{\langle s,(|y|^q,|y|^p)\rangle}\,\mathbf{N}_{p,\tau(x)}(\mathrm{d}y)\right)\\
&= \left(\int_{\mathbb{R}} |y|^q\, e^{\langle s,(|y|^q,|y|^p)\rangle}\,\mathbf{N}_{p,\tau(x)}(\mathrm{d}y), \int_{\mathbb{R}} |y|^p\, e^{\langle s,(|y|^q,|y|^p)\rangle}\,\mathbf{N}_{p,\tau(x)}(\mathrm{d}y)\right),
\end{aligned}$$

thus,

$$\begin{aligned}
\nabla_s \mathbb{E}\Big[e^{\left\langle s,\mathcal{V}_i^{(n)}\right\rangle}\Big]\Big|_{s=(0,0)} &= \left(\int_{\mathbb{R}} |y|^q\,\mathbf{N}_{p,\tau(x)}(\mathrm{d}y), \int_{\mathbb{R}} |y|^p\,\mathbf{N}_{p,\tau(x)}(\mathrm{d}y)\right)\\
&= \left(\mathbb{E}\Big[|\mathcal{Y}_i^{(n)}|^q\Big], \mathbb{E}\Big[|\mathcal{Y}_i^{(n)}|^p\Big]\right)\\
&= \mathbb{E}\Big[\mathcal{V}_i^{(n)}\Big].
\end{aligned}$$

Furthermore, for some $\tau \in \mathrm{Dom}(\Lambda_p)$,

$$\begin{aligned}
\nabla_\tau \Lambda_p(\tau) &= \left(\frac{\partial}{\partial \tau_1}\log\int_{\mathbb{R}} e^{\langle \tau,(|y|^q,|y|^p)\rangle} f_{\mathbf{N}_p}(y)\,\mathrm{d}y,\ \frac{\partial}{\partial \tau_2}\log\int_{\mathbb{R}} e^{\langle \tau,(|y|^q,|y|^p)\rangle} f_{\mathbf{N}_p}(y)\,\mathrm{d}y\right)\\
&= e^{-\Lambda_p(\tau)}\left(\int_{\mathbb{R}} |y|^q\, e^{\langle \tau,(|y|^q,|y|^p)\rangle} f_{\mathbf{N}_p}(y)\,\mathrm{d}y, \int_{\mathbb{R}} |y|^p\, e^{\langle \tau,(|y|^q,|y|^p)\rangle} f_{\mathbf{N}_p}(y)\,\mathrm{d}y\right)\\
&= \left(\int_{\mathbb{R}} |y|^q\,\mathbf{N}_{p,\tau}(\mathrm{d}y), \int_{\mathbb{R}} |y|^p\,\mathbf{N}_{p,\tau}(\mathrm{d}y)\right). \qquad (5.41)
\end{aligned}$$

Thus,

$$\begin{aligned}\nabla_\tau \Lambda_p(\tau(x)) &= \left(\int_{\mathbb{R}} |y|^q \, \mathbf{N}_{p,\tau(x)}(\mathrm{d}y), \int_{\mathbb{R}} |y|^p \, \mathbf{N}_{p,\tau(x)}(\mathrm{d}y)\right) \\ &= \left(\mathbb{E}\Big[|\mathcal{Y}_1^{(n)}|^q\Big], \mathbb{E}\Big[|\mathcal{Y}_1^{(n)}|^p\Big]\right) \\ &= \mathbb{E}\Big[\mathcal{V}_1^{(n)}\Big].\end{aligned}$$

Recalling Lemma 2.2.1 (4), $\tau(x)$ was defined to be the argument, where the supremum of $[\langle x, \tau\rangle - \Lambda_p(\tau)]$ is attained, hence,

$$\nabla_\tau \big[\langle x, \tau(x)\rangle - \Lambda_p(\tau(x))\big] = x - \nabla_\tau \Lambda_p(\tau(x)) = 0,$$

yielding $\nabla_\tau \Lambda_p(\tau(x)) = x$, and thereby $\mathbb{E}[\mathcal{V}_1^{(n)}] = x$. Hence, we see that the $(\mathcal{V}_i^{(n)} - x)$ are centered and $\mathbb{E}[\mathcal{S}^{(n)}] = \sqrt{n}\left(\mathbb{E}[\mathcal{V}_1^{(n)}] - x\right) = 0$, which proves *i)*.

It remains to show that $\mathrm{Cov}[\mathcal{S}^{(n)}] = \mathfrak{H}_x = \mathcal{H}_\tau \Lambda(\tau(x))$. By independence and identical distribution of the $\mathcal{V}_i^{(n)}$ it holds that $\mathrm{Cov}[\mathcal{S}^{(n)}] = \frac{1}{n}\sum_{i=1}^n \mathrm{Cov}[(\mathcal{V}_i^{(n)} - x)] = \mathrm{Cov}[\mathcal{V}_1^{(n)}]$. Hence, we need to show that

$$\left(\mathbb{E}\Big[\Big(\mathcal{V}_1^{(n)}\Big)_j \Big(\mathcal{V}_1^{(n)}\Big)_k\Big] - \mathbb{E}\Big[\Big(\mathcal{V}_1^{(n)}\Big)_j\Big]\mathbb{E}\Big[\Big(\mathcal{V}_1^{(n)}\Big)_k\Big]\right)_{j,k\in\{1,2\}} = \left(\frac{\partial^2}{\partial\tau_k \partial\tau_j}\Lambda_p(\tau(x))\right)_{j,k\in\{1,2\}}.$$

We can reformulate the left-hand side in the above for individual $j, k \in \{1,2\}$ as

a) $(j,k) = (1,1) : \mathbb{E}\left[|\mathcal{Y}_1^{(n)}|^{2q}\right] - \mathbb{E}\left[|\mathcal{Y}_1^{(n)}|^q\right]^2$

b) $(j,k) = (2,2) : \mathbb{E}\left[|\mathcal{Y}_1^{(n)}|^{2p}\right] - \mathbb{E}\left[|\mathcal{Y}_1^{(n)}|^p\right]^2$

c) $(j,k) = (1,2)$ and $(j,k) = (2,1) : \mathbb{E}\left[|\mathcal{Y}_1^{(n)}|^{q+p}\right] - \mathbb{E}\left[|\mathcal{Y}_1^{(n)}|^q\right]\mathbb{E}\left[|\mathcal{Y}_1^{(n)}|^p\right]$.

Now, using the calculations from (5.41), it holds for general $\tau \in \mathrm{Dom}(\Lambda_p)$ that

$$\begin{aligned}\frac{\partial^2}{\partial\tau_1\partial\tau_1}\Lambda_p(\tau) &= \frac{\partial}{\partial\tau_1}\left(e^{-\Lambda_p(\tau)}\int_{\mathbb{R}} |y|^q \, e^{\langle\tau,(|y|^q,|y|^p)\rangle} \, f_{\mathbf{N}_p}(y)\,\mathrm{d}y\right) \\ &= \frac{\partial}{\partial\tau_1}\left(e^{-\Lambda_p(\tau)}\right)\int_{\mathbb{R}} |y|^q \, e^{\langle\tau,(|y|^q,|y|^p)\rangle} \, f_{\mathbf{N}_p}(y)\,\mathrm{d}y \\ &\quad + e^{-\Lambda_p(\tau)}\frac{\partial}{\partial\tau_1}\left(\int_{\mathbb{R}} |y|^q \, e^{\langle\tau,(|y|^q,|y|^p)\rangle} \, f_{\mathbf{N}_p}(y)\,\mathrm{d}y\right)\end{aligned}$$

$$\begin{aligned} &= \frac{\partial}{\partial \tau_1}\left(e^{-\Lambda_p(\tau)}\right) \int\limits_{\mathbb{R}} |y|^q \, e^{\langle \tau, (|y|^q, |y|^p) \rangle} \, f_{\mathbf{N}_p}(y) \, \mathrm{d}y \\ &+ \; e^{-\Lambda_p(\tau)} \int\limits_{\mathbb{R}} |y|^{2q} \, e^{\langle \tau, (|y|^q, |y|^p) \rangle} \, f_{\mathbf{N}_p}(y) \, \mathrm{d}y, \end{aligned}$$

and

$$\begin{aligned} \frac{\partial}{\partial \tau_1}\left(e^{-\Lambda_p(\tau)}\right) &= \frac{\partial}{\partial \tau_1}\left(\int\limits_{\mathbb{R}} e^{\langle \tau, (|y|^q, |y|^p) \rangle} \, f_{\mathbf{N}_p}(y) \, \mathrm{d}y\right)^{-1} \\ &= -\left(\int\limits_{\mathbb{R}} e^{\langle \tau, (|y|^q, |y|^p) \rangle} \, f_{\mathbf{N}_p}(y) \, \mathrm{d}y\right)^{-2} \frac{\partial}{\partial \tau_1} \int\limits_{\mathbb{R}} e^{\langle \tau, (|y|^q, |y|^p) \rangle} \, f_{\mathbf{N}_p}(y) \, \mathrm{d}y \\ &= -\left(\int\limits_{\mathbb{R}} e^{\langle \tau, (|y|^q, |y|^p) \rangle} \, f_{\mathbf{N}_p}(y) \, \mathrm{d}y\right)^{-2} \int\limits_{\mathbb{R}} |y|^q \, e^{\langle \tau, (|y|^q, |y|^p) \rangle} \, f_{\mathbf{N}_p}(y) \, \mathrm{d}y \\ &= -\left(\int\limits_{\mathbb{R}} e^{\langle \tau, (|y|^q, |y|^p) \rangle - \Lambda_p(\tau)} \, f_{\mathbf{N}_p}(y) \, \mathrm{d}y\right)^{-2} e^{-\Lambda_p(\tau)} \\ &\quad \times \int\limits_{\mathbb{R}} |y|^q \, e^{\langle \tau, (|y|^q, |y|^p) \rangle - \Lambda_p(\tau)} \, f_{\mathbf{N}_p}(y) \, \mathrm{d}y \\ &= -\left(\int\limits_{\mathbb{R}} \mathbf{N}_{p,\tau}(\mathrm{d}y)\right)^{-2} e^{-\Lambda_p(\tau)} \int\limits_{\mathbb{R}} |y|^q \, \mathbf{N}_{p,\tau}(\mathrm{d}y) \\ &= -\, e^{-\Lambda_p(\tau)} \int\limits_{\mathbb{R}} |y|^q \, \mathbf{N}_{p,\tau}(\mathrm{d}y). \end{aligned}$$

Hence,

$$\begin{aligned} \frac{\partial^2}{\partial \tau_1 \partial \tau_1} \Lambda_p(\tau) &= \int\limits_{\mathbb{R}} |y|^{2q} \, e^{\langle \tau, (|y|^q, |y|^p) \rangle - \Lambda_p(\tau)} \, f_{\mathbf{N}_p}(y) \, \mathrm{d}y \\ &\quad -\left(\int\limits_{\mathbb{R}} |y|^q \, e^{\langle \tau, (|y|^q, |y|^p) \rangle - \Lambda_p(\tau)} \, f_{\mathbf{N}_p}(y) \, \mathrm{d}y\right)^2 \\ &= \int\limits_{\mathbb{R}} |y|^{2q} \, \mathbf{N}_{p,\tau}(\mathrm{d}y) - \left(\int\limits_{\mathbb{R}} |y|^q \, \mathbf{N}_{p,\tau}(\mathrm{d}y)\right)^2, \end{aligned}$$

which, evaluated at $\tau = \tau(x) \in \mathrm{Dom}(\Lambda_p)$, yields

$$\begin{aligned}
\frac{\partial^2}{\partial\tau_1\partial\tau_1}\,\Lambda_p(\tau(x)) &= \int\limits_{\mathbb{R}} |y|^{2q}\,\mathbf{N}_{p,\tau(x)}(\mathrm{d}y) - \Bigg(\int\limits_{\mathbb{R}} |y|^{q}\,\mathbf{N}_{p,\tau(x)}(\mathrm{d}y)\Bigg)^2 \\
&= \mathbb{E}\Big[\big|\mathcal{Y}^{(n)}\big|^{2q}\Big] - \mathbb{E}\big[\big|\mathcal{Y}^{(n)}\big|^{q}\big]^2.
\end{aligned}$$

Analogously it follows that

$$\frac{\partial}{\partial\tau_2}\,\big(e^{-\Lambda_p(\tau(x))}\big) = e^{-\Lambda_p(\tau(x))}\,\mathbb{E}\Big[\big|\mathcal{Y}_1^{(n)}\big|^{p}\Big],$$

and thus

$$\frac{\partial^2}{\partial\tau_2\partial\tau_2}\,\Lambda_p(\tau(x)) = \mathbb{E}\Big[\big|\mathcal{Y}_1^{(n)}\big|^{p}\Big]^2 + \mathbb{E}\Big[\big|\mathcal{Y}_1^{(n)}\big|^{2p}\Big].$$

Finally, using the derivatives of $e^{-\Lambda_p(\tau)}$ from the previous two cases, yields

$$\begin{aligned}
\frac{\partial^2}{\partial\tau_1\partial\tau_2}\,\Lambda_p(\tau) &= \frac{\partial^2}{\partial\tau_2\partial\tau_1}\,\Lambda_p(\tau) \\
&= \frac{\partial}{\partial\tau_2}\,\big(e^{-\Lambda_p(\tau)}\big)\int\limits_{\mathbb{R}} |y|^{q}\,e^{\langle\tau,(|y|^q,|y|^p)\rangle}\,f_{\mathbf{N}_p}(y)\,\mathrm{d}y \\
&\qquad + e^{-\Lambda_p(\tau)}\,\frac{\partial}{\partial\tau_2}\Bigg(\int\limits_{\mathbb{R}} |y|^{q}\,e^{\langle\tau_2,(|y|^q,|y|^p)\rangle}\,f_{\mathbf{N}_p}(y)\,\mathrm{d}y\Bigg) \\
&= \int\limits_{\mathbb{R}} |y|^{q+p}\,\mathbf{N}_{p,\tau}(\mathrm{d}y) - \Bigg(\int\limits_{\mathbb{R}} |y|^{p}\,\mathbf{N}_{p,\tau}(\mathrm{d}y)\Bigg)\Bigg(\int\limits_{\mathbb{R}} |y|^{q}\,\mathbf{N}_{p,\tau}(\mathrm{d}y)\Bigg).
\end{aligned}$$

Then, for $\tau = \tau(x) \in \mathrm{Dom}(\Lambda_p)$ we get

$$\begin{aligned}
\frac{\partial^2}{\partial\tau_1\partial\tau_2}\,\Lambda_p(\tau(x)) &= \int\limits_{\mathbb{R}} |y|^{q+p}\,\mathbf{N}_{p,\tau(x)}(\mathrm{d}y) - \Bigg(\int\limits_{\mathbb{R}} |y|^{p}\,\mathbf{N}_{p,\tau(x)}(\mathrm{d}y)\Bigg)\Bigg(\int\limits_{\mathbb{R}} |y|^{q}\,\mathbf{N}_{p,\tau(x)}(\mathrm{d}y)\Bigg) \\
&= \mathbb{E}\Big[\big|\mathcal{Y}_1^{(n)}\big|^{q+p}\Big] - \mathbb{E}\Big[\big|\mathcal{Y}_1^{(n)}\big|^{q}\Big]\mathbb{E}\Big[\big|\mathcal{Y}_1^{(n)}\big|^{p}\Big],
\end{aligned}$$

proving *ii)*. □

We proceed to prove the integral approximation in Lemma 5.2.23.

*Proof of Lemma 5.2.23.* By Lemma 5.2.24, we can rewrite the integral $\mathcal{I}^{(n)}(x)$ from (5.38) via the Fourier transform of the density $g_x^{(n)}$ of $\mathcal{G}^{(n)}$ as

$$\mathcal{I}^{(n)}(x) = \int_{\mathbb{R}^2} e^{-n\langle it,x\rangle} \left(\frac{\varphi_p(\tau(x)+it)}{\varphi_p(\tau(x))}\right)^n \mathrm{d}t = \int_{\mathbb{R}^2} \mathcal{F}\left(g_x^{(n)}\right)(t)\,\mathrm{d}t.$$

It remains to show that

$$\int_{\mathbb{R}^2} \mathcal{F}\left(g_x^{(n)}\right)(t)\,\mathrm{d}t = \left(\frac{n}{2\pi}\right)^{-1} (\det \mathfrak{H}_x)^{-1/2}\,(1+o(1)).$$

To do so, we will show that most of the mass of the integral is concentrated in a neighbourhood of the origin, outside of which it drops exponentially in $n$. It holds that

$$\frac{\varphi_p(\tau(x)+it)}{\varphi_p(\tau(x))} = e^{-\Lambda_p(\tau(x))} \int_{\mathbb{R}} e^{\langle \tau(x)+it,(|y|^q,|y|^p)\rangle} f_{\mathbf{N}_p}(y)\,\mathrm{d}y = \mathbb{E}\left[e^{\langle it,\mathcal{V}_1^{(n)}\rangle}\right],$$

i.e., the term can also be writen as a Fourier transform. Hence, for $t \in \mathbb{R}$ we have

$$\left|\frac{\varphi_p(\tau(x)+it)}{\varphi_p(\tau(x))}\right| = \left|\mathbb{E}\left[e^{\langle it,\mathcal{V}_1^{(n)}\rangle}\right]\right| \le \mathbb{E}\left[\left|e^{\langle it,\mathcal{V}_1^{(n)}\rangle}\right|\right] = 1, \tag{5.42}$$

with $\mathbb{E}[e^{\langle it,\mathcal{V}_1^{(n)}\rangle}] = 1$ only for $t = (0,0)$. Also, from the density property of $\mathbf{N}_{p,\tau(x)}$ follows that $e^{\langle \tau(x),(|y|^q,|y|^p)\rangle - \Lambda_p(\tau(x))} f_{\mathbf{N}_p}(y) \in L_1(\mathbb{R}^2)$. By the Riemann-Lebesgue lemma [42, Proposition 2.2.17] we hence know that

$$\lim_{\|t\|\to+\infty} \mathbb{E}\left[e^{\langle it,\mathcal{V}_1^{(n)}\rangle}\right] = 0.$$

Together with (5.42) we thus conclude that for every neighbourhood $B$ of the origin, there is a $C < 1$, such that $\mathbb{E}[e^{\langle it,\mathcal{V}_1^{(n)}\rangle}] < C$ for all $t \in B^c$. Also, we have seen in Lemma 5.2.20, that there is some $s > 1$ such that $\varphi_p(\tau(x)+it) \in L_s(\mathbb{R}^2)$ and the same extends to $\mathbb{E}\big[e^{\langle it,\mathcal{V}_1^{(n)}\rangle}\big]$. Thereby, for some sufficiently large $s_0 > 1$ it follows that

$$\begin{aligned}
\left|\int_{B^c} \mathcal{F}\left(g_x^{(n)}\right)(t)\,\mathrm{d}t\right| &\le \int_{B^c} \left|\mathcal{F}\left(g_x^{(n)}\right)(t)\right|\,\mathrm{d}t \\
&= \int_{B^c} \left|e^{-n\langle it,x\rangle}\left(\frac{\varphi_p(\tau(x)+it)}{\varphi_p(\tau(x))}\right)^n\right|\,\mathrm{d}t \\
&= \int_{B^c} \left|\frac{\varphi_p(\tau(x)+it)}{\varphi_p(\tau(x))}\right|^n\,\mathrm{d}t \\
&\le C^{(n-s_0)} \int_{B^c} \left|\frac{\varphi_p(\tau(x)+it)}{\varphi_p(\tau(x))}\right|^{s_0}\,\mathrm{d}t,
\end{aligned}$$

which goes to zero exponentially fast in $n \in \mathbb{N}$, that is,

$$\int_{B^c} \mathcal{F}\left(g_x^{(n)}\right)(t)\,\mathrm{d}t = o(1). \tag{5.43}$$

We see, that the contribution of the integral outside of a neighbourhood of the origin can be neglected for large $n \in \mathbb{N}$. Hence, it remains to consider the integral over said neighbourhood. By substituting $t$ by $\tilde{t} = \sqrt{n}\,t$, we get

$$\int_B \mathcal{F}\left(g_x^{(n)}\right)(t)\,\mathrm{d}t = \frac{1}{n}\int_{\sqrt{n}B} \mathcal{F}\left(g_x^{(n)}\right)\left(\frac{\tilde{t}}{\sqrt{n}}\right)\mathrm{d}\tilde{t}. \tag{5.44}$$

The integrand on the right-hand side is just the characteristic function of $\frac{1}{\sqrt{n}}\mathcal{G}^{(n)} = \mathcal{S}^{(n)}$. By Lemma 5.2.25 $\mathcal{S}^{(n)}$ is the $\sqrt{n}$-multiple of the empirical average of centered i.i.d. random vectors $\big(\mathcal{V}_i^{(n)} - x\big)$ with covariance matrix $\mathrm{Cov}\big(\mathcal{S}^{(n)}\big) = \mathfrak{H}_x$. Thus, by the central limit theorem, it holds that $\mathcal{S}^{(n)}$ converges in distribution to a centered Gaussian distribution in $\mathbb{R}^n$ with covariance matrix $\mathfrak{H}_x$, denoted as $\mathcal{N}^{(n)}(0,\mathfrak{H}_x)$. Thus, the characteristic function of the distribution of $\mathcal{S}^{(n)}$ will converge pointwise to that of $\mathcal{N}^{(n)}(0,\mathfrak{H}_x)$, i.e.,

$$\lim_{n\to\infty} \mathcal{F}\left(g_x^{(n)}\right)\left(\frac{\tilde{t}}{\sqrt{n}}\right) = \exp\left(-\frac{1}{2}\left\langle \mathfrak{H}_x\,\tilde{t},\tilde{t}\,\right\rangle\right). \tag{5.45}$$

To use the above on (5.44), we show the conditions of the dominated convergence theorem. Using Taylor expansion of $\mathcal{F}\left(g_x^{(n)}\right)\left(\frac{\tilde{t}}{\sqrt{n}}\right)$ around the origin (see, e.g., [67, Lemma 4.10]), we have that for $\alpha \in \mathbb{N}_0^2$, $k \in \mathbb{N}_0$, and $t$ in a neighbourhood of the origin

$$\mathcal{F}\left(g_x^{(n)}\right)\left(\frac{\tilde{t}}{\sqrt{n}}\right) = \sum_{||\alpha||_1 \le k} \frac{(i\tilde{t}\,)^{\alpha}}{\alpha!}\,\mathbb{E}\big[\mathcal{S}(n)^{\alpha}\big] + o\big(\tilde{t}^{\,k}\big),$$

using the multi-index notation $t^{\alpha} = t_1^{\alpha_1} t_2^{\alpha_2}$ and $\alpha! = \alpha_1!\,\alpha_2!$. For $k = 2$, by Lemma 5.2.25, this gives

$$\begin{aligned}
&\mathcal{F}\big(g_x^{(n)}\big)\left(\frac{\tilde{t}}{\sqrt{n}}\right)\\
&= 1 + (i\tilde{t}_1)\,\mathbb{E}\Big[\Big(\mathcal{S}^{(n)}\Big)_1\Big] + (i\tilde{t}_2)\,\mathbb{E}\Big[\Big(\mathcal{S}^{(n)}\Big)_2\Big]\\
&\quad + \frac{(i\tilde{t}_1)^2}{2}\,\mathbb{E}\left[\Big(\mathcal{S}^{(n)}\Big)_1^2\right] + \frac{(i\tilde{t}_2)^2}{2}\,\mathbb{E}\left[\Big(\mathcal{S}^{(n)}\Big)_2^2\right] + (i\tilde{t}_1)(i\tilde{t}_2)\,\mathbb{E}\Big[\Big(\mathcal{S}^{(n)}\Big)_1\Big(\mathcal{S}^{(n)}\Big)_2\Big] + o\big(\tilde{t}^{\,k}\big)\\
&= 1 - \frac{1}{2}\left(\tilde{t}_1^2\,\mathbb{E}\left[\Big(\mathcal{V}_1^{(n)}\Big)_1^2\right] + \tilde{t}_2^2\,\mathbb{E}\left[\Big(\mathcal{V}_1^{(n)}\Big)_2^2\right] + 2\,\tilde{t}_1\tilde{t}_2\,\mathbb{E}\Big[\Big(\mathcal{V}_1^{(n)}\Big)_1\Big(\mathcal{V}_1^{(n)}\Big)_2\Big]\right) + o\big(\tilde{t}^{\,k}\big)\\
&= 1 - \frac{1}{2}\big\langle \mathfrak{H}_x\,\tilde{t},\tilde{t}\,\big\rangle + o\big(\tilde{t}^{2}\big).
\end{aligned} \tag{5.46}$$

By [67, Lemma 4.14] and the Jensen inequality for $\langle \tilde{t}, \mathcal{S}^{(n)} \rangle$, we can give the following upper bound on the error term

$$\begin{aligned}
\left| \mathcal{F}\big(g_x^{(n)}\big)\left(\frac{\tilde{t}}{\sqrt{n}}\right) - 1 + \frac{1}{2}\left\langle \mathfrak{H}_x \tilde{t}, \tilde{t}\,\right\rangle \right| &= \left| \mathbb{E}\Big[e^{i\langle \tilde{t}, \mathcal{S}^{(n)}\rangle}\Big] - \sum_{||\alpha||_1 \le 2} \frac{\big(i\tilde{t}\,\big)^{\alpha}}{\alpha!}\, \mathbb{E}\Big[\big(\mathcal{S}^{(n)}\big)^{\alpha}\Big] \right| \\
&\le \mathbb{E}\left[\left| e^{i\langle \tilde{t}, \mathcal{S}^{(n)}\rangle} - \sum_{||\alpha||_1 \le 2} \frac{\big(i\tilde{t}\,\big)^{\alpha}}{\alpha!} \big(\mathcal{S}^{(n)}\big)^{\alpha} \right|\right] \\
&\le \mathbb{E}\Big[\, \big|\langle \tilde{t}, \mathcal{S}^{(n)}\rangle\big|^2 \Big] \\
&\le \mathbb{E}\Big[\, \big\|\tilde{t}\,\big\|_2^2\, \big\|\mathcal{S}^{(n)}\big\|_2^2 \Big] \\
&= \big\|\tilde{t}\,\big\|_2^2\, \mathbb{E}\Big[\, \big\|\mathcal{S}^{(n)}\big\|_2^2 \Big]. \qquad (5.47)
\end{aligned}$$

The next step is to show that $\mathbb{E}\Big[\, \big\|\mathcal{S}^{(n)}\big\|_2^2 \Big]$ is bounded, which will be done along the lines of [85, Lemma 6.3]. It holds that

$$\begin{aligned}
\mathbb{E}\Big[\, \big\|\mathcal{S}^{(n)}\big\|_2^2 \,\Big] &= \mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \Big(\mathcal{V}_i^{(n)} - x\Big)_1\right)^2 + \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \Big(\mathcal{V}_i^{(n)} - x\Big)_2\right)^2\right] \\
&= \frac{1}{n}\,\mathbb{E}\left[\left(\sum_{i=1}^{n} \Big(\mathcal{V}_i^{(n)} - x\Big)_1\right)^2\right] + \frac{1}{n}\,\mathbb{E}\left[\left(\sum_{i=1}^{n} \Big(\mathcal{V}_i^{(n)} - x\Big)_2\right)^2\right],
\end{aligned}$$

where, by independence of the $\mathcal{V}_i^{(n)}$ and Lemma 5.2.25, we have for $k \in \{1, 2\}$ that

$$\begin{aligned}
&\frac{1}{n}\,\mathbb{E}\left[\left(\sum_{i=1}^{n} \Big(\mathcal{V}_i^{(n)} - x\Big)_k\right)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[\left(\Big(\mathcal{V}_i^{(n)} - \mathbb{E}\Big[\mathcal{V}_i^{(n)}\Big]\Big)_k\right)^2\right] \\
&\quad + \frac{1}{n}\sum_{1\le i<j\le n} \mathbb{E}\Big[\Big(\mathcal{V}_i^{(n)} - \mathbb{E}\Big[\mathcal{V}_i^{(n)}\Big]\Big)_k\Big]\, \mathbb{E}\Big[\Big(\mathcal{V}_j^{(n)} - \mathbb{E}\Big[\mathcal{V}_j^{(n)}\Big]\Big)_k\Big] \\
&= (\mathfrak{H}_x)_{kk} + \frac{1}{n}\sum_{1\le i<j\le n} \mathrm{Cov}\Big[\Big(\mathcal{V}_i^{(n)}\Big)_k, \Big(\mathcal{V}_j^{(n)}\Big)_k\Big] \\
&= (\mathfrak{H}_x)_{kk}\,.
\end{aligned}$$

Thus, $\mathbb{E}\Big[\big\|\mathcal{S}^{(n)}\big\|_2^2\Big]$ is bounded, and thereby the same holds for the error term in (5.47).

Recall, that $\mathfrak{H}_x := \mathcal{H}_\tau \Lambda_p(\tau(x))$ is positive definite, as it is invertible by Lemma 5.2.1 and positive semi-definite by the convexity of $\Lambda_p$ on its effective domain. For $n \in \mathbb{N}$ sufficiently large, we can now always choose a small enough neighbourhood $B$ of the origin, such that there is an $\varepsilon > 0$ with

$$\frac{1}{2}\left\langle \varepsilon\, I_2\, \tilde{t}, \tilde{t}\right\rangle \geq \left\|\tilde{t}\right\|_2^2 \mathbb{E}\left[\left\|\mathcal{S}^{(n)}\right\|_2^2\right],$$

and $\mathfrak{H}_x - \varepsilon\, I_2$ positive definite, where $I_2$ denotes the $(2 \times 2)$ identity matrix in $\mathbb{R}^2$. Together with the well-known inequality $1 + x \leq e^x$, $x \in \mathbb{R}$, this yields with 5.46 that

$$\begin{aligned}
\mathcal{F}\left(g_x^{(n)}\right)\left(\frac{\tilde{t}}{\sqrt{n}}\right) &= 1 - \frac{1}{2}\left\langle \mathfrak{H}_x\, \tilde{t}, \tilde{t}\right\rangle + o\big(\tilde{t}^2\big) \\
&\leq 1 - \frac{1}{2}\left\langle \mathfrak{H}_x\, \tilde{t}, \tilde{t}\,\right\rangle + \frac{1}{2}\left\langle \varepsilon\, I_2\, \tilde{t}, \tilde{t}\,\right\rangle \\
&= 1 - \frac{1}{2}\left\langle (\mathfrak{H}_x - \varepsilon\, I_2)\, \tilde{t}, \tilde{t}\,\right\rangle \\
&\leq \exp\left(-\frac{1}{2}\left\langle (\mathfrak{H}_x - \varepsilon\, I_2)\, \tilde{t}, \tilde{t}\,\right\rangle\right).
\end{aligned}$$

Furthermore, it holds that

$$\int_{\mathbb{R}^2} \exp\left(-\frac{1}{2}\left\langle (\mathfrak{H}_x - \varepsilon\, I_2)\, \tilde{t}, \tilde{t}\,\right\rangle\right) \mathrm{d}\tilde{t} = \frac{2\pi}{\big(\det(\mathfrak{H}_x - \varepsilon\, I_2)\big)^{1/2}},$$

i.e., the function is integrable. The conditions of the dominated convergence theorem are thus fulfilled, and thereby it follows with (5.45) that

$$\begin{aligned}
\int_{\sqrt{n}B} \mathcal{F}\left(g_x^{(n)}\right)\left(\frac{\tilde{t}}{\sqrt{n}}\right) \mathrm{d}\tilde{t} &= \int_{\mathbb{R}^2} \exp\left(-\frac{1}{2}\langle \mathfrak{H}_x\, \tilde{t}, \tilde{t}\,\rangle\right) \mathrm{d}\tilde{t}\,(1 + o(1)) \\
&= \frac{2\pi}{\big(\det \mathfrak{H}_x\big)^{1/2}}\,(1 + o(1)). \qquad (5.48)
\end{aligned}$$

Combining (5.43), (5.44), and (5.48) to get

$$\begin{aligned}
\int_{\mathbb{R}^2} \mathcal{F}\left(g_x^{(n)}\right)(t)\,\mathrm{d}t &= \int_{B} \mathcal{F}\left(g_x^{(n)}\right)(t)\,\mathrm{d}t + \int_{B^c} \mathcal{F}\left(g_x^{(n)}\right)(t)\mathrm{d}t \\
&= \frac{1}{n}\int_{\sqrt{n}B} \mathcal{F}\left(g_x^{(n)}\right)\left(\frac{\tilde{t}}{\sqrt{n}}\right) \mathrm{d}\tilde{t} + o(1) \\
&= \left(\frac{n}{2\pi}\right)^{-1}\big(\det \mathfrak{H}_x\big)^{-1/2}\,(1 + o(1))
\end{aligned}$$

finishes the proof of Lemma 5.2.23. □

Combining Proposition 5.2.22 and Lemma 5.2.23 then yields the asymptotic density estimate in Proposition 5.2.19. We shall now consider the case of $\ell_p^n$-balls, that is, derive an asymptotic density estimate for $\mathscr{S}^{(n)}$ as in (5.23).

**Proposition 5.2.26** *For $\mathscr{S}^{(n)} = \frac{1}{n}\sum_{i=1}^n \mathscr{V}_i^{(n)}$ with $\mathscr{V}_i^{(n)} = \big(|Y_i^{(n)}|^q, |Y_i^{(n)}|^p, U^{1/n}\big)$, $Y_i^{(n)} \sim \mathbf{N}_p$ i.i.d., $U$ uniformly distributed on $[0,1]$ independently of the $Y_i^{(n)}$, and $x = (x_1, x_2) \in \mathrm{Dom}(\Lambda_p^*)$, $y \in (0,1]$, it holds that for sufficiently large $n \in \mathbb{N}$ the distribution of $\mathscr{S}^{(n)}$ has Lebesgue density*

$$\hbar^{(n)}(x_1, x_2, y) = \frac{n^2}{2\pi}\, y^{-1} (\det \mathfrak{H}_x)^{-1/2}\, e^{-n\,\mathcal{I}_{\mathscr{S}}(x_1,x_2,y)}\,(1+o(1)),$$

*where $\mathcal{I}_{\mathscr{S}}(x_1,x_2,y) := \big[\Lambda_p^*(x) - \log(y)\big]$ and $\mathfrak{H}_x := \mathcal{H}_\tau \Lambda_p(\tau(x))$ as in (5.10).*

*Proof.* By direct calculation we can see for $y \in [0,1]$ that $\mathbb{P}\big(U^{1/n} \le y\big) = \mathbb{P}(U \le y^n) = y^n$, hence the density of $U^{1/n}$ is given by $f_{U^{1/n}}(y) = n\,y^{n-1}$. As $U^{1/n}$ is independent of the $Y_i^{(n)}$, and thereby also of $S^{(n)} = \frac{1}{n}\sum_{i=1}^n \big(|Y_i^{(n)}|^q, |Y_i^{(n)}|^p\big)$, the density of $\mathscr{S}^{(n)} = \frac{1}{n}\sum_{i=1}^n \big(|Y_i^{(n)}|^q, |Y_i^{(n)}|^p, U^{1/n}\big)$ is given by the product of their densities, hence

$$\hbar^{(n)}(x_1,x_2,y) = h^{(n)}(x_1,x_2) f_{U^{1/n}}(y) = \frac{n^2}{2\pi}\, y^{-1}\big(\det \mathfrak{H}_x\big)^{-1/2}\, e^{-n\left[\Lambda_p^*(x) - \log(y)\right]}\,(1+o(1)).$$

This completes the proof. □

### 5.2.8 Proof of the SLD results for $\ell_p^n$-spheres

The third and final step to prove Theorem 5.2.11 is to calculate the integral in (5.25) over the deviation area $D_z$ using the density estimates from Proposition 5.2.19. Based on the density estimate, one can tell that the integral in (5.25) is of Laplace-type, i.e.,

$$\int_{D_z} h^{(n)}(x)\,\mathrm{d}x = \frac{n}{2\pi}\int_{D_z} \big(\det \mathfrak{H}_x\big)^{-1/2}\, e^{-n\,\Lambda_p^*(x)}\,\mathrm{d}x\,(1+o(1)).$$

Hence, for large $n \in \mathbb{N}$ one would assume it behaves like the integrand evaluated at the infimum of $\Lambda_p^*$. However, as we have seen in Lemma 5.2.18, the infimum of $\Lambda_p^*$ over $\bar{D}_z$ is attained at $z^* = (z^q, 1)$ and lies on the boundary $\partial D_z$ of the area of integration, which needs to be accounted for. This is done by a result of Andriani and Baldi [7], which construes the boundary of the deviation area $\partial D_z$ and the level sets of $\Lambda_p^*$ as hypersurfaces (which are just planar curves in our setting), and uses their Weingarten maps (i.e., absolute value of their curvature) to give a Laplace-type integral approximation with the critical point lying on the boundary of the area of integration.

The resulting factors accounting for the minimizer $z^*$ of $\Lambda_p^*$ lying on $\partial D_z$ will partly make up the prefactor functions from Theorem 5.2.11. This will thereby confirm what was said in Remark 5.2.17, that is, that the geometric shape of $D_z$ – which in itself is a result of the functional structure of the quantity being considered – indeed does have an influence on the sharp large deviation result. So while this influence is very direct for LDPs via the infimum operator from the contraction principle, for sharp large deviations it is considerably more subtle in the form of the absolute curvature of the boundary of the deviation area.

Similar to the proof of the density estimate itself, the proof will again rely on splitting up the integral onto some domain where the mass of the integral concentrates and its complement on which the integral is negligible. Specifically, the integral will be split up into a neighbourhood $B_z$ of $z^*$ and its complement $B_z^c$. The LDP from Proposition 5.2.2 will be used to show the comparative negligibility of the integral on $B_z^c$. On $B_z$ we then apply the Laplace-integration result of Andriani and Baldi [7]. Following that, we derive the Weingarten maps used therein explicitly, thus finishing the proof.

*Proof of Theorem 5.2.11.* We assume the set-up of Theorem 5.2.11 and use the reformulation (5.25) to proceed by considering $\mathbb{P}\big(S^{(n)} \in D_z\big)$. Let $B_z \subset \mathbb{R}^2$ be an open neighbourhood around $z^*$, small enough that $B_z \subset \mathrm{Dom}(\Lambda_p^*)$. Then it holds that

$$\mathbb{P}\big(S^{(n)} \in D_z\big) = \int\limits_{D_z} h^{(n)}(x)\,\mathrm{d}x = \int\limits_{D_z \cap B_z} h^{(n)}(x)\,\mathrm{d}x + \int\limits_{D_z \cap B_z^c} h^{(n)}(x)\,\mathrm{d}x. \tag{5.49}$$

Since $z^* \notin B_z^c$, by Lemma 5.2.18 , there exists an $\eta > 0$, such that

$$\inf_{y \in D_z \cap B_z^c} \Lambda_p^*(y) > \Lambda_p^*(z^*) + \eta,$$

and thus, by the LDP in Proposition 5.2.2, it follows

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(S^{(n)} \in D_z \cap B_z^c\big) \leq -\inf_{y \in D_z \cap B_z^c} \Lambda_p^*(y) \leq -\Lambda_p^*(z^*) - \eta.$$

This gives

$$\mathbb{P}\left(S^{(n)} \in D_z \cap B_z^c\right) \leq e^{-n\,\Lambda_p^*(z^*) - n\,\eta}\,(1 + o(1)) = \frac{1}{e^{n\,\eta}}\, e^{-n\Lambda_p^*(z^*)}(1 + o(1)). \tag{5.50}$$

Furthermore, by the density estimate in Proposition 5.2.19, we have

$$\int\limits_{D_z \cap B_z} h^{(n)}(x)\,\mathrm{d}x = \frac{n}{2\pi} \int\limits_{D_z \cap B_z} \big(\det \mathfrak{H}_x\big)^{-1/2}\, e^{-n\,\Lambda_p^*(x)}\,\mathrm{d}x\,(1 + o(1)). \tag{5.51}$$

As stated above, to calculate this explicitly we rely on a technique established by Andriani and Baldi in [7, Proof of Theorem 4.4]. Therein, an asymptotic integral expansion of Bleistein and Handelsmann [15, Equation (8.3.63)] for Laplace integrals with critical point on the boundary of the area of integration is reformulated via the Weingarten maps of the integration area's boundary and the level set of the exponential function at its critical point, both seen as hypersurfaces. We will present it as one concise result, similar to that formulated in [85, Lemma 5.6].

**Proposition 5.2.27** *Let $D \subset \mathbb{R}^d$ be a bounded domain such that $\partial D$ is a differentiable hypersurface in $\mathbb{R}^d$. Furthermore, let $g : \mathbb{R}^d \to \mathbb{R}$ be a differentiable function and $\phi : D \to [0,\infty)$ a non-negative function that is twice differentiable and attains a unique infimum over $\overline{D}$ at $x^* \in \partial D$. Define the hypersurfaces*

$$\mathscr{C}_D = \partial D \qquad \text{and} \qquad \mathscr{C}_\phi = \{x \in \mathbb{R}^d : \phi(x) = \phi(x^*)\},$$

*and denote by $L_D$ and $L_\phi$ their respective Weingarten maps at $x^*$. Then, for sufficiently large $n \in \mathbb{N}$, it holds that*

$$\int_D g(x)\, e^{-n\,\phi(x)}\,\mathrm{d}x = \frac{(2\pi)^{(d-1)/2}\,\det\big(L_\phi^{-1}\big(L_\phi - L_D\big)\big)^{-1/2}}{n^{(d+1)/2}\big\langle \mathfrak{H}_x\,\phi(x^*)^{-1}\,\nabla_x\phi(x^*), \nabla_x\phi(x^*)\big\rangle^{1/2}}\, g(x^*)\, e^{-n\,\phi(x^*)}(1+o(1)).$$

The proof of this is given by first applying the result from [15, Equation (8.3.63)] for Laplace-type integrals and then using the reformulation of the terms therein from [7, Equation (4.6)] with respect to the Weingarten map.

We shall check that the above conditions hold for the integral in (5.51). $D_z \cap B_z$ is bounded and for $z > m_{p,q}$, we can write $\partial D_z$ as the graph of the infinitely differentiable function $f : (0,\infty) \to (0,\infty)$ with $f(t_1) = z^{-p}\, t_1^{p/q}$ (see (5.24)), thus both $\partial D_z$ and $\partial(D_z \cap B_z)$ are differentiable planar curves. For $B_z$ chosen small enough such that $D_z \cap B_z \subset \mathrm{Dom}(\Lambda_p^*)$, by Lemma 2.2.1 and (5.9),it follows that $\Lambda_p^*$ is twice differentiable on $D_z \cap B_z$. From Lemma 2.2.1 we also know that $\langle x, \tau\rangle - \Lambda_p(\tau)$ has a unique argument $\tau(x)$ of its supremum in $\tau$, i.e., $x - \nabla_\tau \Lambda_p(\tau) = 0$ has a unique solution in $(x, \tau(x))$. Further, by Lemma 2.2.1 and Lemma 5.2.1 we have that $\mathcal{H}_\tau \Lambda_p(\tau)$ is invertible for all $\tau \in \mathrm{Dom}(\Lambda_p)$. Thus, it follows from the implicit function theorem that $x \mapsto \tau(x)$ is as differentiable in $x$ as $(x,\tau) \mapsto (x - \nabla_\tau \Lambda_p(\tau))$ is in $\tau$, yielding that $\tau(x)$ is infinitely differentiable on $\mathrm{Dom}(\Lambda_p^*)$. Hence, we get that $\big(\det \mathfrak{H}_x\big)^{-1/2} = \big(\det \mathcal{H}_\tau \Lambda_p(\tau(x)\big)^{-1/2}$ is differentiable in $x$. Lastly, by Lemma 5.2.18, $z^* \in \partial(D_z \cap B_z)$ is the unique argument at which the infimum of $\Lambda_p^*$ on $\overline{D}_z$ and $\overline{D_z \cap B_z}$ is attained.

Thus, in view of the above, we can use Proposition 5.2.27 for $D = D_z \cap B_z \subset \mathbb{R}^2$ with $g(x) = \big(\det \mathfrak{H}_x\big)^{-1/2}$, $\phi(x) = \Lambda_p^*(x)$, and $x^* = z^*$, to get that

$$\begin{aligned}&\int\limits_{D_z \cap B_z} h^{(n)}(x)\,\mathrm{d}x \\ &= \frac{n}{2\pi}\,\frac{(2\pi)^{1/2}\,\det\!\big(L_\Lambda^{-1}\big(L_\Lambda - L_D\big)\big)^{-1/2}\,(\det \mathfrak{H}_{z^*})^{-1/2}\,e^{-n\,\Lambda_p^*(z^*)}}{n^{3/2}\,\big\langle \mathfrak{H}_x\,\Lambda_p^*(z^*)^{-1}\,\nabla_x\,\Lambda_p^*(z^*), \nabla_x\,\Lambda_p^*(z^*)\big\rangle^{1/2}}\,(1+o(1)), \end{aligned} \tag{5.52}$$

for the respective Weingarten maps $L_D$ and $L_\Lambda$ at $z^*$ of the curves

$$\mathscr{C}_D = \partial\big(D_z \cap B_z\big) \qquad \text{and} \qquad \mathscr{C}_\Lambda = \big\{x \in \mathbb{R}^2 : \Lambda_p^*(x) = \Lambda_p^*(z^*)\big\}.$$

Via Lemma 5.2.1 we get

$$\Big\langle \mathfrak{H}_x\,\Lambda_p^*(z^*)^{-1}\,\nabla_x\,\Lambda_p^*(z^*), \nabla_x\,\Lambda_p^*(z^*)\Big\rangle = \Big\langle \mathfrak{H}_{z^*}\,\tau(z^*), \tau(z^*)\Big\rangle.$$

With the definition of $\xi(z)^2$ in (5.18) the integral in (5.52) hence simplifies to

$$\int\limits_{D_z \cap B_z} h^{(n)}(x)\,\mathrm{d}x = \frac{1}{\sqrt{2\pi n}\,\xi(z)}\,\big(\det(L_\Lambda^{-1}(L_\Lambda - L_D)\big)^{-1/2}\,e^{-n\,\Lambda_p^*(z^*)}\,(1+o(1)). \tag{5.53}$$

It only remains to prove that $\det\!\big(L_\Lambda^{-1}\big(L_\Lambda - L_D\big)\big) = \kappa(z)^2$. We proceed to calculate the Weingarten maps of the curves $\mathscr{C}_D$ and $\mathscr{C}_\Lambda$ explicitly. As discussed in Section 5.2.2, the Weingarten map of a planar curve at a point $x$ reduces to the absolute value of its curvature in $x$. As previously mentioned, $\partial D_z$ is the graph of the function $f(t_1) = z^{-p} t_1^{p/q}$. Thus, the same holds locally for $\mathscr{C}_D = \partial(D_z \cap B_z)$ in a neighbourhood of $z^*$, so the curvature formula for graphs of functions in Corollary 5.2.10 ii) gives

$$L_D = \frac{|f''(z^q)|}{\big(1 + f'(z^q)^2\big)^{3/2}},$$

where

$$f'(t_1)^2 = \big(pq^{-1}\,z^{-p}\,{t_1}^{p/q-1}\big)^2 \Rightarrow f'(z^q)^2 = p^2 q^{-2} z^{-2q},$$

and

$$f''(t_1) = pq^{-1}\,\big(pq^{-1} - 1\big)\,z^{-p}\,{t_1}^{p/q-2} \Rightarrow f''(z^q) = \big(p^2 - pq\big)q^{-2}\,z^{-2q}.$$

This yields

$$L_D = \frac{\big|\big(p^2 - pq\big)q^{-2}\,z^{-2q}\big|}{\big(1 + p^2 q^{-2} z^{-2q}\big)^{3/2}} = \frac{\big|pq(p-q)z^q\big|}{\big(z^{2q} + p^2 q^{-2}\big)^{3/2}}. \tag{5.54}$$

The curve $\mathscr{C}_\Lambda$ is the zero set of the function $F(x) := \Lambda_p^*(x) - \Lambda_p^*(z^*)$. From Lemma 5.2.1 we know that

$$\big(F_{[1,0]}, F_{[0,1]}\big) = \left(\frac{\partial}{\partial x_1}\Lambda_p^*(z^*),\ \frac{\partial}{\partial x_2}\Lambda_p^*(z^*)\right) = \tau(z^*),$$

and

$$\begin{pmatrix} F_{[2,0]} & F_{[1,1]} \\ F_{[1,1]} & F_{[0,2]} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial^2}{\partial x_1^2}\Lambda_p^*(z^*) & \dfrac{\partial^2}{\partial x_2 \partial x_1}\Lambda_p^*(z^*) \\[2ex] \dfrac{\partial^2}{\partial x_1 \partial x_2}\Lambda_p^*(z^*) & \dfrac{\partial^2}{\partial x_2^2}\Lambda_p^*(z^*) \end{pmatrix} = \mathfrak{H}_{z^*}^{-1},$$

for derivatives $F_{[i,j]} = F_{[i,j]}(z^*)$ as in (2.1). Hence, by the curvature formula for implicit curves from Lemma 5.2.9 and Corollary 5.2.10 i), we get

$$L_\Lambda = \frac{\left|\tau(z^*)_2^2 \left(\mathfrak{H}_{z^*}^{-1}\right)_{11} - 2\tau(z^*)_1\tau(z^*)_2 \left(\mathfrak{H}_{z^*}^{-1}\right)_{12} + \tau(z^*)_1^2 \left(\mathfrak{H}_{z^*}^{-1}\right)_{22}\right|}{\left(\tau(z^*)_1^2 + \tau(z^*)_2^2\right)^{3/2}}. \tag{5.55}$$

Since both $L_D$ and $L_\Lambda$ are one-dimensional, it follows from (5.54) and (5.55) that

$$\det(L_\Lambda^{-1}(L_\Lambda - L_D)) = L_\Lambda^{-1}(L_\Lambda - L_D) = 1 - \frac{L_D}{L_\Lambda} = \kappa(z)^2.$$

for $\kappa(z)^2$ as in (5.19). It now follows with (5.53) that

$$\int\limits_{D_z \cap B_z} h^{(n)}(x)\,\mathrm{d}x = \frac{1}{\sqrt{2\pi n}\,\xi(z)\,\kappa(z)}\, e^{-n\Lambda_p^*(z^*)}\,(1 + o(1)). \tag{5.56}$$

Comparing (5.56) with the upper bound of the integral outside of $B_z$ in (5.50), we can see that the integral over $B_z^c$ is negligible for large $n \in \mathbb{N}$. Thus, combining (5.49), (5.50), and (5.56) finishes the proof of Theorem 5.2.11. □

### 5.2.9 Proof of the SLD results for $\ell_p^n$-balls

We now turn to proving Theorem 5.2.12, that is, consider $\|\mathscr{X}\|$ distributed in the $\ell_p^n$-ball according to $\mathbf{U}_{n,p}$. We assume the set-up of Theorem 5.2.12 and proceed similarly to the previous proof, using the reformulation of the target probability as a density integral over $\mathscr{D}_z$ from (5.26) in conjunction with the density approximation from Proposition 5.2.26. The resulting integral over $\mathscr{D}_z$ is again split into a neighbourhood of the minimizer of $\mathcal{I}_{\mathscr{S}}$ over $\overline{\mathscr{D}}_z$ and its complement, which, according to Lemma 5.2.18, is attained at $z^{**} = (z^q, 1, 1)$. However, in this setting the integral approximation of Andriani and Baldi [7] from Proposition 5.2.27 is not applicable, as certain differentiability conditions are no longer met. Hence, for the integral within that neighbourhood we apply a result of Breitung and Hohenbichler [19], which yields a Laplace integral approximation under less restrictive differentiability conditions. This result is again geometric in nature, as the behaviour of the density on $\partial\mathscr{D}_z$ still heavily dictates the value of the overall approximation. However, since this result is formulated for a certain neighbourhood of the origin, we first need to construct a sufficient transformation, mapping our deviation area onto such a neighbourhood. After that, we calculate the specific approximation in the setting of $\ell_p^n$-balls.

*Proof of Theorem 5.2.12.* We assume the set-up of Theorem 5.2.12 and use the reformulation (5.26) to proceed by considering $\mathbb{P}\left(\mathscr{S}^{(n)} \in \mathscr{D}_z\right)$. Let $\mathscr{B}_z \subset \mathbb{R}^3$ be an open neighbourhood around $z^{**} = (z^q, 1, 1)$ small enough such that the first two coordinates of points within $\mathscr{B}_z$ lie in $\mathrm{Dom}(\Lambda_p^*)$ and the third is positive. Then it holds that

$$\mathbb{P}\left(\mathscr{S}^{(n)} \in \mathscr{D}_z\right) = \int\limits_{\mathscr{D}_z \cap \mathscr{B}_z} \hbar^{(n)}(x, y)\,\mathrm{d}x\,\mathrm{d}y \;+ \int\limits_{\mathscr{D}_z \cap \mathscr{B}_z^c} \hbar^{(n)}(x, y)\,\mathrm{d}x\,\mathrm{d}y, \tag{5.57}$$

for $x = (x_1, x_2) \in \mathbb{R}^2$ and $y \in (0, 1]$. As in the proof of Theorem 5.2.11, it follows from Lemma 5.2.18 ii) and the LDP in Proposition 5.2.5 that there is an $\eta > 0$, such that

$$\mathbb{P}\left(\mathscr{S}^{(n)} \in \mathscr{D}_z \cap \mathscr{B}_z^c\right) \le e^{-n\,\mathcal{I}_{\mathscr{S}}(z^{**}) - n\eta}\,(1 + o(1)) = \frac{1}{e^{n\,\eta}}\,e^{-n\,\Lambda_p^*(z^*)}(1 + o(1)), \tag{5.58}$$

with $\mathcal{I}_{\mathscr{S}}(t) = \left[\Lambda_p^*(t_1, t_2) - \log(t_3)\right]$ as defined in Lemma 5.2.6. We will again use this to show the comparative negligibility of the integral over $\mathscr{D}_z \cap \mathscr{B}_z^c$.

Let us now consider the first integral in (5.57). Since $z^* \in \mathrm{Dom}(\Lambda_p^*)$, for sufficiently small $\mathscr{B}_z$, we have that $x = (x_1, x_2) \in \mathrm{Dom}(\Lambda_p^*)$ and $y \in (0, 1]$. By the density approximation from Proposition 5.2.26, it then holds that

$$\begin{aligned}\int\limits_{\mathscr{D}_z \cap B_z} &\hbar^{(n)}(x_1, x_2, y)\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}y \\ &= \frac{n^2}{2\pi} \int\limits_{\mathscr{D}_z \cap B_z} y^{-1}\left(\det \mathfrak{H}_x\right)^{-1/2} e^{-n\,\mathcal{I}_{\mathscr{S}}(x_1, x_2, y)}\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}y\,(1 + o(1)).\end{aligned}$$

As we have seen in Lemma 5.2.18, $\mathcal{I}_{\mathscr{S}}$ attains its infimum on $\overline{\mathscr{D}}_z$ at $z^{**}$. However, we cannot use the result of Andriani and Baldi from Proposition 5.2.27 here, since at $z^{**}$ the boundary of $\mathscr{D}_z \cap \mathscr{B}_z$ is not differentiable, and thereby not a smooth planar curve.

As a substitute we use the following asymptotic integral approximation result based on Breitung and Hohenbichler [19, Lemma 4], which gives a Laplace integral approximation very similar to that in Liao and Ramanan [85, v2, Lemma 5.1], but under weaker conditions. Note that this version of [19, Lemma 4] is not to be confused with that in Proposition 3.1.4, where the setting was simpler and merely one-dimensional.

**Proposition 5.2.28** *Let $F \subset \mathbb{R}^3$ be a compact set containing the origin in its interior. If*

*(a) $f : F \to \mathbb{R}$ and $g : F \to \mathbb{R}$ are continuous functions with $g(\mathbf{0}) \neq 0$, where $\mathbf{0} := (0, 0, 0)$,*

*(b) $f(x) > f(\mathbf{0})$ for all $x \in F \cap \left(\mathbb{R}_+^2 \times \mathbb{R}\right) \setminus \{\mathbf{0}\}$,*

*(c) there is a neighbourhood $V \subset F$ of $\mathbf{0}$ in which $f$ is twice continuously differentiable,*

*(d) $f_{[1,0,0]} > 0$, $f_{[0,1,0]} > 0$, and $f_{[0,0,2]} > 0$, with $f_{[i,j,k]} = f_{[i,j,k]}(\mathbf{0})$ as in (2.1),*

*then it holds that*

$$\int\limits_{F \cap \left(\mathbb{R}_+^2 \times \mathbb{R}\right)} g(x)\, e^{-nf(x)}\, \mathrm{d}x = \frac{\sqrt{2\pi}}{n^{5/2}} \frac{g(x^*)}{f_{[1,0,0]}\, f_{[0,1,0]} \sqrt{f_{[0,0,2]}}}\, e^{-nf(x^*)}\, (1 + o(1)).$$

**Remark 5.2.29** This is the result from [19, Lemma 4] for $n = 3, k = 2$ and functions $g$ and $(-f)$ instead of $h$ and $f$. The parameter $\lambda$ from [19, Lemma 4] in our setting is replaced by the integer $n \in \mathbb{N}$. Furthermore, a typo within said result has been corrected, namely the sum in [19, Equation (11)] is replaced by a product (compare proof therein). This proposition is quite close to [85, v2, Lemma 5.1], but does not require the same level of smoothness of $f$ and $g$, and $g$ does not depend on $n \in \mathbb{N}$.

To apply this, we use a transformation of $\mathscr{D}_z \cap \mathscr{B}_z$, mapping $z^{**} = (z^q, 1, 1)$ to $\mathbf{0}$. Consider

$$\mathfrak{I} : \mathbb{R}^3 \to \mathbb{R}^3 \qquad \text{with} \qquad \mathfrak{I}(x_1, x_2, y) = \left(y^q x_1 - z^q x_2^{q/p}, 1 - y, x_2 - 1\right) = (t_1, t_2, t_3).$$

It then holds that

$$\mathfrak{I}(z^{**}) = \mathbf{0} \quad \text{and} \quad \mathfrak{I}(\mathscr{D}_z) = \tilde{\mathscr{D}}_z := \left\{t \in \mathbb{R}^3 : t_1 > 0, t_2 \in [0, 1), t_3 > -1\right\}.$$

Furthermore, in a neighbourhood of $z^{**}$ small enough such that $t_2 < 1$, $\mathfrak{J}$ is invertible with

$$\mathfrak{J}^{-1}(t_1,t_2,t_3) = \left(\frac{t_1 + z^q(t_3+1)^{q/p}}{(1-t_2)^q},\, t_3+1,\, 1-t_2\right).$$

Let us calculate the Jacobian of $\mathfrak{J}^{-1}$:

$$J_t\mathfrak{J}^{-1}(t) = \begin{pmatrix} \frac{1}{(1-t_2)^q} & \frac{q(t_1+z^q(t_3+1)^{q/p})}{(1-t_2)^{q+1}} & \frac{z^q \frac{q}{p}(t_3+1)^{q/p\,-\,1}}{(1-t_2)^q} \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}. \tag{5.59}$$

Thus, we have that $\left|\det J_t\mathfrak{J}^{-1}(t)\right| = (1-t_2)^{-q}$. We set $\mathscr{g}(x_1,x_2,y) := y^{-1}\big(\det \mathfrak{H}_x\big)^{-1/2}$, as well as $\tilde{\mathscr{B}}_z := \mathfrak{J}\big(\mathscr{B}_z\big)$, and transform the area of integration via $\mathfrak{J}^{-1}$, yielding

$$\begin{aligned}\mathbb{P}\left(\mathcal{S}^{(n)} \in \mathscr{D}_z \cap \mathscr{B}_z\right) &= \int\limits_{\mathscr{D}_z \cap \mathscr{B}_z} \mathscr{h}^{(n)}(x_1,x_2,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \frac{n^2}{2\pi} \int\limits_{\mathscr{D}_z \cap \mathscr{B}_z} y^{-1}(\det \mathfrak{H}_x)^{-1/2}\, e^{-n\,[\Lambda_p^*(x_1,x_2)-\log(y)]}\,\mathrm{d}x\,\mathrm{d}y\,(1+o(1)) \\ &= \frac{n^2}{2\pi} \int\limits_{\mathscr{D}_z \cap \mathscr{B}_z} \mathscr{g}(x_1,x_2,y)\, e^{-n\,\mathcal{I}_{\mathcal{S}}(x_1,x_2,y)}\,\mathrm{d}x\,\mathrm{d}y\,(1+o(1)) \\ &= \frac{n^2}{2\pi} \int\limits_{\tilde{\mathscr{D}}_z \cap \tilde{\mathscr{B}}_z} \mathscr{g} \circ \mathfrak{J}^{-1}(t)\, e^{-n\,\mathcal{I}_{\mathcal{S}} \circ \mathfrak{J}^{-1}(t)}\,(1-t_2)^{-q}\,\mathrm{d}t\,(1+o(1)).\end{aligned}$$

We now set $\tilde{g}(t) := (1-t_2)^{-q}\,\mathscr{g} \circ \mathfrak{J}^{-1}(t)$ and $\tilde{f}(t) := \mathcal{I}_{\mathcal{S}} \circ \mathfrak{J}^{-1}(t)$, then

$$\mathbb{P}\left(\mathcal{S}^{(n)} \in \mathscr{D}_z \cap \mathscr{B}_z\right) = \frac{n^2}{2\pi} \int\limits_{\tilde{\mathscr{D}}_z \cap \tilde{\mathscr{B}}_z} \tilde{g}(t)\, e^{-n\,\tilde{f}(t)}\,\mathrm{d}t\,(1+o(1)). \tag{5.60}$$

We intend to apply Proposition 5.2.28 to the integral in (5.60). It holds that $\tilde{\mathscr{D}}_z \cap \tilde{\mathscr{B}}_z$ is bounded and since the value of the integral is the same if we integrate over the open set $\tilde{\mathscr{B}}_z$ or its closure, we will continue to work with $\tilde{\mathscr{B}}_z$. Further, we have that $\tilde{\mathscr{B}}_z$ contains the origin in its (relative) interior, as the interior point $z^{**}$ of $\mathscr{B}_z$ is again mapped by the continuous function $\mathfrak{J}$ onto an interior point, which is $\mathfrak{J}(z^{**}) = \mathbf{0}$. Since we have chosen the neighbourhood $\mathscr{B}_z$ of $z^{**}$ small enough for $\tilde{\mathscr{B}}_z$ to not contain $(t_1, 1, t_3)$, it holds that

$$\tilde{g}(t) = (1-t_2)^{-q}\,\left[(1-t_2)^{-1}\left(\det \mathfrak{H}_{(\mathfrak{J}^{-1}(t)_1,\mathfrak{J}^{-1}(t)_3)}\right)^{-1/2}\right]$$

is also differentiable on $\tilde{\mathscr{D}}_z \cap \tilde{\mathscr{B}}_z$ as a composition of differentiable functions, and is thereby continuous on $\tilde{\mathscr{B}}_z$. The differentiability of $\mathfrak{J}^{-1}$, together with that of $\Lambda_p^*$ given by Lemma 2.2.1, yields the differentiability (and thereby the continuity) of $\tilde{f}(t) := \mathcal{I}_{\mathscr{S}} \circ \mathfrak{J}^{-1}(t)$ on $\tilde{\mathscr{B}}_z$. It holds furthermore that

$$\tilde{g}(\mathbf{0}) = (\det \mathfrak{H}_{z^*})^{-1/2}, \tag{5.61}$$

which is positive, since $\mathfrak{H}_{z^*}$ is positive definite on $\mathrm{Dom}(\Lambda_p^*)$, as argued in Section 5.2.8. Again, for $\mathscr{B}_z$ small enough, it also holds (up to a null set) that $\tilde{\mathscr{B}}_z \cap (\mathbb{R}_+^2 \times \mathbb{R}) = \tilde{\mathscr{B}}_z \cap \tilde{\mathscr{D}}_z$, on which we know from Lemma 5.2.18 that $\mathbf{0} = \mathfrak{J}(z^{**})$ is the unique infimum of $\tilde{f}$ since

$$\tilde{f}(\mathbf{0}) = \mathcal{I}_{\mathscr{S}} \circ \mathfrak{J}^{-1}(\mathbf{0}) = \mathcal{I}_{\mathscr{S}}(z^{**}) = \Lambda_p^*(z^*). \tag{5.62}$$

We can see from (5.59) that all partial derivatives of $\mathfrak{J}^{-1}$ are themselves continuously differentiable in a sufficiently small neighbourhood of $\mathbf{0}$. Thereby, $\mathfrak{J}^{-1}$ is twice continuously differentiable in such a neighbourhood. The two-fold continuous differentiability of $\Lambda_p^*$ on $\mathrm{Dom}(\Lambda_p^*)$ has already been shown in the proof of Theorem 5.2.11. Finally, by Lemma 5.2.1 i), it holds that

$$\begin{aligned}
\nabla_{(x_1,x_2,y)} \mathcal{I}_{\mathscr{S}}(z^{**}) &= \left( \frac{\partial}{\partial x_1} \Lambda_p^*(x_1,x_2), \frac{\partial}{\partial x_2} \Lambda_p^*(x_1,x_2), -\frac{1}{y} \right) \Big|_{(x_1,x_2,y)=z^{**}} \\
&= \left( \tau(x)_1, \tau(x)_2, -\frac{1}{y} \right) \Big|_{(x_1,x_2,y)=z^{**}} \\
&= \left( \tau(z^*)_1, \tau(z^*)_2, -1 \right),
\end{aligned}$$

from which we can deduce that

$$\begin{aligned}
\nabla_t \tilde{f}(\mathbf{0}) &= \nabla_{(x_1,x_2,y)} \mathcal{I}_{\mathscr{S}}(z^{**}) \, J_t \mathfrak{J}^{-1}(\mathbf{0}) \\
&= (\tau(z^*)_1, \tau(z^*)_2, -1) \begin{pmatrix} 1 & qz^q & z^q \frac{q}{p} \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \\
&= \left( \tau(z^*)_1, \; qz^q \tau(z^*)_1 + 1, \; z^q \frac{q}{p} \tau(z^*)_1 + \tau(z^*)_2 \right).
\end{aligned} \tag{5.63}$$

It thereby follows that $\nabla_t \tilde{f}(\mathbf{0}) \neq \mathbf{0}$, as the first two components cannot be equal to zero simultaneously. But since $\tilde{f}(t)$ attains its infimum on $\tilde{\mathscr{B}}_z \cap (\mathbb{R}_+^2 \times \mathbb{R})$ in $t = \mathbf{0}$, it holds that $\tilde{f}_{[1,0,0]} > 0$ and $\tilde{f}_{[0,1,0]} > 0$, as otherwise a step into either direction $t_1$ or $t_2$ would maintain or decrease the value of $\tilde{f}$, contradicting the unique infimum property.

On the other hand, by the same argument it has to hold that $\tilde f_{[0,0,1]} = 0$ and $\tilde f_{[0,0,2]} > 0$, as otherwise a step into either direction $t_3$ or $(-t_3)$ would maintain or decrease $\tilde f$, again contradicting the unique infimum property of $\mathbf{0}$. Hence, we have shown all conditions for Proposition 5.2.28, whereby it now follows for the integral in (5.60) that

$$\begin{aligned}
\mathbb{P}\left(\mathscr{S}^{(n)} \in \mathscr{D}_z \cap \mathscr{B}_z\right) &= \frac{n^2}{2\pi} \int\limits_{\tilde{\mathscr{D}}_z \cap \tilde{\mathscr{B}}_z} \tilde g(t)\, e^{-n\,\tilde f(t)}\,\mathrm{d}t\,(1+o(1)) \\
&= \frac{n^2}{2\pi} \int\limits_{\tilde{\mathscr{B}}_z \cap (\mathbb{R}_+^2 \times \mathbb{R})} \tilde g(t)\, e^{-n\,\tilde f(t)}\,\mathrm{d}t\,(1+o(1)) \\
&= \frac{1}{\sqrt{2\pi n}} \frac{\tilde g(\mathbf{0})}{\tilde f_{[1,0,0]}\,\tilde f_{[0,1,0]}\,\sqrt{\tilde f_{[0,0,2]}}}\, e^{-n\tilde f(\mathbf{0})}\,(1+o(1)). \qquad (5.64)
\end{aligned}$$

The final term that remains to be calculated explicitly is $\tilde f_{[0,0,2]}$, as $\tilde f_{[1,0,0]}$ and $\tilde f_{[0,1,0]}$ are given in (5.63). We start by noting that

$$\frac{\partial}{\partial t_3}\mathfrak{I}^{-1}(t)\Big|_{t=\mathbf{0}} = \left(\frac{z^q \frac{q}{p}(t_3+1)^{q/p-1}}{(1-t_2)^q}, 1, 0\right)\Bigg|_{t=\mathbf{0}} = \left(z^q\frac{p}{q}, 1, 0\right),$$

and

$$\frac{\partial^2}{\partial t_3^2}\mathfrak{I}^{-1}(t)\Big|_{t=\mathbf{0}} = \left(\frac{z^q \frac{q}{p}\left(\frac{q}{p}-1\right)(t_3+1)^{q/p-2}}{(1-t_2)^q}, 0, 0\right)\Bigg|_{t=\mathbf{0}} = \left(\frac{z^q q^2}{p^2} - \frac{z^q q}{p}, 0, 0\right).$$

By Lemma 5.2.1 ii), we get that

$$\mathcal{H}_{(x_1,x_2,y)}\mathcal{I}_{\mathscr{S}}(z^{**}) = \begin{pmatrix} \left(\mathfrak{H}_{z^*}^{-1}\right)_{11} & \left(\mathfrak{H}_{z^*}^{-1}\right)_{12} & 0 \\ \left(\mathfrak{H}_{z^*}^{-1}\right)_{21} & \left(\mathfrak{H}_{z^*}^{-1}\right)_{22} & 0 \\ 0 & 0 & y^{-2} \end{pmatrix}.$$

It thereby follows that

$$\begin{aligned}
\tilde f_{[0,0,2]} &= \frac{\partial^2}{\partial^2 t_3}\mathcal{I}_{\mathscr{S}} \circ \mathfrak{I}^{-1}(\mathbf{0}) \\
&= \frac{\partial}{\partial t_3}\left[\nabla_{(x_1,x_2,y)}\mathcal{I}_{\mathscr{S}}\left(\mathfrak{I}^{-1}(t)\right)\frac{\partial}{\partial t_3}\mathfrak{I}^{-1}(t)\right]\Bigg|_{t=\mathbf{0}} \\
&= \frac{\partial}{\partial t_3}\left[\nabla_{(x_1,x_2,y)}\mathcal{I}_{\mathscr{S}}\left(\mathfrak{I}^{-1}(t)\right)\right]\Big|_{t=\mathbf{0}}\frac{\partial}{\partial t_3}\mathfrak{I}^{-1}(\mathbf{0}) + \nabla_{(x_1,x_2,y)}\mathcal{I}_{\mathscr{S}}(z^{**})\,\frac{\partial^2}{\partial t_3^2}\mathfrak{I}^{-1}(\mathbf{0})
\end{aligned}$$

$$
\begin{aligned}
&= \left(z^q \frac{q}{p}, 1, 0\right) \mathcal{H}_{(x_1,x_2,y)} \mathcal{I}_{\mathcal{S}}(z^{**}) \ \left(z^q \frac{q}{p}, 1, 0\right) \\
&\qquad + \Big(\tau(z^*)_1, \tau(z^*)_2, -1\Big) \left(\frac{z^q q^2}{p^2} - \frac{z^q q}{p}, 0, 0\right) \\
&= \left(\frac{z^q q}{p}, 1, 0\right) \left(\frac{z^q q}{p}\left(\mathfrak{H}_{z^*}^{-1}\right)_{11} + \left(\mathfrak{H}_{z^*}^{-1}\right)_{12}, \frac{z^q q}{p}\left(\mathfrak{H}_{z^*}^{-1}\right)_{21} + \left(\mathfrak{H}_{z^*}^{-1}\right)_{22}, 0\right) \\
&\qquad + \tau(z^*)_1 \left(\frac{z^q q^2}{p^2} - \frac{z^q q}{p}\right) \\
&= \frac{z^{2q} q^2}{p^2}\left(\mathfrak{H}_{z^*}^{-1}\right)_{11} + \frac{2 z^q q}{p}\left(\mathfrak{H}_{z^*}^{-1}\right)_{12} + \left(\mathfrak{H}_{z^*}^{-1}\right)_{22} + \tau(z^*)_1 \left(\frac{z^q q^2}{p^2} - \frac{z^q q}{p}\right). \qquad (5.65)
\end{aligned}
$$

Plugging the terms from (5.61), (5.63) and (5.65) into the fraction in (5.64) yields

$$
\begin{aligned}
&\frac{\tilde{g}(\mathbf{0})}{\tilde{f}_{[1,0,0]} \tilde{f}_{[0,1,0]} \sqrt{|\tilde{f}_{[0,0,2]}|}} \\
&= \left(\det \mathfrak{H}_{z^*}\right)^{-1/2} \left(\tau(z^*)_1\right)^{-1} \left(q z^q \tau(z^*)_1 + 1\right)^{-1} \\
&\quad \times \left[\frac{z^{2q} q^2}{p^2}\left(\mathfrak{H}_{z^*}^{-1}\right)_{11} + \frac{2 z^q q}{p}\left(\mathfrak{H}_{z^*}^{-1}\right)_{12} + \left(\mathfrak{H}_{z^*}^{-1}\right)_{22} + \tau(z^*)_1 \left(\frac{z^q q^2}{p^2} - \frac{z^q q}{p}\right)\right]^{-1/2} \\
&= \Bigg[\det \mathfrak{H}_{z^*} \ \left(\tau(z^*)_1\right)^2 \left(q z^q \tau(z^*)_1 + 1\right)^2 \\
&\quad \times \left(\frac{z^{2q} q^2}{p^2}\left(\mathfrak{H}_{z^*}^{-1}\right)_{11} + \frac{2 z^q q}{p}\left(\mathfrak{H}_{z^*}^{-1}\right)_{12} + \left(\mathfrak{H}_{z^*}^{-1}\right)_{22} + \tau(z^*)_1 \frac{z^q q (q-p)}{p^2}\right)\Bigg]^{-1/2} \\
&= \gamma(z)^{-1}, \qquad (5.66)
\end{aligned}
$$

with $\gamma(z)$ as in (5.20). Hence, it follows with (5.62), (5.64), and (5.66) that

$$
\mathbb{P}\left(\mathcal{S}^{(n)} \in \mathscr{D}_z \cap \mathscr{B}_z\right) = \frac{1}{\sqrt{2\pi n}\, \gamma(z)} e^{-n \Lambda_p^*(z^*)} \left(1 + o(1)\right). \qquad (5.67)
$$

Combining (5.57) with the two integral estimates from (5.58) and (5.67) shows that the integral over the complement of $\mathscr{B}_z$ can be comparatively neglected and it follows

$$
\mathbb{P}\left(n^{1/p - 1/q} \left\|\mathscr{Z}^{(n)}\right\|_q > z\right) = \mathbb{P}\left(\mathcal{S}^{(n)} \in \mathscr{D}_z \cap \mathscr{B}_z\right) = \frac{1}{\sqrt{2\pi n}\, \gamma(z)} e^{-n \Lambda_p^*(z^*)} \left(1 + o(1)\right),
$$

which proves the second main result of this section for $\ell_p^n$-balls. $\square$

## 5.3 Sharpening the $p$-AGM inequality using SLDs

In this section we use sharp large deviation results to derive improvements for a probabilistic version of the $p$-generalized arithmetic-geometric mean ($p$-AGM) inequality, expanding on works of Kabluchko, Prochno, and Vysotsky [66] and Thäle [111]. We will begin by laying out the setting and the main results of this section and then proceed to prove them via the same three-pronged approach of the proof in the previous section, i.e., giving probabilistic representations for the target variables, deriving asymptotic density estimates for those representations, and then integrating over those densities using the geometric Laplace integration results of Andriani and Baldi [7] to finish the proof. Due to the similar nature of Section 5.2 and Section 5.3, we refer to proofs in the former whenever possible for the sake of brevity.

The well-established AGM inequality states that for $n \in \mathbb{N}$ and $x_1, \dots, x_n \in \mathbb{R}^n$

$$\left(\prod_{i=1}^{n} |x_i|\right)^{1/n} \leq \frac{1}{n}\sum_{i=1}^{n} |x_i|,$$

which can be generalized for some $p > 0$ to

$$\left(\prod_{i=1}^{n} |x_i|\right)^{1/n} \leq \left(\frac{1}{n}\sum_{i=1}^{n} |x_i|^p\right)^{1/p}, \tag{5.68}$$

referred to as the $p$-generalized AGM inequality. As layed out in Section 1.2, if the vector one applies the inequality to is chosen randomly from $\mathbb{B}_p^n$ with directional distribution $\mathbf{C}_{n,p}$, the $p$-AGM inequality can be improved or reversed up to a respective scalar constant with high probability. By a random vector $X^{(n)} \in \mathbb{B}_p^n$ having directional distribution $\mathbf{C}_{n,p}$ we mean that there is a distribution $\mathbf{R}$ on $[0, 1]$ with $\mathbf{R}(\{0\}) = 0$ such that for a random variable $R$ with distribution $\mathbf{R}$ and a random variable $Z^{(n)}$ with distribution $\mathbf{C}_{n,p}$ independent of $R$ we have that

$$X^{(n)} \stackrel{\mathcal{D}}{=} R \cdot Z^{(n)}. \tag{5.69}$$

This can also be expanded to sequences of $p$-radial distributions $\left(\mathbf{R}^{(n)}\right)_{n\in\mathbb{N}}$, if the limiting distribution also has no atom at zero. Furthermore, note that it can be conjectured that all such distributions are already characterized by the class of $p$-radial distributions $\mathbf{P}_{n,p,\mathbf{W}}$ from Proposition 2.4.4, however, it has not been proven at this point that this is in fact the case. The relevant quantity to consider when deriving such a result is the ratio of the two sides of the inequality in (5.68), which is analyzed with respect to is distributional properties. For a random vector $X^{(n)} \in \mathbb{B}_p^n$ with directional distribution $\mathbf{C}_{n,p}$ and $p$-radial distribution $\mathbf{R}$ without an atom at zero we now want to give sharp asymptotics for the probability that the ratio of the two sides of the $p$-AGM inequality is, respectively, bigger or smaller than a constant $\theta \in [0,1]$. Effectively this means that we provide asymptotically exact estimates on a non-logarithmic scale for the probability of the $p$-AGM inequality being improvable or reversible up to a constant, respectively.

To state this section's main result, we define the following functions: For $\tau \in \mathbb{R}^2$, set

$$\tilde{\Lambda}_p(\tau) := \log \int_{\mathbb{R}} e^{\tau_1 \log(|y|)+\tau_2 |y|^p} f_{\mathbf{N}_p}(y)\,\mathrm{d}y, \tag{5.70}$$

and denote by $\tilde{\Lambda}_p^*$ the Legendre-Fenchel transform of $\tilde{\Lambda}_p$, where we employ the notation "$\tilde{\Lambda}_p$" etc. to avoid confusion with $\Lambda_p$ from (5.8). However, the two functions are indeed very similar to each other, as only the first summand in the exponent of $\Lambda_p$ from (5.8) is exchanged from $|y|^q$ to $\log(|y|)$. Hence, we will often refer to arguments from Section 5.2 regarding $\Lambda_p$ when making the same point for $\tilde{\Lambda}_p$ if the argument remains the same. As we will see in Section 5.3.1, $\tilde{\Lambda}_p$ is the cumulant generating function of the probabilistic representation used in this section, so by Lemma 2.2.1 there is a $\tau(x) \in \mathrm{Dom}(\tilde{\Lambda}_p)$ for every $x \in \mathrm{Dom}(\tilde{\Lambda}_p^*)$, such that

$$\tilde{\Lambda}_p^*(x) = \langle x, \tau(x)\rangle - \tilde{\Lambda}_p(\tau(x)).$$

For $x \in \mathrm{Dom}(\tilde{\Lambda}_p^*)$ set

$$\tilde{\mathfrak{H}}_x := \mathcal{H}_\tau \tilde{\Lambda}_p(\tau(x)) \tag{5.71}$$

analogue to (5.10). By the same arguments as in Lemma 5.2.1 it holds that

$$\nabla_x \tilde{\Lambda}_p^*(x) = \tau(x), \qquad \text{and} \qquad \mathcal{H}_x \tilde{\Lambda}_p^*(x) = \tilde{\mathfrak{H}}_x^{-1}. \tag{5.72}$$

For $x > 0$ we define the digamma function $\psi$ and a constant $m_p$ based on $p$ and $\psi$ as

$$\psi(x) := \frac{\Gamma'(x)}{\Gamma(x)}, \qquad \text{and} \qquad m_p := \frac{1}{p}\left(\psi\left(\frac{1}{p}\right) + \log(p)\right) < 0.$$

As Section 5.3.1 will show, $e^{m_p}$ is the limit towards which the expectations of the ratio of the $p$-AGM inequality converge in $n \in \mathbb{N}$, thereby filling the same role as $m_{p,q}$ from (5.5) in Section 5.2. Furthermore, we define the prefactor functions $\tilde{\xi}(\theta)$ and $\tilde{\kappa}(\theta)$ for $\theta \in [0,1]$. For $\theta \in (0,1]$, we denote $\theta^* := (\log\theta, 1) \in \mathbb{R}^2$ and for $\theta \in (0,1)$ set

$$\tilde{\xi}(\theta)^2 := \langle \tilde{\mathfrak{H}}_{\theta^*}\,\tau(\theta^*), \tau(\theta^*)\rangle \,\det \tilde{\mathfrak{H}}_{\theta^*},$$

and

$$\tilde{\kappa}(\theta)^2 := 1 - c_{\tilde{\kappa}}(\theta),$$

with $c_{\tilde{\kappa}}(\theta)$ given by

$$\frac{\left(\tau(\theta^*)_1^2 + \tau(\theta^*)_2^2\right)^{3/2} p^2 e^{p\theta}\theta^{-p}}{\left|\tau(\theta^*)_2^2 \left(\tilde{\mathfrak{H}}_{\theta^*}^{-1}\right)_{11} - 2\tau(\theta^*)_1\tau(\theta^*)_2 \left(\tilde{\mathfrak{H}}_{\theta^*}^{-1}\right)_{12} + \tau(\theta^*)_1^2 \left(\tilde{\mathfrak{H}}_{\theta^*}^{-1}\right)_{22}\right| \left(1 + p^2 e^{2p\theta}\theta^{-2p}\right)^{3/2}}.$$

With the necessary definitions and notation set up, we now proceed to formulate the main result of Section 5.3.

**Theorem 5.3.1** *Let $1 \le p < \infty$, $n \in \mathbb{N}$, and $X^{(n)}$ be a random vector in $\mathbb{B}_p^n$ with directional distribution $\mathbf{C}_{n,p}$ in the sense of* (5.69). *It then holds*

*i) for $\theta \in (e^{m_p}, 1)$ and $n$ sufficiently large that*

$$\mathbb{P}\left[\left(\prod_{i=1}^n |X_i^{(n)}|\right)^{1/n} > \theta \cdot \left(\sum_{i=1}^n |X_i^{(n)}|\right)^{1/p}\right] = \frac{1}{\sqrt{2\pi n}\,\tilde{\kappa}(\theta)\tilde{\xi}(\theta)}\, e^{-n\,\mathcal{I}_p(\theta)}\,(1 + o(1)),$$

*ii) and for $\theta \in (0, e^{m_p})$ and $n$ sufficiently large that*

$$\mathbb{P}\left[\left(\prod_{i=1}^n |X_i^{(n)}|\right)^{1/n} < \theta \cdot \left(\sum_{i=1}^n |X_i^{(n)}|\right)^{1/p}\right] = \frac{1}{\sqrt{2\pi n}\,\tilde{\kappa}(\theta)\tilde{\xi}(\theta)}\, e^{-n\,\mathcal{I}_p(\theta)}\,(1 + o(1)),$$

*where*

$$\begin{aligned}\mathcal{I}_p(\theta) \;:=\;& \big[pG_p(\theta) - 1\big]\log(\theta) + G_p(\theta)\big[\log G_p(\theta) - 1\big] - \log\Gamma(G_p(\theta)) \\ &+ \frac{1}{p}\big(1 + \log(p)\big) + \log\Gamma\left(\frac{1}{p}\right),\end{aligned} \tag{5.73}$$

*with $G_p(\theta) := H^{-1}\big(p\log(\theta)\big)$, where $H : (0,\infty) \to (-\infty, 0)$ is an increasing bijection given by*

$$H(x) := \psi(x) - \log(x). \tag{5.74}$$

The two parts of the above theorem describe the decay of the probability that the $p$-AGM inequality is either reversible with a prefactor $\theta \in (e^{m_p}, 1)$ [part i)] or can be sharpened with a prefactor $\theta \in (0, e^{m_p})$ [part ii)]. Conversely, their respective opposites, i.e., the probabilities that the inequality can be reversed with a prefactor $\theta \in (0, e^{m_p})$ or sharpened with a prefactor $\theta \in (e^{m_p}, 1)$ tend to 1 in $n \in \mathbb{N}$. This will be pointed out in further detail in Section 5.3.1.

Note that this result is not dependent on the $p$-radial distribution of $X^{(n)}$, as is also the case in [66, 111], even though SLD results usually tend to be more sensitive to the idiosyncrasies of the underlying distributions.

These results are consistent with the LDP of Kabluchko, Prochno, and Vysotsky, as taking the logarithm of the probability in the above theorem, dividing by $n$, and then considering the limit, yields what they have shown in [66, Theorem 1.2], namely that

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left[\left(\prod_{i=1}^{n}\left|X_i^{(n)}\right|\right)^{1/n}>\theta\cdot\left(\sum_{i=1}^{n}\left|X_i^{(n)}\right|\right)^{1/p}\right]=-\mathcal{I}_p(\theta) \tag{5.75}$$

for $\theta\in(e^{m_p},1)$ and

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left[\left(\prod_{i=1}^{n}\left|X_i^{(n)}\right|\right)^{1/n}>\theta\cdot\left(\sum_{i=1}^{n}\left|X_i^{(n)}\right|\right)^{1/p}\right]=-\mathcal{I}_p(\theta) \tag{5.76}$$

for $\theta\in(0,e^{m_p})$. However, we do provide a refinement of their findings in the classic sense of sharp large deviation results refining LDPs, as layed out in Section 5.1: Theorem 5.3.1 gives estimates on a non-logarithmic scale and we can thereby give concrete and asymptotically exact probability estimates for the reversibility and improvability of the $p$-AGM inequality for a specific (sufficiently large) $n\in\mathbb{N}$, whereas on the logarithmic scale of an LDP, as in (5.75) and (5.76), this is not possible and the prefactor in Theorem 5.3.1 vanishes.

### 5.3.1 Probabilistic representation for $p$-AGM ratios

We now turn to the first of the three steps in proving Theorem 5.3.1, which is providing a probabilistic representation for the ratio of the two sides of the $p$-AGM inequality in terms of $p$-generalized Gaussian random vectors. Furthermore, the large deviation results of Kabluchko, Prochno, and Vysotsky [66] for this ratio will be given explicitly and expanded to general distributions with directional component $\mathbf{C}_{n,p}$.

For a random vector $X^{(n)}\in\mathbb{B}_p^n$ with directional distribution $\mathbf{C}_{n,p}$ in the sense of (5.69) the main variable of interest is the ratio of the two sides of the $p$-AGM inequality given as

$$\mathcal{R}_n:=\frac{\left(\prod\limits_{i=1}^{n}\left|X_i^{(n)}\right|\right)^{1/n}}{\left(\frac{1}{n}\sum\limits_{i=1}^{n}\left|X_i^{(n)}\right|^p\right)^{1/p}}. \tag{5.77}$$

We want to formulate the target probabilities $\mathbb{P}(\mathcal{R}_n > \theta)$ and $\mathbb{P}(\mathcal{R}_n < \theta)$ via a random vector $Y^{(n)}$ with $p$-generalized Gaussian distribution $\mathbf{N}_p$ of its coordinates. It directly follows from Proposition 2.4.2 for a random vector $X^{(n)} \in \mathbb{B}_p^n$ with directional distribution $\mathbf{C}_{n,p}$ and $p$-radial distribution $\mathbf{R}$ on $[0,1]$ in the sense of (5.69) that

$$\mathcal{R}_n \stackrel{\mathcal{D}}{=} \frac{\left(\prod_{i=1}^{n}\left|R\,\frac{Y_i^{(n)}}{\|Y_i^{(n)}\|_p}\right|\right)^{1/n}}{\left(\frac{1}{n}\sum_{i=1}^{n}\left|R\,\frac{Y_i^{(n)}}{\|Y_i^{(n)}\|_p}\right|^p\right)^{1/p}} = \frac{\left(\prod_{i=1}^{n}|Y_i^{(n)}|\right)^{1/n}}{\left(\frac{1}{n}\sum_{i=1}^{n}|Y_i^{(n)}|^p\right)^{1/p}}, \tag{5.78}$$

with i.i.d. $Y_i^{(n)} \sim \mathbf{N}_p$. Thus, we see that $\mathcal{R}_n$ does not depend on the $p$-radial distribution $\mathbf{R}$, which is why the rate function in the main result is universal for all random vectors in $\mathbb{B}_p^n$ with directional distribution $\mathbf{C}_{n,p}$. This calculation, previously given in [111], also shows that the central limit theorem and the LDP established in [66] and the MDP shown in [111] also hold for any random vector in $\mathbb{B}_p^n$ with directional distribution $\mathbf{C}_{n,p}$ as in (5.69). In the light of the representation in (5.78), let us present the LDP based on [66, Theorem 1.4] here in this more general form.

**Proposition 5.3.2** *Let $1 \le p < \infty$ and $X^{(n)}$ be a random vector in $\mathbb{B}_p^n$ with directional distribution $\mathbf{C}_{n,p}$ in the sense of* (5.69)*. Then the sequence $(\mathcal{R}_n)_{n\in\mathbb{N}}$ with $\mathcal{R}_n$ as defined in* (5.77) *based on $X^{(n)}$ satisfies an LDP on $[0,1]$ with speed $n$ and rate function $\mathcal{I}_p$ as in* (5.73)*.*

It is furthermore shown in [66] that $\mathcal{I}_p(e^{m_p}) = 0$ and $\mathcal{I}_p(0+) = \mathcal{I}_p(1-) = +\infty$, where $\mathcal{I}_p(0+)$ and $\mathcal{I}_p(1-)$ denote the limits of $\mathcal{I}_p$ for sequences that converge to zero and one from above and below, respectively. As suggested by the central limit theorem in [66, Theorem 1.1], the expectations of $\mathcal{R}_n$ converge to $e^{m_p}$, that is, $e^{m_p}$ is the value from which deviation probabilities are given in the above LDP and the sharp large deviation results in Theorem 5.3.1.

Proposition 5.3.2 is proven in [66] by showing an LDP for the sequence of empirical averages of the coordinates of the random vector

$$\tilde{V}^{(n)} := \left(\tilde{V}_1^{(n)}, \ldots, \tilde{V}_n^{(n)}\right), \qquad \text{with} \qquad \tilde{V}_i^{(n)} := \left(\log|Y_i^{(n)}|, |Y_i^{(n)}|^p\right), \tag{5.79}$$

with $Y_i^{(n)}$ i.i.d. and $Y_i^{(n)} \sim \mathbf{N}_p$ (we again employ the notation "$\tilde{V}^{(n)}$" to avoid confusion with the terms from (5.6)). This is done via Cramér's theorem in Proposition 2.3.3, i.e., by showing that the cumulant generating function of the $\tilde{V}_i^{(n)}$, which is given by $\tilde{\Lambda}_p$

from (5.70), is finite in a neighbourhood of the origin, hence the sequence of empirical averages of the coordinates

$$\tilde{S}^{(n)} := \frac{1}{n}\sum_{i=1}^{n}\tilde{V}_i^{(n)} = \frac{1}{n}\sum_{i=1}^{n}\Big(\log\big|Y_i^{(n)}\big|, \big|Y_i^{(n)}\big|^p\Big) \tag{5.80}$$

satisfies an LDP with speed $n$ and rate function given by $\tilde{\Lambda}_p^*$. This LDP is then mapped to the sequence $(\mathcal{R}_n)_{n\in\mathbb{N}}$ via the probabilistic representation from (5.78) and the contraction principle in Proposition 2.3.5, considering the map $F(x_1,x_2) := e^{x_1}{x_2}^{-1/p}$, yielding an LDP for $(\mathcal{R}_n)_{n\in\mathbb{N}}$ with speed $n$ and rate function

$$\inf_{(x_1,x_2):\,F(x_1,x_2)=\theta} \tilde{\Lambda}_p^*(x_1,x_2), \qquad \theta\in[0,1].$$

This is then finalized by showing that the above infimum is attained uniquely at $\theta^* := (\log\theta, 1)$ and that this infimum can be given explicitly as

$$\inf_{(x_1,x_2):\,F(x_1,x_2)=\theta} \tilde{\Lambda}_p^*(x_1,x_2) = \tilde{\Lambda}_p^*(\theta^*) = \mathcal{I}_p(\theta), \tag{5.81}$$

with $\mathcal{I}_p$ as in (5.73). Further, it is shown that the effective domain of $\mathcal{I}_p$ is $(0,1)$ and that for $x\in \mathrm{Dom}(\tilde{\Lambda}_p^*)$ it holds that

$$\tau(x) = \Big(p\,H^{-1}\big(px_1 - \log x_2\big) - 1,\ p^{-1} - x_2^{-1}\,H^{-1}\big(px_1 - \log x_2\big)\Big) \tag{5.82}$$

with $H$ as in (5.74) (see [66, p. 11 f.]).

We use the same probabilistic representations from (5.79) and (5.80), but proceed with them in a different fashion. Hence, as in Section 5.2, we rewrite the target probability as the probability of $\tilde{S}^{(n)}$ lying in some domain dependent on $\theta$, which we again refer to as the deviation area. It holds that for $\theta\in[0,1]$

$$\mathbb{P}\big(\mathcal{R}_n > \theta\big) = \mathbb{P}\Big(\tilde{S}^{(n)} \in D_{\theta,>}\Big) \quad\text{and}\quad \mathbb{P}\big(\mathcal{R}_n < \theta\big) = \mathbb{P}\Big(\tilde{S}^{(n)} \in D_{\theta,<}\Big), \tag{5.83}$$

with

$$D_{\theta,>} := \Big\{x\in\mathbb{R}^2 : x_2>0, e^{x_1}{x_2}^{-1/p} > \theta\Big\}, \tag{5.84}$$

and

$$D_{\theta,<} := \Big\{x\in\mathbb{R}^2 : x_2>0, e^{x_1}{x_2}^{-1/p} < \theta\Big\}. \tag{5.85}$$

**Remark 5.3.3** Again the points satisfying the infimum condition $F(x_1,x_2)=\theta$ in (5.81) are exactly those on the boundary $\{(x_1,x_2)\in\mathbb{R}^2 : x_2>0, e^{x_1}{x_2}^{-1/p}=\theta\}$ of $D_{\theta,>}$ and $D_{\theta,<}$ (which coincide). Hence, (5.81) shows that the infimum of $\tilde{\Lambda}_p^*$ over this boundary is uniquely attained at $\theta^*$.

### 5.3.2 Asymptotic density estimate for $p$-AGM ratios

We now give an asymptotic estimate for the density of $\tilde{S}^{(n)}$, denoted by $\tilde{h}^{(n)}$, such that for sufficiently large $n \in \mathbb{N}$ we can write the probabilities in (5.83) as integrals of $\tilde{h}^{(n)}$ over $D_{\theta,>}$ and $D_{\theta,<}$.

**Proposition 5.3.4** *Let $p \in [1,\infty)$ and $n \in \mathbb{N}$. For $\tilde{S}^{(n)} = \frac{1}{n}\sum_{i=1}^n \tilde{V}_i^{(n)}$ with $\tilde{V}_i^{(n)} = \big(\log\big|Y_i^{(n)}\big|, \big|Y_i^{(n)}\big|^p\big)$, $Y_i^{(n)} \sim \mathbf{N}_p$ i.i.d., $x \in \operatorname{Dom}(\tilde{\Lambda}_p^*)$, and $n$ sufficiently large, it holds that the distribution of $\tilde{S}^{(n)}$ has Lebesgue density*

$$\tilde{h}^{(n)}(x) = \frac{n}{2\pi}\left(\det\tilde{\mathfrak{H}}_x\right)^{-1/2} e^{-n\,\tilde{\Lambda}_p^*(x)}\,(1+o(1)),$$

*with $\tilde{\mathfrak{H}}_x$ as in* (5.71).

Deriving the density estimate in Section 5.2.7 was quite involved, so to prove the above we only adapt the parts of the proof therein in which the specific nature of the probabilistic representation, that is, the change from $\big|Y_i^{(n)}\big|^q$ to $\log\big|Y_i^{(n)}\big|$, is relevant. As pointed out in Section 5.2.7, the proof of Proposition 5.2.19 follows along the lines of Borovkov and Rogozin [17]. However, since the conditions therein, namely the random variables having a common bounded density, are not satisfied in either of our settings, one mainly needs to show the integrability of the Fourier transform (see [7, Remark 3.2]). The main part of the proof of Proposition 5.2.19 for which the structure of the first component of the $V_i^{(n)} = \big(\big|Y_i^{(n)}\big|^q, \big|Y_i^{(n)}\big|^p\big)$ plays a role is in Lemma 5.2.20. Therein it was proven that there is an $s > 1$ such that the Fourier transform of the sum of two $V_i^{(n)}$ (times some exponential term) is in $L_s(\mathbb{R}^2)$. This can be proven for $\tilde{V}_i^{(n)} = \big(\log\big|Y_i^{(n)}\big|, \big|Y_i^{(n)}\big|^p\big)$ in a mostly analogue fashion by rewriting their sum as a transformation of the $Y_i^{(n)}$ and making a transformation of densities argument, and we will briefly argue why that is the case. For $i = 1, j = 2$ we can write

$$\tilde{V}_1^{(n)} + \tilde{V}_2^{(n)} = \Big(\log\big|Y_1^{(n)}\big| + \log\big|Y_2^{(n)}\big|,\ \big|Y_1^{(n)}\big|^p + \big|Y_2^{(n)}\big|^p\Big) =: \tilde{T}\Big(Y_1^{(n)}, Y_2^{(n)}\Big).$$

For some given $x \in \mathbb{R}^2$ we solve $x = \tilde{T}(y)$ for $y \in \mathbb{R}^2$, which amounts to characterizing

$$\tilde{T}^{-1}(x) = \big\{y \in \mathbb{R}^2 : \log(|y_1|) + \log(|y_2|) = x_1,\ |y_1|^p + |y_2|^p = x_2\big\}.$$

If $x_2 \leq 0$, the above set is clearly empty, hence we assume $x_2 > 0$, in which case the set $\big\{y \in \mathbb{R}^2 : |y_1|^p + |y_2|^p = x_2\big\}$ is an $\ell_p^n$-sphere with radius $x_2^{1/p}$, again denoted as $\mathbb{S}_p^{n-1}\big(x_2^{1/p}\big)$. For $x_1 \in \mathbb{R}$ the set $G(x_1) := \big\{y \in \mathbb{R}^2 : \log(|y_1|) + \log(|y_2|)) = x_1\big\}$ can be interpreted as the graph of the function $f(y) = e^{x_1}y^{-1}$ mirrored along the axes and at the origin into all orthans of $\mathbb{R}^2$ (see Figure 5.3).

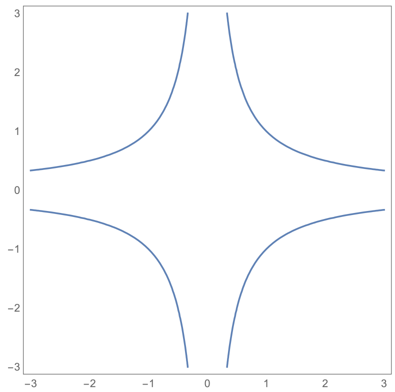


FIGURE 5.3: The set $G(r)$ for $r = 0$.

Therefore, we have

$$\tilde{T}^{-1}(x) = G(x_1) \cap \mathbb{S}_p^1\big(x_2^{1/p}\big).$$

As one can see in Figure 5.4, the number of points in the above set is zero in case $\exp\big(\frac{x_1}{2}\big) > \big(\frac{x_2}{2}\big)^{1/p}$, four if $\exp\big(\frac{x_1}{2}\big) = \big(\frac{x_2}{2}\big)^{1/p}$, and eight if $\exp\big(\frac{x_1}{2}\big) < \big(\frac{x_2}{2}\big)^{1/p}$. Since the case $\exp(\frac{x_1}{2}) = \big(\frac{x_2}{2}\big)^{1/p}$ only holds for a zero set of $x \in \mathbb{R} \times (0, \infty)$, we only consider $x \in \mathbb{R} \times (0, \infty)$ with $\exp(\frac{x_1}{2}) < \big(\frac{x_2}{2}\big)^{1/p}$.

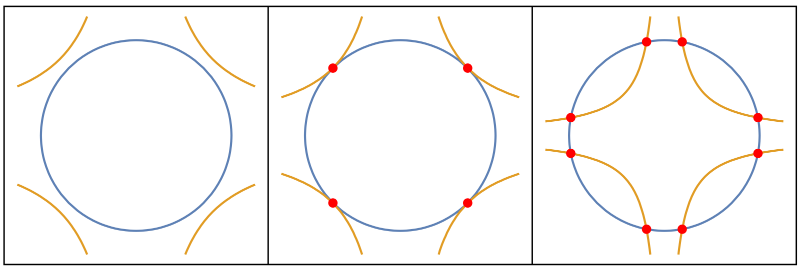

FIGURE 5.4: Intersection points (red) of $G(r)$ (orange) and $\mathbb{S}_2^1(1)$ (blue) for different values of $r$ with $r \in \big\{0.25,\ 0,\ -1\big\}$ (from left to right).

The transformation $\tilde{T}$ is continuously differentiable outside of $(0, 0)$ with Jacobian

$$J_y\tilde{T}(y) = \begin{pmatrix} \operatorname{sgn}(y_1)|y_1|^{-1} & \operatorname{sgn}(y_2)|y_2|^{-1} \\ \operatorname{sgn}(y_1)p|y_1|^{p-1} & \operatorname{sgn}(y_2)p|y_2|^{p-1} \end{pmatrix},$$

and

$$\det\big[J_y\tilde{T}(y)\big] = \operatorname{sgn}(y_1)\operatorname{sgn}(y_2)\,|y_1|^{-1}|y_2|^{-1}\,p\big(|y_2|^p - |y_1|^p\big),$$

which is only zero for $|y_1| = |y_2|$, thus can be disregarded, since it geometrically amounts to $G(x_1)$ and $\mathbb{S}_p^1(x_2^{1/p})$ intersecting at the points $2^{-1/p}x_2^{1/p}(\pm1, \pm1)$, which we excluded. For the remainder of the proof, one can proceed just as was done in the proof of Lemma 5.2.20, that is, apply the density transformation argument from [54, Section 4.5, p. 151 ff.] via $\tilde{T}$ and use the same integral estimates therein.

Thus, we see that one can show an analogue version of Lemma 5.2.20 in this setting, and hence, we can proceed in an analogue fashion as the remainder of the proof of Proposition 5.2.19 to show the density estimate in Proposition 5.3.4. The details of this are therefore omitted.

### 5.3.3 Proof of the sharpening of the $p$-AGM inequality

Assuming the set-up of Theorem 5.3.1 and combining the probabilistic representation results in (5.78) and (5.83) with the local density approximation in Proposition 5.3.4, we get that

$$\begin{aligned}
\mathbb{P}(\mathcal{R}_n > \theta) &= \mathbb{P}\Big(\tilde{S}^{(n)} \in D_{\theta,>}\Big) \\
&= \int\limits_{D_{\theta,>}} \tilde{h}^{(n)}(x)\,\mathrm{d}x \\
&= \frac{n}{2\pi} \int\limits_{D_{\theta,>}} \big(\det \tilde{\mathfrak{H}}_x\big)^{-1/2} e^{-n\,\tilde{\Lambda}_p^*(x)}\,\mathrm{d}x\,(1+o(1)), \qquad (5.86)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}(\mathcal{R}_n < \theta) &= \mathbb{P}\Big(\tilde{S}^{(n)} \in D_{\theta,<}\Big) \\
&= \int\limits_{D_{\theta,>}} \tilde{h}^{(n)}(x)\,\mathrm{d}x \\
&= \frac{n}{2\pi} \int\limits_{D_{\theta,<}} \big(\det \tilde{\mathfrak{H}}_x\big)^{-1/2} e^{-n\,\tilde{\Lambda}_p^*(x)}\,\mathrm{d}x\,(1+o(1)), \qquad (5.87)
\end{aligned}$$

with $D_{\theta,>}$ and $D_{\theta,<}$ as in (5.84) and (5.85). The final step of the proof of Theorem 5.3.1 now is to calculate the above integrals explicitly. We will only do this in detail for the integral in (5.86), as the calculation for the integral in (5.87) proceeds in a mostly analogue fashion, and we will merely point out the specific differences at the end of the proof. As in [7, 69, 85], the first step is to split up the integration area into a neighbourhood around the point $\theta^*$, at which the exponent in the integrand attains its infimum on the boundary of $D_{\theta,>}$, and its complement. On this neighbourhood we then employ the Laplace integral approximation from Proposition 5.2.27 by Andriani and Baldi [7], and on the complement we use the large deviation principle from Proposition 5.3.2 to show the comparative negligibility of the corresponding integral.

*Proof of Theorem 5.3.1.* We begin by proving the statement in Theorem 5.3.1 i). Let us assume the setting therein and let $B_\theta$ be an open neighbourhood of $\theta^*$, small enough such that $B_\theta \subset \mathrm{Dom}(\tilde{\Lambda}_p^*)$. The fact that $\theta^* \in \mathrm{Dom}(\tilde{\Lambda}_p^*)$ follows from the fact that $\tilde{\Lambda}_p^*(\theta^*) = \mathcal{I}_p(\theta) < \infty$ for $\theta \in (0,1)$, as seen in Proposition 5.3.2. Splitting up the reformulation of our target probability in (5.86) into integrals of $\tilde{h}^{(n)}$ over $B_\theta$ and $B_\theta^c$ yields

$$\mathbb{P}(\mathcal{R}_n > \theta) = \int\limits_{D_{\theta,>} \cap B_\theta} \tilde{h}^{(n)}(x)\,\mathrm{d}x + \int\limits_{D_{\theta,>} \cap B_\theta^c} \tilde{h}^{(n)}(x)\,\mathrm{d}x. \tag{5.88}$$

We begin by showing the comparative negligibility of the second integral term. We know from Remark 5.3.3 that $\tilde{\Lambda}_p^*$ attains its unique infimum on $\partial D_{\theta,>}$ at $\theta^*$. This property can be shown to hold for the closure $\overline{D}_{\theta,>}$ as follows: assume $t \in \mathbb{R}^2$ with $t \in D_{\theta,>}^\circ$, i.e., $e^{t_1} t_2^{-1/p} > \theta$. We then consider $\vartheta := e^{t_1} t_2^{-1/p}$. If $\vartheta^* \notin \mathrm{Dom}(\tilde{\Lambda}_p^*)$, it trivially holds that $\tilde{\Lambda}_p^*(\theta^*) < \tilde{\Lambda}_p^*(t) = \infty$. Hence, assume that $\vartheta^* \in \mathrm{Dom}(\tilde{\Lambda}_p^*)$. It now follows that $t \in \partial D_{\vartheta,>}$, which yields that $\tilde{\Lambda}_p^*(t) > \tilde{\Lambda}_p^*(\vartheta^*) = \mathcal{I}_p(\vartheta)$ by Remark 5.3.3. By Lemma 2.2.1 $\tilde{\Lambda}_p^*$ is strictly convex on $\mathrm{Dom}(\tilde{\Lambda}_p^*)$. From Proposition 5.3.2 we have that $\tilde{\Lambda}_p^*(e^{m_p *}) = \mathcal{I}_p(e^{m_p}) = 0$, thus $\mathcal{I}_p$ is strictly increasing on the interval $(e^{m_p}, 1)$, hence for $\vartheta > \theta$ we have $\tilde{\Lambda}_p^*(t) > \mathcal{I}_p(\vartheta) > \mathcal{I}_p(\theta) = \tilde{\Lambda}_p^*(\theta^*)$, thereby proving that $\tilde{\Lambda}_p^*$ attains its unique infimum on $\overline{D}_{\theta,>}$ at $\theta^*$.

Therefore, it follows from $\theta^* \notin B_\theta^c$ that there is an $\eta > 0$, such that

$$\inf_{t \in D_{\theta,>} \cap B_\theta^c} \tilde{\Lambda}_p^*(t) > \tilde{\Lambda}_p^*(\theta^*) + \eta.$$

The LDP in Proposition 5.3.2 then implies that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\tilde{S}^{(n)} \in D_{\theta,>} \cap B_\theta^c\Big) \le -\inf_{y \in D_\theta \cap B_\theta^c} \tilde{\Lambda}_p^*(y) \le -\tilde{\Lambda}_p^*(\theta^*) - \eta,$$

from which it follows that

$$\mathbb{P}\Big(\tilde{S}^{(n)} \in D_{\theta,>} \cap B_\theta^c\Big) \le e^{-n\,\tilde{\Lambda}_p^*(\theta^*) - n\,\eta}\,(1 + o(1)) = \frac{1}{e^{n\,\eta}}\, e^{-n\tilde{\Lambda}_p^*(\theta^*)}(1 + o(1)). \tag{5.89}$$

Due to the leading exponential term $e^{-n\,\eta}$, the above will be comparatively negligible compared to the other integral term

$$\int\limits_{D_{\theta,>} \cap B_\theta} \tilde{h}^{(n)}(x)\,\mathrm{d}x = \frac{n}{2\pi} \int\limits_{D_{\theta,>} \cap B_\theta} \big(\det \tilde{\mathfrak{H}}_x\big)^{-1/2} e^{-n\,\tilde{\Lambda}_p^*(x)}\,\mathrm{d}x\,(1 + o(1)), \tag{5.90}$$

which we will concretely calculate in the following. The clear course of action for this will be to apply Proposition 5.2.27 to the integral in (5.90) with $D = D_{\theta,>} \cap B_\theta \subset \mathbb{R}^2$,

$x^* = \theta^*$, $g(x) := \big(\det \tilde{\mathfrak{H}}_x\big)^{-1/2}$ and $\phi(x) = \tilde{\Lambda}_p^*(x)$. The area of integration is clearly bounded and since for sufficiently small $B_\theta$ it follows from (5.84) that $\partial(D_{\theta,>} \cap B_\theta)$ around $\theta^*$ is a section of the graph of the differentiable function $f(t_1) = \theta^{-p} e^{pt_1}$, it is indeed a differentiable planar curve. Also, $\tilde{\Lambda}_p^*$ attains a unique infimum on $\partial(D_{\theta,>} \cap B_\theta)$ in $\theta^*$ (see (5.81) and Remark 5.3.3), which also holds for the entirety of $\overline{D_{\theta,>} \cap B_\theta}$, as was shown above. The other conditions of Proposition 5.2.27 follow by the same arguments put forth in the proof of Theorem 5.2.11 in Section 5.2.8. Hence, we can apply Proposition 5.2.27 as intended, which gives

$$\int\limits_{D_{\theta,>} \cap B_\theta} \tilde{h}^{(n)}(x)\,\mathrm{d}x$$
$$= \frac{1}{\sqrt{2\pi n}} \frac{\det\big(L_{\tilde{\Lambda}}^{-1}\big(L_{\tilde{\Lambda}} - L_D\big)\big)^{-1/2} e^{-n\,\tilde{\Lambda}_p^*(\theta^*)}}{\big\langle \mathcal{H}_x\, \tilde{\Lambda}_p^*(\theta^*)^{-1}\, \nabla_x\, \tilde{\Lambda}_p^*(\theta^*), \nabla_x\, \tilde{\Lambda}_p^*(\theta^*)\big\rangle^{1/2} \big(\det \tilde{\mathfrak{H}}_{\theta^*}\big)^{1/2}} (1 + o(1)), \tag{5.91}$$

where $L_{\tilde{\Lambda}}$ and $L_D$ are the respective Weingarten maps of the curves

$$\mathscr{C}_D = \partial(D_{\theta,>} \cap B_\theta) \qquad \text{and} \qquad \mathscr{C}_{\tilde{\Lambda}} = \{x \in \mathbb{R}^2 : \tilde{\Lambda}_p^*(x) = \tilde{\Lambda}_p^*(\theta^*)\}$$

at $\theta^*$. We now need to resolve the different components in this fraction. As stated in (5.72), by the same arguments as in Lemma 5.2.1 it holds that $\nabla_x \tilde{\Lambda}_p^*(x) = \tau(x)$, and $\mathcal{H}_x \tilde{\Lambda}_p^*(x) = \tilde{\mathfrak{H}}_x^{-1}$. This allows rewriting the term in the denominator in (5.91) as

$$\Big\langle \mathcal{H}_x\, \tilde{\Lambda}_p^*(\theta^*)^{-1}\, \nabla_x\, \tilde{\Lambda}_p^*(\theta^*), \nabla_x\, \tilde{\Lambda}_p^*(\theta^*)\Big\rangle \det \tilde{\mathfrak{H}}_{\theta^*}$$
$$= \Big\langle \tilde{\mathfrak{H}}_{\theta^*}\, \tau(\theta^*), \tau(\theta^*)\Big\rangle \det \tilde{\mathfrak{H}}_{\theta^*} = \tilde{\xi}(\theta)^2. \tag{5.92}$$

It remains to give the Weingarten maps of the curves $\mathscr{C}_D$ and $\mathscr{C}_{\tilde{\Lambda}}$ explicitly. We start by noting that $\det(L_{\tilde{\Lambda}}^{-1}(L_{\tilde{\Lambda}} - L_D)) = 1 - L_D {L_{\tilde{\Lambda}}}^{-1} = \tilde{\kappa}(\theta)^2$, the determinant falling away due to the Weingarten maps being one-dimensional. As discussed in Section 5.2.2, the Weingarten map of a planar curve at a given point reduces to the absolute value of its curvature at that point. Since $\mathscr{C}_D = \partial(D_{\theta,>} \cap B_\theta)$ around $\theta^*$ is a segment of the graph of $f(t_1) = \theta^{-p} e^{pt_1}$, we get from Corollary 5.2.10 ii) that

$$L_D = \frac{|f''(\theta)|}{(1 + f'(\theta)^2)^{3/2}} = \frac{p^2 e^{p\theta} \theta^{-p}}{\big(1 + p^2 e^{2p\theta} \theta^{-2p}\big)^{3/2}}. \tag{5.93}$$

The curve $\mathscr{C}_{\tilde{\Lambda}}$ can be written as the zero set of the function $F(x) := \tilde{\Lambda}_p^*(x) - \tilde{\Lambda}_p^*(\theta^*)$, and its derivatives $F_{[i,j]}$ at $\theta^*$ as in Corollary 5.2.10 i) are known from the identities $\nabla_x \tilde{\Lambda}_p^*(x) = \tau(x)$ and $\mathcal{H}_x \tilde{\Lambda}_p^*(x) = \tilde{\mathfrak{H}}_x^{-1}$. (Note, that for $\theta = e^{m_p}$ we have that $\mathscr{C}_{\tilde{\Lambda}}$ is the zero set of $F(x) = \tilde{\Lambda}_p^*(x)$, since $\tilde{\Lambda}_p^*(e^{m_p *}) = 0$. By (5.82) it follows that $\tau(x) = 0$ only if $x = e^{m_p *} = (m_p, 1)$. Hence, the zero set of $F(x) = \tilde{\Lambda}_p^*(x)$ is solely the point $e^{m_p *}$, which is not a differentiable curve, and hence is not accessible by these geometric methods). It thus follows that

$$L_{\tilde{\Lambda}} = \frac{\left|\tau(\theta^*)_2^2 \left(\tilde{\mathfrak{H}}_{\theta^*}^{-1}\right)_{11} - 2\tau(\theta^*)_1 \tau(\theta^*)_2 \left(\tilde{\mathfrak{H}}_{\theta^*}^{-1}\right)_{12} + \tau(\theta^*)_1^2 \left(\tilde{\mathfrak{H}}_{\theta^*}^{-1}\right)_{22}\right|}{\left(\tau(\theta^*)_1^2 + \tau(\theta^*)_2^2\right)^{3/2}}.$$

This, together with (5.93), now yields that $1 - L_D {L_{\tilde{\Lambda}}}^{-1} = \tilde{\kappa}(\theta)^2$, which combined with (5.81) and (5.92) and applied to (5.91) gives

$$\int\limits_{D_{\theta,>} \cap B_\theta} \tilde{h}^{(n)}(x)\,\mathrm{d}x = \frac{1}{\sqrt{2\pi n}\,\tilde{\xi}(\theta)\,\tilde{\kappa}(\theta)} e^{-n\mathcal{I}_p(\theta)}\,(1 + o(1)). \tag{5.94}$$

Comparing (5.94) with the upper bound of the integral outside of $B_\theta$ in (5.89), we can see that the integral over $B_\theta^c$ is negligible, as it is of order $o(1)$. Thus, combining (5.88), (5.89) and (5.94) finishes the proof of Theorem 5.3.1 i).

The proof of Theorem 5.3.1 ii) is almost completely the same regarding probabilistic representation, local density estimation, and integral approximation, as hardly any of the steps therein use the fact that we are working on $D_{\theta,>}$ for $\theta \in (e^{m_p}, 1)$ instead of $D_{\theta,<}$ for $\theta \in (0, e^{m_p})$, but rather consider a neighbourhood of $\partial D_{\theta,>}$ around $\theta^*$, which coincides with that same neighbourhood of $\partial D_{\theta,<}$ around $\theta^*$, and are therefore the same in both settings. The only notable difference is that one shows the fact that $\theta^*$ minimizes $\tilde{\Lambda}_p^*$ not only on $\partial D_{\theta,<}$, as in (5.81), but also on $\overline{D}_{\theta,<}$, by using the fact that $\tilde{\Lambda}_p^*$ is strictly decreasing on $(0, e^{m_p})$ instead of it being strictly increasing on $\theta \in (e^{m_p}, 1)$. Beyond this, the proof is the same as for Theorem 5.3.1 i) and is hence omitted here. $\square$

# Bibliography


[1] Aldaz, J. M. Selfimprovement of the inequality between arithmetic and geometric means. *J. Math. Inequal. 2*, 4 (2008), 473–477.

[2] Aldaz, J. M. Concentration of the ratio between the geometric and arithmetic means. *J. Theor. Probab. 23*, 2 (2010), 498–508.

[3] Alonso-Gutiérrez, D., and Prochno, J. Thin-shell concentration for random vectors in Orlicz balls via moderate deviations and Gibbs measures.

[4] Alonso-Gutiérrez, D., Prochno, J., and Thäle, C. Large deviations for high-dimensional random projections of $\ell_p^n$-balls. *Adv. in Appl. Math. 99* (2018), 1–35.

[5] Alonso-Gutiérrez, D., Prochno, J., and Thäle, C. Gaussian fluctuations for high-dimensional random projections of $\ell_p^n$-balls. *Bernoulli 25*, 4A (2019), 3139–3174.

[6] Anderson, G. W., Guionnet, A., and Zeitouni, O. *An Introduction to Random Matrices*, vol. 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[7] Andriani, C., and Baldi, P. Sharp estimates of deviations of the sample mean in many dimensions. *Ann. Inst. H. Poincaré Probab. Statist. 33*, 3 (1997), 371–385.

[8] Anttila, M., Ball, K., and Perissinaki, I. The central limit problem for convex bodies. *Trans. Amer. Math. Soc. 355*, 12 (2003), 4723–4735.

[9] Artstein-Avidan, S., Giannopoulos, A., and Milman, V. D. *Asymptotic Geometric Analysis. Part I*, vol. 202 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2015.

[10] Bahadur, R. R., and Ranga Rao, R. On deviations of the sample mean. *Ann. Math. Statist. 31*, 4 (1960), 1015–1027.

[11] Barthe, F., and Cordero-Erausquin, D. Invariances in variance estimates. *Proc. Lond. Math. Soc. (3) 106*, 1 (2013), 33–64.

[12] Barthe, F., Gamboa, F., Lozada-Chang, L., and Rouault, A. Generalized Dirichlet distributions on the ball and moments. *ALEA Lat. Am. J. Probab. Math. Stat. 7* (2010), 319–340.

[13] Barthe, F., Guédon, O., Mendelson, S., and Naor, A. A probabilistic approach to the geometry of the $\ell_p^n$-ball. *Ann. Probab. 33*, 2 (2005), 480–513.

[14] Barthe, F., and Wolff, P. Volume properties of high-dimensional Orlicz balls. *arXiv:2106.01675* (2021).

[15] Bleistein, N., and Handelsmann, R. A. *Asymptotic Expansions of Integrals*. Dover Publications Inc., New York, 1975.

[16] Bogachev, V. I., and Ruas, M. A. S. *Measure Theory*, vol. 1. Springer, 2007.

[17] Borovkov, A. A., and Rogozin, B. A. On the central limit theorem in the higher-dimensional case. *Teor. Verojatnost. i Primenen. 10* (1965), 61–69.

[18] Brazitikos, S., Giannopoulos, A., Valettas, P., and Vritsiou, B.-H. *Geometry of Isotropic Convex Bodies*, vol. 196 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2014.

[19] Breitung, K., and Hohenbichler, M. Asymptotic approximations for multivariate integrals with an application to multinormal probabilities. *J. Multivar. Anal. 30*, 1 (1989), 80–97.

[20] Chafaï, D., Guédon, O., Lecué, G., and Pajor, A. *Interactions between Compressed Sensing Random Matrices and High Dimensional Geometry.* Société Mathématique de France, 2012.

[21] Chaganty, N. R., and Sethuraman, J. Large deviation local limit theorems for arbitrary sequences of random variables. *Ann. Probab.* (1985), 97–114.

[22] Chaganty, N. R., and Sethuraman, J. Multidimensional large deviation local limit theorems. *J. Multivar. Anal. 20*, 2 (1986), 190–204.

[23] Chaganty, N. R., and Sethuraman, J. Strong large deviation and local limit theorems. *Ann. Probab.* (1993), 1671–1690.

[24] Chaganty, N. R., and Sethuraman, J. Multidimensional strong large deviation theorems. *J. Statist. Plann. Inference 55*, 3 (1996), 265–280.

[25] Cramér, H. Sur un nouveau théorème-limite de la théorie des probabilités. *Actualité Sci. Indust. 736* (1938), 5–23.

[26] Daniels, H. E. Saddlepoint approximations in statistics. *Ann. Math. Statist. 25*, 4 (1954), 631–650.

[27] DasGupta, A. *Fundamentals of Probability: a First Course.* Springer Science & Business Media, 2010.

[28] Debye, P. Näherungsformeln für die Zylinderfunktionen für große Werte des Arguments und unbeschränkt veränderliche Werte des Index. *Math. Ann. 67* (1909), 535–558.

[29] Dembo, A., and Zeitouni, O. *Large Deviations. Techniques and Applications*, vol. 38 of *Stochastic Modelling and Applied Probability.* Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.

[30] den Hollander, F. *Large Deviations*, vol. 14 of *Fields Institute Monographs.* American Mathematical Society, Providence, RI, 2000.

[31] Diaconis, P., and Freedman, D. A dozen de Finetti-style results in search of a theory. *Ann. Inst. H. Poincaré Probab. Statist. 23*, 2, suppl. (1987), 397–423.

[32] Dirichlet, P. Sur une nouvelle méthode pour la détermination des intégrales multiples. *J. Math. Pures Appl. 4* (1839), 164–168.

[33] Esscher, F. On the probability function in the collective theory of risk. *Scand. Actuar. J. 1932*, 3 (1932), 175–195.

[34] Foucart, S., and Rauhut, H. *A mathematical Introduction to Compressive Sensing.* Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, 2013.

[35] Frühwirth, L., and Prochno, J. Sanov-type large deviations and conditional limit theorems for high-dimensional Orlicz balls. *arXiv:2111.04691* (2021).

[36] Gantert, N., and Kim, S., and Ramanan, K. Large deviations for random projections of $\ell^p$ balls. *Ann. Probab. 45* (2017), 4419–4476.

[37] Gelfand, I. M., and Fomin, S. *Calculus of Variations.* Dover Publications Inc., Mineola, New York, 2000.

[38] Gluskin, E., and Milman, V. Note on the geometric-arithmetic mean inequality. In *Geometric Aspects of Functional Analysis.* Springer, 2003, pp. 131–135.

[39] Goldman, R. Curvature formulas for implicit curves and surfaces. *Comput. Aided Geom. Des. 22*, 7 (2005), 632 – 658. Geometric Modelling and Differential Geometry.

[40] Gowers, W. T. A solution to Banach's hyperplane problem. *Bull. London Math. Soc. 26*, 6 (1994), 523–530.

[41] Gowers, W. T., and Maurey, B. The unconditional basic sequence problem. *J. Am. Math. Soc 6*, 4 (1993), 851–874.

[42] Grafakos, L. *Modern Fourier Analysis.* Graduate Texts in Mathematics. Springer New York, 2009.

[43] Grafakos, L. *Classical Fourier Analysis*, vol. 3. Springer, 2014.

[44] Gromov, M., and Milman, V. D. Generalization of the spherical isoperimetric inequality to uniformly convex Banach spaces. *Compos. Math. 62*, 3 (1987), 263–282.

[45] Guédon, O. Concentration phenomena in high dimensional geometry. In *Journées MAS 2012*, vol. 44 of *ESAIM Proc.* EDP Sci., Les Ulis, 2014, pp. 47–60.

[46] Guédon, O., Nayar, P., and Tkocz, T. Concentration inequalities and geometry of convex bodies. In *Analytical and probabilistic methods in the geometry of convex bodies*, vol. 2 of *IMPAN Lect. Notes.* Polish Acad. Sci. Inst. Math., Warsaw, 2014, pp. 9–86.

[47] Guédon, O., and Paouris, G. Concentration of mass on the Schatten classes. *Ann. Inst. H. Poincaré Probab. Statist. 43*, 1 (2007), 87–99.

[48] Hicks, N. J. *Notes on Differential Geometry.* Van Nostrand Company, 1965.

[49] Hinrichs, A., Krieg, D., Novak, E., Prochno, J., and Ullrich, M. Random sections of ellipsoids and the power of random information. *Trans. Am. Math. Soc. 374*, 12 (2021), 8691–8713.

[50] Hinrichs, A., Prochno, J., and Sonnleitner, M. Random sections of $\ell_p$-ellipsoids, optimal recovery and Gelfand numbers of diagonal operators. *arXiv:2109.14504* (2021).

[51] Hinrichs, A., Prochno, J., and Ullrich, M. The curse of dimensionality for numerical integration on general domains. *J. Complex. 50* (2019), 25–42.

[52] Hinrichs, A., Prochno, J., and Vybíral, J. Entropy numbers of embeddings of Schatten classes. *J. Funct. Anal. 273*, 10 (2017), 3241–3261.

[53] Hinrichs, A., Prochno, J., and Vybíral, J. Gelfand numbers of embeddings of Schatten classes. *Math. Ann.* (2021).

[54] Hogg, R., and Craig, A. *Introduction to Mathematical Statistics*. Macmillan Publishing Co. Inc., 1978.

[55] Jameson, G. Inequalities for gamma function ratios. *Amer. Math. Monthly 120*, 10 (2013), 936–940.

[56] Johnston, S., and Prochno, J. A Maxwell principle for generalized Orlicz balls. *arXiv:2012.11568* (2020).

[57] Joutard, C. Asymptotic approximation for the probability density function of an arbitrary sequence of random variables. *Statist. Probab. Lett. 90* (2014), 100–107.

[58] Joutard, C. A large deviation approximation for multivariate density functions. *Math. Methods Statist. 28*, 1 (2019), 66–73.

[59] Kabluchko, Z., and Prochno, J. Large deviations for random matrices in the orthogonal group and Stiefel manifold with applications to random projections of product distributions. *arXiv:2110.12977* (2021).

[60] Kabluchko, Z., and Prochno, J. The maximum entropy principle and volumetric properties of Orlicz balls. *J. Math. Anal. Appl. 495*, 1 (2021), 124687.

[61] Kabluchko, Z., Prochno, J., and Thäle, C. High-dimensional limit theorems for random vectors in $\ell_p^n$-balls. *Commun. Contemp. Math. 21*, 1 (2019), 1750092, 30.

[62] Kabluchko, Z., Prochno, J., and Thäle, C. Sanov-type large deviations in Schatten classes. *Ann. Inst. H. Poincaré Probab. Statist. 56*, 2 (2019), 928–953.

[63] Kabluchko, Z., Prochno, J., and Thäle, C. Intersection of unit balls in classical matrix ensembles. *Israel J. Math. 239*, 1 (2020), 129–172.

[64] Kabluchko, Z., Prochno, J., and Thäle, C. Exact asymptotic volume and volume ratio of Schatten unit balls. *Journal of Approximation Theory 257* (2020), 105457.

[65] Kabluchko, Z., Prochno, J., and Thäle, C. High-dimensional limit theorems for random vectors in $\ell_p^n$-balls. II. *Commun. Contemp. Math. 23*, 03 (2021), 1950073.

[66] Kabluchko, Z., Prochno, J., and Vysotsky, V. Yet another note on the arithmetic-geometric mean inequality. *Studia Math. 253* (2020), 39–55.

[67] Kallenberg, O. *Foundations of Modern Probability*, second ed. Probability and its Applications. Springer-Verlag, New York, 2002.

[68] Kallenberg, O. *Random Measures, Theory and Applications*, vol. 77 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 2017.

[69] Kaufmann, T. Sharp asymptotics for $q$-norms of random vectors in high-dimensional $\ell_p^n$-balls. *Mod. Stoch.: Theory Appl. 8*, 2 (2021), 239–274.

[70] Kaufmann, T., Sambale, H., and Thäle, C. Large deviations for uniform projections of $p$-radial distributions on $\ell_p^n$-balls. *arXiv:2203.00476* (2022).

[71] Kaufmann, T., and Thäle, C. Sharpening the probabilistic arithmetic-geometric mean inequality. *arXiv:2112.04340* (2021).

[72] Kaufmann, T., and Thäle, C. Weighted $p$-radial distributions on Euclidean and matrix $p$-balls with applications to large deviations. *J. Math. Anal. Appl. 515*, 1 (2022), 126377.

[73] Kim, S. Problems at the interface of probability and convex geometry: Random projections and constrained processes. *Ph.D. thesis, Brown University* (2017).

[74] Kim, S., and Ramanan, K. A conditional limit theorem for high-dimensional $\ell^p$-spheres. *J. Appl. Probab. 55* (2018), 1060–1077.

[75] Kim, S. S., Liao, Y.-T., and Ramanan, K. An asymptotic thin shell condition and large deviations for random multidimensional projections. *Adv. Appl. Math 134* (2022), 102306.

[76] Kim, S. S., and Ramanan, K. Large deviation principles induced by the Stiefel manifold, and random multi-dimensional projections. *arXiv:2105.04685* (2021).

[77] Klartag, B. A central limit theorem for convex sets. *Invent. Math. 168*, 1 (2007), 91–131.

[78] Klartag, B. Power-law estimates for the central limit theorem for convex sets. *J. Funct. Anal. 245*, 1 (2007), 284–310.

[79] Klingenberg, W., and Hoffman, D. *A Course in Differential Geometry.* Graduate Texts in Mathematics. Springer New York, 2013.

[80] König, H., Meyer, M., and Pajor, A. The isotropy constants of the Schatten classes are bounded. *Math. Ann. 312*, 4 (1998), 773–783.

[81] König, W. *Große Abweichungen: Techniken und Anwendungen.* Springer, 2020.

[82] Krantz, S. G., and Parks, H. R. *Geometric Integration Theory.* Springer Science & Business Media, 2008.

[83] Krieg, D., and Sonnleitner, M. Random points are optimal for the approximation of Sobolev functions. *arXiv:2009.11275* (2020).

[84] Lang, S. *Complex Analysis*, vol. 103. Springer Science & Business Media, 2003.

[85] Liao, Y.-T., and Ramanan, K. Geometric sharp large deviations for random projections of $\ell_p^n$ spheres and balls. *arXiv:2001.04053v3* (2020).

[86] Lindenstrauss, J., and Milman, V. The local theory of normed spaces and its applications to convex geometry. In *Handbook of Convex Geometry, Part B*, P. Gruber and J. Wills, Eds. North-Holland Amsterdam, 1993, pp. 1149–1220.

[87] Ludwig, M., Milman, V. D., Pestov, V., and Tomczak-Jaegermann, N. *Asymptotic Geometric Analysis: Proceedings of the Fall 2010 Fields Institute Thematic Program*, vol. 68. Springer Science & Business Media, 2013.

[88] Meyer, C. D. *Matrix Analysis and Applied Linear Algebra*, vol. 71. Siam, 2000.

[89] Milman, V. D. Geometrization of probability. In *Geometry and Dynamics of Groups and Spaces*. Springer, 2007, pp. 647–667.

[90] Milman, V. D., and Tomczak-Jaegermann, N. Asymptotic $l_p$ spaces and bounded distortions. *Contemp. Math. 144* (1993), 173–173.

[91] Mogul'skii, A. A. De Finetti-type results for $\ell_p$. *Sib. Math. J. 32*, 4 (1991), 609–616.

[92] Naor, A. The surface measure and cone measure on the sphere of $\ell_p^n$. *Trans. Amer. Math. Soc. 359*, 3 (2007), 1045–1079.

[93] Naor, A., and Romik, D. Projecting the surface measure of the sphere of $\ell_p^n$. *Ann. Inst. H. Poincaré Probab. Statist. 39*, 2 (2003), 241–261.

[94] Paouris, G., and Werner, E. Relative entropy of cone measures and $L_p$ centroid bodies. *Proc. Lond. Math. Soc. (3) 104*, 2 (2012), 253–286.

[95] Petrov, V. V. On the probabilities of large deviations for sums of independent random variables. *Teor. Verojatnost. i Primenen 10* (1965), 310–322.

[96] Pietsch, A. What is "local theory of Banach spaces"? *Studia Math. 135* (1999), 273–298.

[97] Prochno, J., and Strzelecki, M. Approximation, Gelfand, and Kolmogorov numbers of Schatten class embeddings. *J. Approx. Theory 277* (2022), 105736.

[98] Prochno, J., Thäle, C., and Turchi, N. Geometry of $\ell_p^n$-balls: Classical results and recent developments. In *High Dimensional Probability VIII*, Progress in Probability. N. Gozlan, R. Latala, K. Loucini, M. Madiman eds. Birkhäuser, 2019.

[99] Rachev, S., and Rüschendorf, L. Approximate independence of distributions on spheres and their stability properties. *Ann. Probab. 19*, 3 (1991), 1311–1337.

[100] Radke, J., and Vritsiou, B.-H. On the thin-shell conjecture for the Schatten classes. *Ann. Inst. H. Poincaré Probab. Statist. 56*, 1 (2020), 87–119.

[101] Richter, W. Local limit theorems for large deviations. *Dokl. Akad. Nauk SSSR (N.S.) 115* (1957), 53–56.

[102] Richter, W. Mehrdimensionale lokale Grenzwertsätze für grosse Abweichungen. *Teor. Veroyatnost. i Primenen 3* (1958), 107–114.

[103] Rockafellar, R. T. *Convex Analysis*, vol. 36. Princeton University Press, 1970.

[104] Saint Raymond, J. Le volume des idéaux d'opérateurs classiques. *Studia Math. 80*, 1 (1984), 63–75.

[105] Schechtman, G., and Schmuckenschläger, M. Another remark on the volume of the intersection of two $L_p^n$ balls. In *Geometric aspects of functional analysis (1989–90)*, vol. 1469 of *Lecture Notes in Math.* Springer, Berlin, 1991, pp. 174–178.

[106] Schechtman, G., and Zinn, J. On the volume of the intersection of two $L_p^n$ balls. *Proc. Amer. Math. Soc. 110*, 1 (1990), 217–224.

[107] Schechtman, G., and Zinn, J. Concentration on the $l_p^n$ ball. In *Geometric Aspects of Functional Analysis*, vol. 1745 of *Lecture Notes in Math.* Springer, Berlin, 2000, pp. 245–256.

[108] Schmuckenschläger, M. CLT and the volume of intersections of $l_p^n$-balls. *Geom. Dedicata 85*, 1-3 (2001), 189–195.

[109] Sodin, S. An isoperimetric inequality on the $l_p$ balls. *Ann. Inst. H. Poincaré Probab. Statist. 44*, 2 (2008), 362–373.

[110] Stein, E. M., and Shakarchi, R. *Fourier Analysis: an Introduction*, vol. 1. Princeton University Press, 2011.

[111] Thäle, C. Another note on the inequality between geometric and $p$-generalized arithmetic mean. *Math. Nachr. 294*, 10 (2021), 1977–1986.

[112] Varadhan, S. Asymptotic probabilities and differential equations. *Commun. Pure Appl. Math. 19*, 3 (1966), 261–286.

[113] Varadhan, S. *Large Deviations*, vol. 27 of *Courant Lecture Notes in Mathematics.* Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2016.

[114] Vershynin, R. *High-Dimensional Probability: An Introduction with Applications in Data Science.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.

[115] Vritsiou, B.-H. Selberg-type integrals and the variance conjecture for the operator norm. *arXiv: 1805.03129* (2018).