NAWI Graz
Natural Sciences

TU Graz

Lorenz Frühwirth, BSc

An Asymptotic Thin Shell Condition and Large Deviations for Random Multidimensional Projections

MASTER'S THESIS

to achieve the university degree of

Diplom-Ingenieur

Master's degree programme: Mathematics

submitted to

Graz University of Technology

Supervisor

Dipl.-Math. Dr. Joscha Karl PROCHNO
Institut für Mathematik und wissenschaftliches Rechnen,
Karl-Franzens-Universität Graz

Graz, December 2020

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis.

Date, Signature

Abstract

In this thesis, we are dealing with the paper „An asymptotic thin shell condition and large deviations for random multidimensional projections“ by Kim, Liao, and Ramanan, where a large deviation analogue to Klartag's famous central limit theorem is proven for various convex bodies under an asymptotic thin-shell condition, unifying several of the results that have appeared in the past 3 years.

The introduction provides a brief historical overview of large deviation theory, followed by a short discussion about related works. In the preliminary chapter we introduce the formal background and some fundamental results together with a few examples. In the main part we state the necessary thin-shell assumptions and present most of the results from ["An asymptotic thin shell condition and large deviations for random multidimensional projections". arXiv:1912.13447v2 (2020)]. In particular, we show that certain classes of Orlicz balls and Gibbs measures satisfy those assumptions.

Kurzfassung

In dieser Diplomarbeit beschäftigen wir uns mit der Arbeit „An asymptotic thin shell condition and large deviations for random multidimensional projections“ von Kim, Liao, und Ramanan. Ausgehend von Klartag’s berühmten zentralen Grenzwertsatz werden hier Resultate aus der Theorie der großen Abweichungen für verschiedene konvexe Körper, unter Annahme einer asymptotischen „thin-shell“ Bedingung, bewiesen.

In der Einleitung wird ein Überblick über die historische Entwicklung der Theorie der großen Abweichungen gegeben, gefolgt von einer Disskusion der aktuellen Forschung. Im präliminaren Kapitel wird die Theorie der großen Abweichungen zusammen mit einigen grundlegenden Resultaten und diversen Anwendungen eigeführt. Im Hauptteil der Arbeit formulieren wir die nötigen „thin-shell“ Annahmen und präsentieren die meisten Resultate aus ["An asymptotic thin shell condition and large deviations for random multidimensional projections". arXiv:1912.13447v2 (2020)]. Im Speziellen zeigen wir, dass bestimmte Orlicz Bälle und Gibbs Maße diese Annahmen erfüllen.

1 Danksagung

Mein vorderster Dank geht an meine Eltern, die mich immer unterstützt haben und mir mein Studium ermöglichten.

Ebenso will ich meinem Betreuer Joscha danken, der mich in das spannende Gebiet dieser Arbeit eingeführt hat. Ich freue mich auf die weitere Zusammenarbeit!

Darüber hinaus möchte ich meinen Bruder Samuel und meine engsten Freunde erwähnen, die mir immer eine moralische Stütze waren und die mir im schwierigen vergangenen Jahr viel Kraft gaben.

Contents

2 Introduction

The history of large deviation theory, as many other mathematical topics, has its origins in physics. In the second half of the 19^{th} century one was interested in using probabilistic methods to study the behaviour of systems including many particles, such models are used for example in fluid mechanics or magnetism. Many authors mention in this context the 1877 published work [6], where Boltzmann was interested in entropy of thermodynamic systems (see also [12]). It took some time before a systematic mathematical approach followed. In 1938 an important work of Cramér was published (see [8] for a commented and translated version). Cramér was interested in improving the normal approximation, to this end he was looking for appropriate correction terms. He introduced new techniques for his proofs, like the Esscher transform, also known as exponential tilting, and also the Legendre transform of the cumulant generating function appears for the first time in this context. After this groundbreaking work, it took another 30 years, before Varadhan (see [26]) introduced a general approach to large deviations. In the following years Varadhan and Donsker published several papers and developed the theory (see [27] for references and further information). Parallel to that Freidlin and Wentzell (see [13] for a translated version of the original work) developed their approach. They had the idea to solve ordinary differential equations by considering a similar family of stochastic differential equations containing small random perturbations. Under suitable conditions the limit theorems of probability theory appear if one considers the family of respective solutions. In this context the study of large deviations came into focus. In [11] one finds a formal introduction to statistical mechanics and the relation to large deviation theory.

In recent years large deviations appeared in various mathematical fields. In particular applications in the context of asymptotic geometric analysis became increasingly popular. In [4] the authors defined a certain concentration hypothesis, referred to as the „thin shell“ condition. The latter plays a major role in Klartag's famous CLT (see [20]) for log-concave, isotropic distributions. Roughly speaking, Klartag proved that for sufficiently large n and „most" directions $\theta^{(n)} \in \mathbb{S}^{n-1}$ we have

$$\langle \theta^{(n)}, X^{(n)} \rangle \approx \mathcal{N}(0,1).$$

Based on this, the question of large deviation results concerning random projections of high dimensional random vectors came into focus. In 2016 Gantert, Kim and Ramanan (see [14]) looked at Cramér's theorem from a geometrical point of view and investigated that in the case of vectors $X^{(n)}$ with iid components, $\langle \theta^{(n)}, X^{(n)} \rangle$ satisfies an LDP for „most" sequences of directions $\{\theta^{(n)}\}_{n\in\mathbb{N}}$. The corresponding rate function is independent of $\theta^{(n)}$ and does not coincide with the one from Cramér's theorem. In that sense Cramér's theorem is atypical. In 2017 Gantert, Kim and Ramanan [15] proved respective large deviation principles for $X^{(n)}$ uniformly distributed on n dimensional ℓ_p balls and independent random directions $\theta^{(n)}$. Moreover they showed „quenched" large deviation principles for fixed directions $\theta^{(n)}$. In a short time several related works were published, where either one dimensional projections are analysed or scaled Euclidean norms of multidimensional projections (see [1], [16]). In most papers one restricts to random vectors uniformly distributed on ℓ_p balls or spheres (see [22]), where one uses well known probabilistic representations. An exception had been proved by Kabluchko, Prochno and Thäle for empirical spectral measures of random matrices in Schatten class unit balls [17]. In the recent work [2] the authors established a relation of large deviations for isotropic log-concave random vectors to the famous KLS conjecture. Eventually, in [18] the authors succeeded to treat quite general classes of sequences of random variables $X^{(n)}$ (including the special case of uniform distributions on ℓ_p balls). Moreover, large deviation principles for non-scalar projections on k-dimensional subspaces are treated, where also the case $k = k_n$ with growing k_n is analysed. Especially the situation of $k_n \to \infty$ is not easy to handle, the crucial idea is the transition to empirical measures, where one can use a more familiar setting.

3 Preliminaries

3.1 Results and techniques from large deviation theory

3.1.1 Formal setting and basic properties

In this section we will establish several auxiliary results and present techniques, which will be needed for the main part of the thesis. Let us start with the formal setting required in order to define an abstract large deviation principle.

Definition 3.1.1. *Let $(\mathcal{X}, \tau)$ be a topological space, which we will call a Polish space iff all of the following are fulfilled:*

1. *$\mathcal{X}$ contains a countable dense subset.*
2. *There is a metric d such that τ is induced by d.*
3. *The metric space $(\mathcal{X}, d)$ is complete.*

The following definitions and results can be found in [9, Chapter 1].

Definition 3.1.2. *Let $\mathbb{I} : \mathcal{X} \to [0, \infty]$ not identically ∞ be a lower semi-continuous function, i.e. for all $\alpha \geq 0$*

$$\psi_{\mathbb{I}}(\alpha) := \{x \in \mathcal{X} \mid \mathbb{I}(x) \leq \alpha\}$$

is a closed subset of $\mathcal{X}$. Then $\mathbb{I}$ is called a rate function. Moreover, if $\psi_{\mathbb{I}}(\alpha)$ is compact, then $\mathbb{I}$ is called a good rate function (GRF).

Lemma 3.1.1. *A rate function $\mathbb{I}$ attains its infimum over every compact non-empty set. If $\mathbb{I}$ is a GRF, then this is even true for any closed set.*

Proof. Let $K \subseteq \mathcal{X}$ be a non-empty compact set and consider the case $\inf_{x \in K} \mathbb{I}(x) < \infty$. Then there exists a convergent and strictly decreasing sequence $\{a_n\}_{n \in \mathbb{N}} \subseteq \mathbb{R}$ with

$$\lim_{n \to \infty} a_n = \inf_{x \in K} \mathbb{I}(x).$$

We introduce the sets $K_n = K \cap \psi_{\mathbb{I}}(a_n)$ and claim that they are non-empty. Assume there exists $n \in \mathbb{N}$ with $K_n = \varnothing$, then for all $x \in K : \mathbb{I}(x) > a_n$ and hence

$$\inf_{x \in K} \mathbb{I}(x) \geq a_n > \inf_{x \in K} \mathbb{I}(x),$$

a contradiction. Moreover, K_n is closed as intersection of a compact set with a closed set. The sets K_n are even compact since $\mathcal{X}$ is a metric space, where closed subsets of compact sets are again compact. Cantor's theorem 6.0.5 implies that their intersection is non-empty, hence we find $x_0 \in \bigcap_{n \in \mathbb{N}} K_n$ with $x_0 \in K$ and $\mathbb{I}(x_0) = \inf_{x \in K} \mathbb{I}(x)$.

Now assume $\mathbb{I}$ is a GRF and take a closed set $C \subseteq \mathcal{X}$. If $\inf_{x \in C} \mathbb{I}(x) = \infty$, then the assertion is true. If $\inf_{x \in C} \mathbb{I}(x) = M < \infty$, then for the non-empty compact set $K = C \cap \psi_{\mathbb{I}}(M+1)$ we have

$$\inf_{x \in C} \mathbb{I}(x) = \inf_{x \in K} \mathbb{I}(x)$$

and the result follows from the one for compact sets. □

Using the previous notation, we can now define a general LDP.

Definition 3.1.3. *Let $(\mathcal{X}, d)$ be a Polish space with Borel σ-algebra $\mathcal{B}(\mathcal{X})$. A family $\{\mu_\epsilon\}_{\epsilon > 0}$ of probability measures satisfies an LDP with rate function $\mathbb{I}$ iff for all $\Gamma \in \mathcal{B}$*

$$-\inf_{x \in \mathring{\Gamma}} \mathbb{I}(x) \leq \liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(\Gamma) \leq \limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(\Gamma) \leq -\inf_{x \in \overline{\Gamma}} \mathbb{I}(x),$$

where $\mathring{\Gamma}$ and $\overline{\Gamma}$ denote the interior and the closure of Γ respectively.

Alternatively one can define an LDP via the following inequalities.
a) (Upper bound) For every closed $F \subseteq \mathcal{X}$

$$\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(F) \leq -\inf_{x \in F} \mathbb{I}(x). \tag{3.1.1}$$

b) (Lower bound) For every open $G \subseteq \mathcal{X}$

$$-\inf_{x \in G} \mathbb{I}(x) \leq \liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(G). \tag{3.1.2}$$

Remark. This definition is possible for any topological space, we do not need to restrict on Polish spaces here. But we will usually deal with Polish spaces in this thesis.

Remark. One can easily show that a rate function for an LDP is unique (see e.g. Lemma 4.1.4 in [9]).

Most of the time we will work with random variables, where we want to find associated LDPs, hence we extend our definition.

Definition 3.1.4. *Consider a family of random variables $\{X_\epsilon\}_{\epsilon>0}$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in a Polish space $\mathcal{X}$. We say that $\{X_\epsilon\}_{\epsilon>0}$ satisfies an LDP on $\mathcal{X}$ if the respective family of distributions $\{\mu_\epsilon\}_{\epsilon>0}$ with $\mu_\epsilon(\cdot) := \mathbb{P}[X_\epsilon \in \cdot]$ satisfies an LDP.*

Since we want to establish some techniques, beginning with the so-called contraction principle, we will now define the notion of weak LDP.

Definition 3.1.5. *(weak LDP)*
A family of probability measures $\{\mu_\epsilon\}_{\epsilon>0}$ is said to satisfy the weak LDP with rate function $\mathbb{I}$ if the upper bound 3.1.1 holds for all compact sets and the lower bound 3.1.2 holds for all open sets.

Definition 3.1.6. *(exponential tightness)*
A family of probability measures $\{\mu_\epsilon\}_{\epsilon>0}$ on $\mathcal{X}$ is exponentially tight if for every $\alpha < \infty$, there exists a compact set $K_\alpha \subseteq \mathcal{X}$ such that

$$\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(K_\alpha) < -\alpha.$$

The next lemma gives a useful relation between weak and full LDPs as defined in Definition 3.1.3.

Lemma 3.1.2. *(Lemma 1.2.18 of [9])*
Let $\{\mu_\epsilon\}_{\epsilon>0}$ be an exponentially tight family.
a) If the upper bound 3.1.1 holds for all compact subsets of $\mathcal{X}$, then it also holds for all closed sets.
b) If the lower bound 3.1.2 holds for all open sets, then $\mathbb{I}$ is a good rate function.

This means, if an exponentially tight family of measures satisfies the weak LDP with rate function $\mathbb{I}$, then $\mathbb{I}$ is a GRF and the full LDP holds.

3.1.2 Contraction principles and exponential equivalence

One may be interested in proving that a certain family of measures satisfies an LDP, which seems quite delicate at this point. A very useful tool is the contraction principle, which we will establish in this subsection. Suppose a Polish space $(\mathcal{X}, d)$ with induced topology τ. Recall that a base $\mathcal{A} \subseteq \tau$ of τ is a system of open sets such that for every $G \in \tau$ and $x \in G$ there exists an $A \in \mathcal{A}$ with $x \in A \subseteq G$.

Theorem 3.1.3. *(Theorem 4.1.11 in [9])*
Let $\mathcal{A}$ be a base of the topology of $\mathcal{X}$. For every $A \in \mathcal{A}$ define

$$\mathcal{L}_A := -\liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(A)$$

and for $x \in \mathcal{X}$

$$\mathbb{I}(x) := \sup_{\{A \in \mathcal{A} : x \in A\}} \mathcal{L}_A.$$

Suppose that for all $x \in \mathcal{X}$

$$\mathbb{I}(x) = \sup_{A \in \mathcal{A}: x \in A} -\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(A).$$

Then $\{\mu_\epsilon\}_{\epsilon>0}$ satisfies the weak LDP with rate function $\mathbb{I}$.

The next theorem may be considered as the converse statement to Theorem 3.1.3.

Theorem 3.1.4. *(Theorem 4.1.18 in [9])*
Suppose that $\{\mu_\epsilon\}_{\epsilon>0}$ satisfies an LDP with rate function $\mathbb{I}$. Then for any base $\mathcal{A}$ of the topology of $\mathcal{X}$, and for any $x \in \mathcal{X}$,

$$\begin{aligned}\mathbb{I}(x) &= \sup_{A \in \mathcal{A}: x \in A} -\liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(A) \\ &= \sup_{A \in \mathcal{A}: x \in A} -\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(A).\end{aligned}$$

Theorem 3.1.5. *(Contraction principle, Theorem 4.2.1 in [9])*
Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces and $f : \mathcal{X} \to \mathcal{Y}$ be a continuous function. Consider a good rate function $\mathbb{I} : \mathcal{X} \to [0, \infty]$.
a) For each $y \in \mathcal{Y}$ define

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

Then $\mathbb{I}'$ is a GRF in $\mathcal{Y}$, where the infimum over the empty set is taken as ∞.
b) If $\mathbb{I}$ controls the LDP associated with a family of probability measures $\{\mu_\epsilon\}_{\epsilon>0}$ on $\mathcal{X}$, then $\mathbb{I}'$ controls the LDP associated with the family of probability measures $\{\mu_\epsilon \circ f^{-1}\}_{\epsilon>0}$ on $\mathcal{Y}$.

Proof. a) Clearly, $\mathbb{I}'$ is non-negative. Since $f^{-1}(\{y\})$ is a closed set for $y \in f(\mathcal{X})$, we can use that a GRF attains its infimum over all non-empty closed sets by Lemma 3.1.1. Hence, we see

$$\psi_{\mathbb{I}'}(\alpha) = \{y : \mathbb{I}'(y) \leq \alpha\} = \{f(x) : \mathbb{I}(x) \leq \alpha\} = f(\psi_{\mathbb{I}}(\alpha)).$$

Since $\psi_{\mathbb{I}}(\alpha)$ is compact in $\mathcal{X}$, $f(\psi_{\mathbb{I}}(\alpha))$ is compact in $\mathcal{Y}$ an thus $\mathbb{I}'$ is a GRF.
b)The definition of $\mathbb{I}'$ implies that for any $A \subseteq \mathcal{X}$,

$$\inf_{y \in A} \mathbb{I}'(y) = \inf_{x \in f^{-1}(A)} \mathbb{I}(x).$$

Since f is continuous, the set $f^{-1}(A)$ is open (closed) in $\mathcal{X}$ for any open (closed) $A \subseteq \mathcal{Y}$. Thus the LDP for $\{\mu_\epsilon \circ f^{-1}\}_{\epsilon>0}$ follows as a consequence of the LDP for $\{\mu_\epsilon\}_{\epsilon>0}$. □

The contraction principle will appear frequently, since we often use continuous mappings to transfer random variables from one space to another. Hence, the following corollary is very useful and can be deduced directly from Theorem 3.1.5.

Corollary. Let $\mathcal{X}$ and $\mathcal{Y}$ be two Polish spaces and $f : \mathcal{X} \to \mathcal{Y}$ a continuous mapping. Further we consider a family of random variables $\{X_\epsilon\}_{\epsilon>0}$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in $\mathcal{X}$. We assume that $\{X_\epsilon\}_{\epsilon>0}$ satisfies an LDP with GRF $\mathbb{I}$. Then $Y_\epsilon := f(X_\epsilon)$ satisfies an LDP in $\mathcal{Y}$ with GRF $\mathbb{I}' : \mathcal{Y} \to [0, \infty]$,

$$\mathbb{I}'(y) := \inf_{x \in f^{-1}(\{y\})} \mathbb{I}(x). \tag{3.1.3}$$

Proof. The claim follows immediately by applying Theorem 3.1.5 to the laws $\mu_\epsilon(\cdot) := \mathbb{P}[X_\epsilon \in \cdot]$. □

Also the following lemma will appear several times in the main part of this thesis. For the proof we make use of the previous theorems.

Lemma 3.1.6. *(Exercise 4.2.7 in [9])*
Consider a Polish space $\mathcal{X}$ and assume that for all $\epsilon > 0$, (X_ϵ, Y_ϵ) is distributed according to the product measure $\mu_\epsilon \otimes \nu_\epsilon$ on $\mathcal{B}_{\mathcal{X}} \otimes \mathcal{B}_{\mathcal{X}}$ (the product of the Borel sigma-algebra's). Assume that $\{\mu_\epsilon\}_{\epsilon>0}$ satisfies the LDP with the GRF $\mathbb{I}_X$ while $\{\nu_\epsilon\}_{\epsilon>0}$ satisfies the LDP with the GRF $\mathbb{I}_Y$, and both $\{\mu_\epsilon\}_{\epsilon>0}$ and $\{\nu_\epsilon\}_{\epsilon>0}$ are exponentially tight. Then for any continuous function $F : \mathcal{X} \times \mathcal{X} \to \mathcal{Y}$, the family of laws induced on $\mathcal{Y}$ by $Z_\epsilon = F(X_\epsilon, Y_\epsilon)$ satisfies the LDP with GRF $\mathbb{I}_Z : \mathcal{X} \times \mathcal{X} \to [0, \infty]$,

$$\mathbb{I}_Z(z) := \inf_{\{(x,y) \in \mathcal{X} \times \mathcal{X} : z = F(x,y)\}} \mathbb{I}_X(x) + \mathbb{I}_Y(y).$$

Proof. We prove that under the assumptions, $\{(X_\epsilon, Y_\epsilon)\}_{\epsilon>0}$ satisfies the LDP with GRF $\mathbb{I}_X + \mathbb{I}_Y$. Then the representation of $\mathbb{I}_Z$ follows as an immediate consequence of the contraction principle.

First we want to show that the product measure satisfies the weak LDP. In the following calculation we will use Theorem 6.0.9 (see Appendix) stating $\mathcal{B}_{\mathcal{X}\times\mathcal{X}} = \mathcal{B}_{\mathcal{X}} \times \mathcal{B}_{\mathcal{X}}$ as well as Theorem 3.1.3 and Theorem 3.1.4. Note that for any $A \in \mathcal{A}$ we have $A = A_1 \times A_2$, where $A_i \in \mathcal{B}_{\mathcal{X}}$ for a given base $\mathcal{A}$ of the product topology. As in Theorem 3.1.3 we consider

$$\tilde{\mathbb{I}}(x,y) = \sup_{\{A\in\mathcal{A}:(x,y)\in A\}} -\liminf_{\epsilon\to 0} \epsilon \log \mu_\epsilon \otimes \nu_\epsilon(A).$$

Using Theorem 3.1.4 for $\mathbb{I}_X$ and $\mathbb{I}_Y$

$$\begin{aligned}
\mathbb{I}_X(x) + \mathbb{I}_Y(y) &= \sup_{A_1\times A_2\in\mathcal{A}:x\in A_1} -\limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon(A_1) + \sup_{A_1\times A_2\in\mathcal{A}:y\in A_2} -\limsup_{\epsilon\to 0} \epsilon \log \nu_\epsilon(A_2) \\
&= \sup_{A_1\times A_2=A\in\mathcal{A}:(x,y)\in A} -\limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon(A_1) - \limsup_{\epsilon\to 0} \epsilon \log \nu_\epsilon(A_2) \\
&\leq \sup_{A_1\times A_2=A\in\mathcal{A}:(x,y)\in A} -\limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon \otimes \nu_\epsilon(A_1 \times A_2) \\
&\leq \sup_{A_1\times A_2=A\in\mathcal{A}:(x,y)\in A} -\liminf_{\epsilon\to 0} \epsilon \log \mu_\epsilon \otimes \nu_\epsilon(A_1 \times A_2) = \tilde{\mathbb{I}}(x,y).
\end{aligned}$$

On the other hand, again applying Theorem 3.1.4, we have that

$$\begin{aligned}
\tilde{\mathbb{I}}(x,y) &= \sup_{A_1\times A_2=A\in\mathcal{A}:(x,y)\in A} -\liminf_{\epsilon\to 0} \epsilon \log \mu_\epsilon \otimes \nu_\epsilon(A_1 \times A_2) \\
&\leq \sup_{A_1\times A_2=A\in\mathcal{A}:(x,y)\in A} -\liminf_{\epsilon\to 0} \epsilon \log \mu_\epsilon(A_1) - \liminf_{\epsilon\to 0} \epsilon \log \nu_\epsilon(A_2) \\
&= \sup_{A_1\times A_2\in\mathcal{A}:x\in A_1} -\liminf_{\epsilon\to 0} \epsilon \log \mu_\epsilon(A_1) + \sup_{A_1\times A_2\in\mathcal{A}:y\in A_2} -\liminf_{\epsilon\to 0} \epsilon \log \nu_\epsilon(A_2) \\
&= \mathbb{I}_X(x) + \mathbb{I}_Y(y).
\end{aligned}$$

This shows $\tilde{\mathbb{I}}(x,y) = \mathbb{I}_X(x) + \mathbb{I}_Y(y)$ and by using Theorem 3.1.3, we can see that $\{(X_\epsilon, Y_\epsilon)\}_{\epsilon>0}$ satisfies the weak LDP with rate function $\tilde{\mathbb{I}}$. Now we can show that $\{\mu_\epsilon \otimes \nu_\epsilon\}_{\epsilon>0}$ is exponentially tight. Choose compact sets K_α^1, K_α^2 such that

$$\begin{aligned}
&\limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon(K_\alpha^1) < -\alpha/2 \\
&\limsup_{\epsilon\to 0} \epsilon \log \nu_\epsilon(K_\alpha^2) < -\alpha/2.
\end{aligned}$$

Then for $K_\alpha = K_\alpha^1 \times K_\alpha^2$ it follows that

$$\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon \otimes \nu_\epsilon(K_\alpha) < -\alpha.$$

Applying Lemma 3.1.2 shows that $\tilde{\mathbb{I}}$ is a GRF and the LDP is satisfied. This concludes the proof. □

There is also an inverse statement to the contraction principle, where one has to assume a more restrictive setting.

Theorem 3.1.7. *(Inverse contraction principle, Theorem 4.2.4 in [9])*
Let $\mathcal{X}$ and $\mathcal{Y}$ be two Polish spaces. Suppose that $g : \mathcal{Y} \to \mathcal{X}$ is a continuous bijection, and that $\{\nu_\epsilon\}_{\epsilon>0}$ is an exponentially tight family of probability measures on $\mathcal{Y}$. If $\{\nu_\epsilon \otimes g^{-1}\}_{\epsilon>0}$ satisfies the LDP with the GRF $\mathbb{I} : \mathcal{X} \to [0, \infty]$, then $\{\nu_\epsilon\}_{\epsilon>0}$ satisfies the LDP with GRF $\mathbb{I}' = \mathbb{I} \circ g$.

The proof is elementary but a little bit lengthy, and so we shall only present it in the appendix (see 6). We want to deduce a corollary, which we will later use in the main part of this thesis.

Corollary. Suppose two Polish spaces $(\mathcal{X}, \tau_1)$ and $(\mathcal{X}, \tau_2)$ with $\tau_2 \subseteq \tau_1$. Let $\{\mu_\epsilon\}_{\epsilon>0}$ be an exponentially tight family of probability measures on $(\mathcal{X}, \tau_1)$. If $\{\mu_\epsilon\}_{\epsilon>0}$ satisfies an LDP with respect to the topology τ_2 that is coarser than τ_1, then the same LDP holds with respect to the topology τ_1.

Proof. The proof follows directly from the inverse contraction principle using the identity mapping $g : (\mathcal{X}, \tau_1) \to (\mathcal{X}, \tau_2), x \mapsto x$, which is continuous because $\tau_2 \subseteq \tau_1$. □

Remark. In [9] Theorem 3.1.7 and Corollary 3.1.2 are stated in a more general setting, but we will only need those results in the case of Polish spaces.

In many situations it is possible to show that a family of laws $\{\mu_\epsilon\}_{\epsilon>0}$ satisfies an LDP by showing that $\{\mu_\epsilon\}_{\epsilon>0}$ is „equivalent“ to $\{\tilde{\mu}_\epsilon\}_{\epsilon>0}$, where the latter is known to satisfy a certain LDP. This leads to the notion of exponential equivalence.

Definition 3.1.7. *Let $\mathcal{Y}$ be a Polish space with metric d. The probability measures $\{\mu_\epsilon\}_{\epsilon>0}$ and $\{\tilde{\mu}_\epsilon\}_{\epsilon>0}$ are called exponentially equivalent if there exist probability spaces $\{(\Omega, \mathcal{B}_\epsilon, P_\epsilon)\}_{\epsilon>0}$ and two families of $\mathcal{Y}$-valued random variables $\{Z_\epsilon\}_{\epsilon>0}$ and $\{\tilde{Z}_\epsilon\}_{\epsilon>0}$ with joint laws $\{P_\epsilon\}_{\epsilon>0}$ and marginals $\{\mu_\epsilon\}_{\epsilon>0}$ and $\{\tilde{\mu}_\epsilon\}_{\epsilon>0}$, respectively, such that the following condition is satisfied: For each $\delta > 0$, the set $\{\omega \in \Omega \mid (Z_\epsilon, \tilde{Z}_\epsilon) \in \Gamma_\delta\}$ is $\mathcal{B}_\epsilon$ measurable, and*

$$\limsup_{\epsilon \to 0} \epsilon \log P_\epsilon(\Gamma_\delta) = -\infty,$$

where

$$\Gamma_\delta = \{(y,\tilde{y}) : d(y,\tilde{y}) > \delta\}.$$

Remark. In [9] this definition is stated in a more general setting, here one could omit the measurability assumption.

Remark. If we work with random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in a Polish space $(\mathcal{X}, d)$, then the previous definition can be simplified to

$$\limsup_{\epsilon\to 0} \epsilon \log \mathbb{P}\big[d(Z_\epsilon, \tilde{Z}_\epsilon) > \delta\big] = -\infty$$

for all $\delta > 0$.

Theorem 3.1.8. *Consider two families of random variables $\{Z_\epsilon\}_{\epsilon>0}$ and $\{\tilde{Z}_\epsilon\}_{\epsilon>0}$, defined on a common probability space mapping into a Polish space $(\mathcal{X}, d)$. If $\{Z_\epsilon\}_{\epsilon>0}$ and $\{\tilde{Z}_\epsilon\}_{\epsilon>0}$ are exponentially equivalent, one of them obeys the LDP with a GRF $\mathbb{I}$ iff the other one does as well.*

Proof. We are providing a direct proof here, in [9] a more general theory is introduced, Theorem 3.1.8 is then a special case.
Assume that $\{Z_\epsilon\}_{\epsilon>0}$ satisfies an LDP with GRF $\mathbb{I}$ and take a closed set C. For $\delta > 0$ we consider the δ neighbourhood C_δ of C, i.e. $C_\delta = \{x \in \mathcal{X} : \exists y \in C : d(x,y) \leq \delta\}$. Then we get the following upper bound,

$$\begin{aligned}\limsup_{\epsilon\to 0} \epsilon \log \mathbb{P}[\tilde{Z}_\epsilon \in C] &= \limsup_{\epsilon\to 0} \epsilon \log \big(\mathbb{P}[\tilde{Z}_\epsilon \in C, d(\tilde{Z}_\epsilon, Z_\epsilon) > \delta] + \mathbb{P}[\tilde{Z}_\epsilon \in C, d(\tilde{Z}_\epsilon, Z_\epsilon) \leq \delta]\big) \\ &\leq \limsup_{\epsilon\to 0} \epsilon \log \big(\mathbb{P}[d(\tilde{Z}_\epsilon, Z_\epsilon) > \delta] + \mathbb{P}[Z_\epsilon \in C_\delta]\big),\end{aligned}$$

where we used that if $d(\tilde{Z}_\epsilon, Z_\epsilon) \leq \delta$, then $[\tilde{Z}_\epsilon \in C] \subseteq [Z_\epsilon \in C_\delta]$. For $a, b > 0$ we note the elementary inequality

$$\log(a+b) \leq \log(2(a \vee b)) = \log 2 + \log a \vee \log b.$$

In the $\limsup$ the $\log 2$ term vanishes, hence

$$\begin{aligned}&\limsup_{\epsilon\to 0} \epsilon \log \big(\mathbb{P}[d(\tilde{Z}_\epsilon, Z_\epsilon) > \delta] + \mathbb{P}[Z_\epsilon \in C_\delta]\big) \\ \leq &\limsup_{\epsilon\to 0} \epsilon \big(\log \mathbb{P}[d(\tilde{Z}_\epsilon, Z_\epsilon) > \delta] \vee \log \mathbb{P}[Z_\epsilon \in C_\delta]\big) \\ = &\limsup_{\epsilon\to 0} \epsilon \big(\log \mathbb{P}[d(\tilde{Z}_\epsilon, Z_\epsilon) > \delta] \vee \log \mathbb{P}[Z_\epsilon \in C_\delta]\big)\end{aligned}$$

$$
\begin{aligned}
&= \limsup_{\epsilon\to 0} \epsilon \log \mathbb{P}[Z_\epsilon \in C_\delta] \\
&\leq - \inf_{x\in C_\delta} \mathbb{I}(x).
\end{aligned}
$$

In the first equality we used the exponential equivalence of $\{\tilde{Z}_\epsilon\}_{\epsilon>0}$ and $\{Z_\epsilon\}_{\epsilon>0}$. If we infimize over all $\delta > 0$, the expression in the last line becomes $-\inf_{x\in C} \mathbb{I}(x)$, since $\mathbb{I}$ is a GRF.

The lower bound can be shown by a local argument, thus take $O \subseteq \mathcal{X}$ open and $x \in O$. Now we can find an open neighbourhood U around x and some $\delta > 0$, such that $U \subseteq U_\delta \subseteq O$. Then

$$
\begin{aligned}
\mathbb{P}[Z_\epsilon \in U] &= \mathbb{P}[Z_\epsilon \in U, d(\tilde{Z}_\epsilon, Z_\epsilon) > \delta] + \mathbb{P}[Z_\epsilon \in U, d(\tilde{Z}_\epsilon, Z_\epsilon) \leq \delta] \\
&\leq \mathbb{P}[d(\tilde{Z}_\epsilon, Z_\epsilon) > \delta] + \mathbb{P}[\tilde{Z}_\epsilon \in U_\delta] \\
&\leq \mathbb{P}[d(\tilde{Z}_\epsilon, Z_\epsilon) > \delta] + \mathbb{P}[\tilde{Z}_\epsilon \in O].
\end{aligned}
$$

Similar to the upper bound, we have

$$
\begin{aligned}
-\mathbb{I}(x) &\leq - \inf_{y\in U} \mathbb{I}(y) \\
&\leq \liminf_{\epsilon\to 0} \epsilon \log \mathbb{P}[Z_\epsilon \in U] \\
&\leq \liminf_{\epsilon\to 0} \epsilon \big(\log \mathbb{P}[d(\tilde{Z}_\epsilon, Z_\epsilon) > \delta] \vee \log \mathbb{P}[\tilde{Z}_\epsilon \in O] \big) \\
&= \liminf_{\epsilon\to 0} \epsilon \log \mathbb{P}[\tilde{Z}_\epsilon \in O].
\end{aligned}
$$

Now we can take the supremum over all $x \in O$ on the left hand side to receive the lower bound. □

We now restrict on a discrete setting. Consider a sequence of probability measures $\{\mu_n\}_{n\in\mathbb{N}} \subseteq P(\mathcal{X})$, where $(\mathcal{X}, d)$ is a Polish space. In this case we are also interested in the speed of the LDP, i.e. we want to find a monotone increasing sequence $\{s_n\}_{n\in\mathbb{N}}$ with $s_n \to \infty$ and a rate function $\mathbb{I} : \mathcal{X} \to [0, \infty]$ such that

$$
- \inf_{x\in\mathring{\Gamma}} \mathbb{I}(x) \leq \liminf_{n\to\infty} \frac{1}{s_n} \log \mu_n(\Gamma) \leq \limsup_{n\to\infty} \frac{1}{s_n} \log \mu_n(\Gamma) \leq - \inf_{x\in\overline{\Gamma}} \mathbb{I}(x),
$$

for a Borel set $\Gamma \subseteq \mathcal{X}$.

Remark. If $\mathbb{I}$ is a GRF and has a unique minimizer $m \in \mathcal{X}$ and we are given a sequence $\{s'_n\}_{n\in\mathbb{N}}$ with $s'_n \ll s_n$ and $s'_n \to \infty$. Then the LDP holds also with speed s'_n, in this case with GRF $\chi_m : \mathcal{X} \to [0, \infty]$, where

$$\chi_m(x) = \begin{cases} 0, & \text{if x=m} \\ \infty & \text{else.} \end{cases}$$

Proof. First, we observe that $\mathbb{I}(m) = 0$, since the LDP inequality yields (note that the whole space $\mathcal{X}$ is closed and open and $\mu_n(\mathcal{X}) = 1$, for all $n \in \mathbb{N}$)

$$-\inf_{x\in\mathcal{X}} \mathbb{I}(x) \le \liminf_{n\to\infty} \frac{1}{s_n} \log \mu_n(\mathcal{X}) = 0 = \limsup_{n\to\infty} \frac{1}{s_n} \log \mu_n(\mathcal{X}) \le -\inf_{x\in\mathcal{X}} \mathbb{I}(x).$$

Hence, $\mathbb{I}(m) = 0$. We first show the large deviation upper bound. Consider a closed $C \subseteq \mathcal{X}$ and $M > 0$. Then, if $m \in C$ (note that $\log \mu_n(C) \le 0, \forall n \in \mathbb{N}$)

$$\limsup_{n\to\infty} \frac{1}{s'_n} \log \mu_n(C) \le \limsup_{n\to\infty} \frac{1}{s_n} \log \mu_n(C) \le -\inf_{x\in C} \mathbb{I}(x) = 0 = -\inf_{x\in C} \chi_m(x).$$

In the case where $m \notin C$, we use that s_n/s'_n exceeds every given $M > 0$ if n is sufficiently large and that a GRF takes its infimum over any closed set, i.e.

$$\limsup_{n\to\infty} \frac{1}{s'_n} \log \mu_n(C) \le \limsup_{n\to\infty} \frac{s_n}{s'_n} \frac{1}{s_n} \log \mu_n(C) \le -M \inf_{x\in C} \mathbb{I}(x) \overset{M\to\infty}{\longrightarrow} -\infty = -\inf_{x\in C} \chi_m(x).$$

For the lower bound, we fix some open set $O \subseteq \mathcal{X}$. If $m \notin O$, then $-\inf_{x\in O} \chi_m(x) = -\infty$ and trivially

$$\liminf_{n\to\infty} \frac{1}{s'_n} \log \mu_n(O) \ge -\infty = -\inf_{x\in O} \chi_m(x).$$

If $m \in O$, then by the upper bound (note that O^c is closed)

$$\limsup_{n\to\infty} \frac{1}{s'_n} \log \mu_n(O^c) \le -\inf_{x\in O^c} \chi_m(x) = -\infty.$$

Which is only possible if $\mu_n(O^c) \overset{n\to\infty}{\longrightarrow} 0$, hence $\mu_n(O) \overset{n\to\infty}{\longrightarrow} 1$. Then we get

$$\liminf_{n\to\infty} \frac{1}{s'_n} \log \mu_n(O) = 0 = -\inf_{x\in O} \chi_m(x).$$

□

A weakness of the general contraction principle might be that the representation of the rate function via an infimum is not closed. We therefore provide a result in a discrete setting, first introduced in [5, Section 6.2], in [18] the authors called it „approximate contraction principle".

We start by introducing the formal background. Let $\mathcal{X}$ be a Polish space and $(\mathfrak{X}, \|.\|)$ be a separable Banach space (i.e, a complete normed vector space over $\mathbb{R}$ with a countable dense subset). Further, let $\mathfrak{X}^*$ be the topological dual, i.e. the space of continuous, linear functionals mapping from $\mathfrak{X}$ to $\mathbb{R}$. We denote by

$$\begin{aligned}\langle .,. \rangle : \mathfrak{X}^* \times \mathfrak{X} &\to \mathbb{R} \\ (x^*, x) &\mapsto x^*(x)\end{aligned}$$

the dual pairing. Let $P(\mathcal{X})$ be the set of probability measures over $\mathcal{X}$ and fix a continuous mapping $c : \mathcal{X} \to \mathfrak{X}$. Further we consider a sequence $\{\mathscr{L}_n\}_{n\in\mathbb{N}}$ of $P(\mathcal{X})$ valued random variables and define $\mathfrak{X}$ valued random variables $\{\mathscr{C}_n\}_{n\in\mathbb{N}}$ with

$$\mathscr{C}_n := \int_{\mathcal{X}} c(x) d\mathscr{L}_n(x) \in \mathfrak{X},$$

assuming the latter exists as pointwise Bochner integral (see Definition 6.0.8 and the following in the Appendix). Let the sequence of vectors $\{(\mathscr{C}_n, \mathscr{L}_n)\}_{n\in\mathbb{N}}$ satisfy an LDP with some speed s_n and GRF $\mathbb{I} : \mathfrak{X} \times P(\mathcal{X}) \to [0, \infty]$.

We are interested in a representation of $\mathbb{I}$, for this urge we provide a motivation for bounded c. In this case the existence of $\mathscr{C}_n$ needs no further assumption and also the mapping $\mu \mapsto \int_{\mathcal{X}} c(x) d\mu(x)$ is well defined and continuous for $\mu \in P(\mathcal{X})$. The contraction principle provides an LDP for $\{\mathscr{L}_n\}_{n\in\mathbb{N}}$ with speed s_n and GRF $\mathbb{I}_0 : P(\mathcal{X}) \to [0, \infty]$

$$\mathbb{I}_0(\mu) = \inf_{x\in\mathfrak{X}} \mathbb{I}(x, \mu).$$

Applying the contraction principle once more for the continuous map

$$\begin{aligned}F : P(\mathcal{X}) &\to \mathfrak{X} \times P(\mathcal{X}) \\ \mu &\mapsto \Big(\int_{\mathcal{X}} c(x) d\mu(x), \mu \Big)\end{aligned}$$

yields a new representation for $\mathbb{I}$. We note that the preimage of $(\mathscr{C}, \mu) \in \mathfrak{X} \times P(\mathcal{X})$ under F is

$$F^{-1}(\{\mathscr{C}, \mu\}) = \begin{cases} \mu, & \text{if } \mathscr{C} = \int_{\mathcal{X}} c(x) d\mu(x) \\ \emptyset & \text{else.} \end{cases}$$

Hence,

$$\mathbb{I}(\mathscr{C},\mu) = \inf_{\nu\in F^{-1}(\{\mathscr{C},\mu\})} \mathbb{I}_0(\nu) = \begin{cases} \mathbb{I}_0(\mu) + \chi_0\Big(\mathscr{C} - \int_{\mathcal{X}} c(x)d\mu(x)\Big) & \text{, if } \mathbb{I}_0(\mu) < \infty \\ \infty & \text{else,} \end{cases} \tag{3.1.4}$$

where $\chi_0(0) = 0$ and $\chi_0(x) = \infty$ for $x \neq 0$. We want to receive a similar representation for $\mathbb{I}$ under additional assumptions in case of unbounded c.
Let us start with some definitions and notation. Let $r \in (0,\infty]$ and $W : \mathcal{X} \to \mathbb{R}$ be a continuous function. Then we consider the Varadhan type limit

$$\Lambda_r(W) := \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\Big[e^{n\big(\int_{\mathcal{X}} W(x)d\mathscr{L}_n(x)\wedge r\big)}\Big] \qquad \text{and} \qquad \bar{\Lambda}(W) := \sup_{r>0} \Lambda_r(W),$$

where $\wedge$ refers to the minimum of the two quantities. Further we will work with the „domain" of $\bar{\Lambda}$ and its „interior", as well as a certain functional. More precisely, for fixed continuous $c : \mathcal{X} \to \mathfrak{X}$, we define

$$\begin{aligned} \mathscr{D} &:= \{\alpha \in \mathfrak{X}^* \mid \bar{\Lambda}(\langle\alpha, c(\cdot)\rangle) < \infty\} \\ \mathscr{D}_o &:= \{\alpha \in \mathfrak{X}^* \mid \exists p > 1 : p\alpha \in \mathscr{D}\} \\ F(x) &:= \sup_{\alpha\in\mathscr{D}_o} \langle\alpha, x\rangle, \ \ x \in \mathfrak{X}. \end{aligned}$$

Now we can state the central result in this subsection, allowing to establish a similar representation of $\mathbb{I}$ as in Equation 3.1.4, but for more general c.

Theorem 3.1.9. *(Proposition 6.4 in [5])*
Let $(\mathcal{X}, d)$ be a Polish space, $(\mathfrak{X}, ||\cdot||)$ be a separable Banach space, $c : \mathcal{X} \to \mathfrak{X}$ a continuous function, $\{\mathscr{L}_n\}_{n\in\mathbb{N}}$ a sequence of $P(\mathcal{X})$ valued random variables and suppose the following assumptions:

1. $\exists \eta > 0 : \Lambda_\infty(\eta||c(\cdot)||) < \infty$.

2. *$\{(\mathscr{C}_n, \mathscr{L}_n)\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathfrak{X} \times P(\mathcal{X})$ with convex GRF $\mathbb{I}$.*

3. *For every sequence $\{W_n\}_{n\in\mathbb{N}}$ with $W_n \in \{V + \langle\alpha, c(\cdot)\rangle \mid V : \mathcal{X} \to \mathbb{R}$ continuous and bounded, $\alpha \in \mathscr{D}_o\}$ and $W_n \downarrow W_\infty$ to a limit $W_\infty : \mathcal{X} \to \mathbb{R}$ that is continuous and bounded from above, we have*
$$\limsup_{n\to\infty} \bar{\Lambda}(W_n) \leq \bar{\Lambda}(W_\infty). \tag{3.1.5}$$

Then $\mathbb{I}(\mathscr{C},\mu)$ satisfies identity 3.1.4 with F instead of χ_0 and $\mathbb{I}_0$ the GRF for the LDP of $\{\mathscr{L}_n\}_{n\in\mathbb{N}}$ in $P(\mathcal{X})$. In particular, $\mathbb{I}(\mathscr{C},\mu) = 0$ iff $\mathscr{C} = \int_{\mathcal{X}} c(x)d\mu(x)$ and $\mathbb{I}_0(\mu) = 0$.

Remark. The assumption $\Lambda_\infty(\eta\|c(\cdot)\|) < \infty$ implies the existence of the Bochner integrals $\mathscr{C}_n = \int_{\mathcal{X}} c(x)d\mathscr{L}_n(x)$, $n \in \mathbb{N}$. Hence, Assumption 1 from the theorem allows to consider the sequence $\{(\mathscr{C}_n, \mathscr{L}_n)\}_{n\in\mathbb{N}}$.

Proof. If $\Lambda_\infty(\eta\|c(\cdot)\|) < \infty$, then

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[e^{n\left(\int_{\mathcal{X}} \eta\|c(x)\|d\mathscr{L}_n(x)\right)}\right] < \infty,$$

and hence for any $n \in \mathbb{N}$,

$$\mathbb{E}\left[e^{n\left(\int_{\mathcal{X}} \eta\|c(x)\|d\mathscr{L}_n(x)\right)}\right] < \infty.$$

In particular $\mathbb{E}\big(\int_{\mathcal{X}} \eta\|c(x)\|d\mathscr{L}_n(x)\big) < \infty$, thus directly implying Bochner-integrability. □

Remark. In [5, Section 6.2], there are several interesting observations together with useful identities including the GRFs $\mathbb{I}_0$ and $\mathbb{I}$, which we will need in the proof of Theorem 3.1.9. Let $C_b(\mathcal{X})$ denote the set of continuous and bounded functions mapping from $\mathcal{X}$ to $\mathbb{R}$. Then under the assumptions of Theorem 3.1.9

$$\mathbb{I}(\mathscr{C},\mu) = \sup_{\alpha\in\mathscr{D}_o} \sup_{V\in C_b(\mathcal{X})} \left[\langle\alpha,\mathscr{C}\rangle + \int_{\mathcal{X}} V(x)d\mu(x) - \bar{\Lambda}(V + \langle\alpha, c(\cdot)\rangle)\right], \tag{3.1.6}$$

for $\mathscr{C} \in \mathfrak{X}$, $\mu \in P(\mathcal{X})$ and

$$\mathbb{I}_0(\mu) = \sup_{V\in C_b(\mathcal{X})} \left[\int_{\mathcal{X}} V(x)d\mu(x) - \bar{\Lambda}(V)\right]. \tag{3.1.7}$$

Proof. (Theorem 3.1.9)
By Assumption 2, $\{(\mathscr{C}_n, \mathscr{L}_n)\}_{n\in\mathbb{N}}$ satisfies an LDP with GRF $\mathbb{I}$. As in the motivation, one can derive an LDP for $\{\mathscr{L}_n\}_{n\in\mathbb{N}}$ via the contraction principle with GRF $\mathbb{I}_0 : P(\mathcal{X}) \to [0,\infty]$

$$\mathbb{I}_0(\mu) = \inf_{x\in\mathfrak{X}} \mathbb{I}(x,\mu).$$

Thus $\mathbb{I}_0(\mu) \leq \mathbb{I}(\mathscr{C},\mu)$ for all $\mathscr{C} \in \mathfrak{X}$. Representation 3.1.4 hence holds in the case $\mathbb{I}_0(\mu) = \infty$. We can now treat the case $\mu \in P(\mathcal{X})$ with $\mathbb{I}_0(\mu) < \infty$.
Using the identity in 3.1.7 and $V^* := \eta\|c(\cdot)\| \wedge M \in C_b(\mathcal{X})$, for some $M > 0$ and $\eta > 0$

from Assumption 1, yields

$$\mathbb{I}_0(\mu) \geq \int_{\mathcal{X}} V^*(x)d\mu(x) - \bar{\Lambda}(V^*)$$

$$\begin{aligned}\iff \int_{\mathcal{X}} \eta\|c(x)\| \wedge M d\mu(x) &\leq \mathbb{I}_0(\mu) + \bar{\Lambda}(\eta\|c(\cdot)\| \wedge M) \\ &\leq \mathbb{I}_0(\mu) + \bar{\Lambda}(\eta\|c(\cdot)\|).\end{aligned}$$

In the last inequality we used that $\bar{\Lambda}$ is monotone increasing, i.e. for $W_1, W_2 \in C_b(\mathcal{X})$ with

$$W_1(x) \leq W_2(x), \forall x \in \mathcal{X}$$

we have $\bar{\Lambda}(W_1(\cdot)) \leq \bar{\Lambda}(W_2(\cdot))$. By monotone convergence for $M \to \infty$ we see that

$$\eta \int \|c(x)\| d\mu(x) \leq \mathbb{I}_0(\mu) + \bar{\Lambda}(\eta\|c(\cdot)\|).$$

Since $\bar{\Lambda} \leq \Lambda_\infty$, we can use the case assumption $\mathbb{I}_0(\mu) < \infty$ and Assumption 1 to deduce

$$\eta \int \|c(x)\| d\mu(x) < \infty.$$

Hence, c is Bochner-integrabel and for $\alpha \in \mathfrak{X}^*$ it follows that $\langle \alpha, c(\cdot)\rangle$ is in $L_1(\mathcal{X}, \mu)$ by Theorem 6.0.18. Moreover, we get the following identity

$$\int_{\mathcal{X}} \langle \alpha, c(x)\rangle d\mu(x) = \langle \alpha, \int_{\mathcal{X}} c(x) d\mu(x)\rangle.$$

In particular, this equality holds for $\alpha \in \mathcal{D}_o$. This can be used to rearrange representation 3.1.6 of $\mathbb{I}$, hence we get

$$\begin{aligned}\mathbb{I}(\mathscr{C}, \mu) &= \sup_{\alpha \in \mathcal{D}_o} \sup_{V \in C_b(\mathcal{X})} \left[\langle \alpha, \mathscr{C}\rangle + \int_{\mathcal{X}} V(x) d\mu(x) - \bar{\Lambda}(V + \langle \alpha, c(\cdot)\rangle) \right] \\ &= \sup_{\alpha \in \mathcal{D}_o} \sup_{V \in C_b(\mathcal{X})} \left[\langle \alpha, \mathscr{C} - \int_{\mathcal{X}} c(x) d\mu(x)\rangle + \int_{\mathcal{X}} V(x) + \langle \alpha, c(x)\rangle d\mu(x) - \bar{\Lambda}(V + \langle \alpha, c(\cdot)\rangle) \right] \\ &= \sup_{\alpha \in \mathcal{D}_o} \left[\langle \alpha, \mathscr{C} - \int_{\mathcal{X}} c(x) d\mu(x)\rangle + \sup_{V \in C_b(\mathcal{X})} \int_{\mathcal{X}} V(x) + \langle \alpha, c(x)\rangle d\mu(x) - \bar{\Lambda}(V + \langle \alpha, c(\cdot)\rangle) \right] \\ &= \sup_{\alpha \in \mathcal{D}_o} \left[\langle \alpha, \mathscr{C} - \int_{\mathcal{X}} c(x) d\mu(x)\rangle + \mathbb{I}_{\langle \alpha, c(\cdot)\rangle}(\mu) \right],\end{aligned}$$

where for a function $g : \mathcal{X} \to \mathbb{R}$ of the form $g = \langle \alpha, c(\cdot) \rangle$, for some $\alpha \in \mathscr{D}_o$, we use

$$\mathbb{I}_g(\mu) := \sup_{V \in C_b(\mathcal{X})} \int_{\mathcal{X}} V(x) + g(x) d\mu(x) - \bar{\Lambda}(V + g),$$

which is consistent with the definition of $\mathbb{I}_0$ in Equation 3.1.7. If one is able to show that $\mathbb{I}_0 = \mathbb{I}_g$, for all such g, then one has established identity 3.1.4. We shall prove this identity now.

1. $\mathbb{I}_0(\mu) \geq \mathbb{I}_g(\mu)$: Define the auxiliary function $\phi_{n,m} : \mathbb{R} \to \mathbb{R}$ with

$$\phi_{n,m}(x) := x \mathbb{1}_{(-m,n)}(x) + n \mathbb{1}_{[n,\infty)}(x) - m \mathbb{1}_{[-m,\infty)}(x), \qquad \text{for } n, m \in \mathbb{N}$$

and denote $\phi_{\infty,m}(x) := \lim_{n\to\infty} \phi_{n,m}(x)$ and $\phi_{n,\infty}(x) := \lim_{m\to\infty} \phi_{n,m}(x)$ for $x \in \mathbb{R}$. Then by construction $\phi_{n,m} \in C_b(\mathbb{R})$. For $V \in C_b(\mathcal{X})$ and $g = \langle \alpha, c(\cdot) \rangle$, for some $\alpha \in \mathscr{D}_o$, we have $\phi_{n,m}(V + g) \in C_b(\mathcal{X})$. Using Equation 3.1.7 yields

$$\mathbb{I}_0(\mu) \geq \int_{\mathcal{X}} \phi_{n,m}(V(x) + g(x)) d\mu(x) - \bar{\Lambda}(\phi_{n,m}(V + g)). \tag{3.1.8}$$

We have the pointwise limit $\phi_{n,m}(V + g) \downarrow \phi_{n,\infty}(V + g)$ as $m \to \infty$ and hence by Assumption 3

$$\limsup_{m\to\infty} \bar{\Lambda}(\phi_{m,n}(V + g)) \leq \bar{\Lambda}(\phi_{\infty,n}(V + g)) \leq \bar{\Lambda}(V + g),$$

where the second inequality holds due to the monotonicity of $\bar{\Lambda}$. We use this to get the lower bound

$$\begin{aligned} \mathbb{I}_0(\mu) &\geq \limsup_{m\to\infty} \int_{\mathcal{X}} \phi_{n,m}(V(x) + g(x)) d\mu(x) - \bar{\Lambda}(V + g) \\ &\geq \int_{\mathcal{X}} \phi_{n,\infty}(V(x) + g(x)) d\mu(x) - \bar{\Lambda}(V + g). \end{aligned}$$

Moreover, $\phi_{n,\infty}(x) \uparrow x$ for $x \in \mathbb{R}$ by construction. Since $g = \langle \alpha, c(\cdot) \rangle$ is Bochner-integrabel and V is bounded, we can apply dominated convergence to receive the inequality

$$\mathbb{I}_0(\mu) \geq \int_{\mathcal{X}} V(x) + g(x) d\mu(x) - \bar{\Lambda}(V + g).$$

Optimization over all $V \in C_b(\mathcal{X})$ yields the desired inequality.

2. $\mathbb{I}_0(\mu) \leq \mathbb{I}_g(\mu)$: We work again with the functions $\{\phi_{n,m}\}_{n,m\in\mathbb{N}}$. For $V \in C_b(\mathcal{X})$ and $g = \langle \alpha, c(\cdot) \rangle$, for some $\alpha \in \mathscr{D}_o$, we have that $V + \phi_{n,m}(g) \in C_b(\mathcal{X})$. Hence, by the definition

of $\mathbb{I}_g$, we have that

$$\mathbb{I}_g(\mu) \geq \int_{\mathcal{X}} (V(x) - g(x) + \phi_{n,m}(g(x)) d\mu(x)) - \bar{\Lambda}(V - g + \phi_{n,m}(g)). \tag{3.1.9}$$

Now we apply essentially the same argument as in the first direction of the proof. $\phi_{m,n}(g) \to g$ pointwise as $n, m \to \infty$, hence

$$\lim_{n\to\infty} \lim_{m\to\infty} \int_{\mathcal{X}} [\phi_{n,m}(g(x)) - g(x)] d\mu(x) = 0$$

by dominated convergence. Also we have $V - g + \phi_{n,m}(g) \downarrow V - g + \phi_{n,\infty}(g)$, so we can use Assumption 3 and the monotonicity of $\bar{\Lambda}$ to get

$$\limsup_{m\to\infty} \bar{\Lambda}(V - g + \phi_{n,m}(g)) \leq \bar{\Lambda}(V - g + \phi_{n,\infty}(g)) \leq \bar{\Lambda}(V).$$

After optimization over all V we showed $\mathbb{I}_g(\mu) \geq \mathbb{I}_0(\mu)$.
For the second claim, we note that $\mathbb{I}(\mathscr{C}, \mu) = 0$ if $\mathbb{I}_0(\mu) = 0$ and $\mathscr{C} = \int_{\mathcal{X}} c(x) d\mu(x)$. The other direction is not immediately clear, assume $\mathscr{C} \neq \int_{\mathcal{X}} c(x) d\mu(x)$. Then we find $\alpha \in \mathfrak{X}^*$, such that $\langle \alpha, \mathscr{C} - \int_{\mathcal{X}} c(x) d\mu(x) \rangle > 0$. Since

$$||\alpha||_\infty := \sup_{x\in\mathfrak{X}, ||x||\leq 1} |\alpha(x)| < \infty,$$

we choose $\epsilon > 0$ with $\epsilon ||\alpha||_\infty < \eta$. Then $\epsilon\alpha \in \mathscr{D}_0$, and we get $F(\mathscr{C} - \int_{\mathcal{X}} c(x) d\mu(x)) > 0$. $\square$

3.1.3 Classical results and some applications

In the following we study a few classical results from large deviation theory, beginning with a version of Cramér's theorem for real-valued random variables.
For a sequence of iid random variables $X_1, X_2, ..$ we introduce the quantities

- $S_n := X_1 + ... + X_n$
- $\Lambda(t) := \log \mathbb{E}[\exp(tX_1)]$
- $\mathcal{D}_\Lambda := \{t \in \mathbb{R} \mid \Lambda(t) < \infty\}$
- $\Lambda^*(x) := \sup_{t\in\mathbb{R}} (xt - \Lambda(t))$
- $\mu_n(\cdot) := \mathbb{P}[S_n/n \in \cdot]$.

Before we formulate Cramér's theorem, we state some properties of Λ and Λ^*.

Lemma 3.1.10. *(Lemma 2.2.5 in [9])*
Denote by $\bar{x} = \mathbb{E}[X_1]$. *Then*
(a) Λ *is a convex function and* Λ^* *is a convex rate function.*
(b) If $\mathcal{D}_\Lambda = \{0\}$, *then* Λ^* *is identically zero. If* $\Lambda(\lambda) < \infty$ *for some* $\lambda > 0$, *then* $\bar{x} < \infty$ *(possibly* $\bar{x} = -\infty$*), and for all* $x \geq \bar{x}$,

$$\Lambda^*(x) = \sup_{\lambda \geq 0} \big(\lambda x - \Lambda(\lambda)\big)$$

is a non-decreasing function for $x > \bar{x}$. *Similarly, if* $\Lambda(\lambda) < \infty$ *for some* $\lambda < 0$, *then* $\bar{x} > -\infty$ *(possibly* $\bar{x} = \infty$*), and for all* $x \leq \bar{x}$,

$$\Lambda^*(x) = \sup_{\lambda \leq 0} \big(\lambda x - \Lambda(\lambda)\big)$$

is a non-increasing function for $x < \bar{x}$. *When* $\bar{x}$ *is finite,* $\Lambda^*(\bar{x}) = 0$, *and always,*

$$\inf_{x \in \mathbb{R}} \Lambda^*(x) = 0.$$

(c) Λ *is differentiable in* $\mathring{\mathcal{D}}_\Lambda$ *with*

$$\Lambda'(\eta) = \frac{1}{\mathbb{E}[e^{\eta X_1}]} \mathbb{E}\big[X_1 e^{\eta X_1}\big]$$

and

$$\Lambda'(\eta) = y \implies \Lambda^*(y) = \eta y - \Lambda(\eta).$$

Theorem 3.1.11. *(Theorem 2.2.3 in [9])*
Let $X_1, X_2, \ldots$ *be iid real-valued random variables. Then* $\{\mu_n\}_{n \in \mathbb{N}}$ *satisfies an LDP with the convex rate function* Λ^*, *namely:*
a) For any closed set $F \subseteq \mathbb{R}$

$$\limsup_{n \to \infty} \frac{1}{n} \log \mu_n(F) \leq -\inf_{x \in F} \Lambda^*(x).$$

b) For any open set $G \subseteq \mathbb{R}$

$$-\inf_{x \in G} \Lambda^*(x) \leq \liminf_{n \to \infty} \frac{1}{n} \log \mu_n(G).$$

Application 1. Consider a sequence $\zeta_1, \zeta_2, \ldots$ of iid standard normal distributed random variables, where we have $\Lambda(t) = t^2/2$. The computation of the Legendre transform is straight forward and gives us

$$\Lambda^*(x) = \frac{x^2}{2}.$$

For each measurable set A we thus have

$$-\inf_{x\in\overline{A}} \frac{x^2}{2} \leq \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(A) \leq \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(A) \leq -\inf_{x\in\mathring{A}} \frac{x^2}{2}.$$

Application 2. Again we consider a sequence $\zeta_1, \zeta_2, \ldots$ of iid standard normal distributed random variables. In this example we study the asymptotic behaviour of

$$\frac{1}{n}\sum_{j=1}^{n} |\zeta_j|.$$

The computation of Λ takes a little bit more effort, but is still elementary, we get

$$\Lambda(t) = \frac{t^2}{2} + \log \Phi(t) + \log 2,$$

where $\Phi(t)$ denotes the standard normal distribution function. There is no closed representation of Λ^*, but we can still provide a few properties using Lemma 3.1.10. Λ^* is convex and infinity over the negative real numbers. For every $t \in \mathbb{R}$ we have $\Lambda(t) < \infty$, thus for $x > \bar{x} = \sqrt{2/\pi}$

$$\Lambda^*(x) = \sup_{t\geq 0} \Big(xt - \frac{t^2}{2} - \log \Phi(t) - \log 2 \Big).$$

Moreover, solving the equation $\Lambda'(\eta) = y$ leads to the relation

$$y = \eta + \frac{\Phi'(\eta)}{\Phi(\eta)}.$$

We see that if $y \to \infty$, then also $\eta \to \infty$ and using properties of Φ implies $\eta = \Theta(y)$. Using

$$\Lambda^*(y) = \eta y - \Lambda(\eta)$$

shows

$$\Lambda^*(y) = \frac{1}{2}\Theta(y)^2 - \log \Theta(y).$$

Hence, Λ^* tends superlinearly to infinity as $y \to \infty$.

Now we introduce the concept of empirical measures and the general theorem of

Sanov. The following is based on [9, Chapter 6].
We consider a Polish vector space Σ and an iid sequence $Y_1, Y_2, ...$ of Σ-valued random variables. Assume that Y_1 is distributed according to $\mu \in P(\Sigma)$, where we recall that $P(\Sigma)$ denotes the space of Borel probability measures on Σ. We are interested in the asymptotic behaviour of the empirical law of $Y_1, Y_2, ..., Y_n$,i.e.

$$L_n^{\mathbf{Y}} = \frac{1}{n}\sum_{k=1}^{n} \delta_{Y_k}, \qquad n \in \mathbb{N}, \tag{3.1.10}$$

where δ_y denotes the Dirac-measure at $y \in \Sigma$. $L_n^{\mathbf{Y}}$ is, as convex combination of probability measures, again a probability measure. The philosophy is to understand $L_n^{\mathbf{Y}}$ as sum of iid $M(\Sigma)$ (space of finite signed measures on Σ) valued random variables. We can then apply the general theorem of Cramér for certain topological vector spaces (see Theorem 3.1.12). The technical difficulty consists in equipping $M(\Sigma)$ with the required topology. We will now leave the safe environment of Polish spaces for a second to state a version of Cramér's theorem in the necessary generality.

Definition 3.1.8. *(locally convex Hausdorff space)*
Let $(\mathcal{X}, \tau)$ be a topological vector space over the real numbers.
1. $(\mathcal{X}, \tau)$ *is called Hausdorff space iff*

$$\forall x, y \in \mathcal{X} : x \neq y : \exists A, B \in \tau : x \in A, y \in B \text{ and } A \cap B = \emptyset.$$

2. $(\mathcal{X}, \tau)$ is called locally convex, iff for every neighbourhood U around 0, there exists an open set $T \subseteq U$ such that

a) T is convex,

b) For all $x \in T : \exists r > 0 : \forall |\alpha| < r : \alpha x \in T$,

c) For all $x \in T : \forall |r| \leq 1 : rx \in T$.

Assumption. a) $\mathcal{X}$ is a locally convex, Hausdorff, topological real vector space. $\mathcal{E}$ is a closed, convex subset of $\mathcal{X}$ such that $\nu(\mathcal{E}) = 1$ and $\mathcal{E}$ can be made into a Polish space with respect to the topology induced by $\mathcal{X}$.
b) The closed convex hull of each $K \subseteq \mathcal{E}$ is compact.

Let ν be a Borel probability measure on such a vector space $\mathcal{X}$. On the space $\mathcal{X}^*$ of continuous linear functionals on $\mathcal{X}$, define the logarithmic moment generating function $\Lambda : \mathcal{X}^* \to (-\infty, \infty]$

$$\lambda \mapsto \log \int_{\mathcal{X}} e^{\langle \lambda, x \rangle} d\nu(x)$$

and let $\Lambda^* : \mathcal{X} \to [0,\infty]$ denote the Legendre transform of Λ, which is, in this setting

$$x \mapsto \sup_{\lambda \in \mathcal{X}^*} \left\{ \langle \lambda, x \rangle - \Lambda(\lambda) \right\}.$$

We now want to consider the empirical mean S_n of iid $X_1, ..., X_n$ $\mathcal{X}$-valued random variables, i.e.

$$S_n := \frac{1}{n} \sum_{i=1}^{n} X_i, \qquad n \in \mathbb{N}, \tag{3.1.11}$$

where each X_i is distributed according to ν. There are some associated measurability issues discussed on p. 252 of [9], we omit the technical details here. Denote by ν_n the distribution of S_n.

Theorem 3.1.12. *(Generalized Cramér, Theorem 6.1.3. in [9])*
Let the previous Assumptions a) and b) hold. Then $\{\nu_n\}_{n\in\mathbb{N}}$ satisfies in $\mathcal{X}$ (and $\mathcal{E}$) a weak LDP with rate function Λ^*. Moreover, for every open, convex subset $A \subseteq \mathcal{X}$,

$$\lim_{n\to\infty} \frac{1}{n} \log \nu_n(A) = -\inf_{x\in A} \Lambda^*(x).$$

Remark. (Corollary 6.1.6 in [9])
Assume $\mathcal{X} = \mathcal{E} = \mathbb{R}^d$ and define ν_n as the distribution of empirical mean of n iid $\mathcal{X}$-valued random variables. Then the sequence $\{\nu_n\}_{n\in\mathbb{N}}$ satisfies a weak LDP with the convex rate function Λ^*. Moreover, if $0 \in \mathcal{D}^o_\Lambda$, then $\{\nu_n\}_{n\in\mathbb{N}}$ satisfies the full LDP with the good, convex rate function Λ^*.

We now come back to our initial problem, where we want to apply Theorem 3.1.12 on the sequence of empirical measures $\{L_n^{\mathbf{Y}}\}_{n\in\mathbb{N}}$ from 3.1.10. After that, our program is as follows:

1. We introduce the weak topology on $M(\Sigma)$ and the induced relative topology on $P(\Sigma)$, which fulfil our assumptions.

2. By applying the generalized Cramér's theorem we get the weak LDP.

3. Exponential tightness leads to the full LDP with GRF Λ^*.

4. Calculate Λ^*.

Definition 3.1.9. *(weak topology)*
The „weak" topology τ^ω on $M(\Sigma)$ is the topology generated by the system of sets $\{U_{\phi,x,\delta} : \phi \in$

$C_b(\Sigma), x \in \mathbb{R}, \delta > 0\}$, *where*

$$U_{\phi,x,\delta} = \{\nu \in M(\Sigma) : |\langle \phi, \nu \rangle - x| < \delta\},$$

with $\langle \phi, \nu \rangle = \int_\Sigma \phi(t) d\nu(t)$.

Lemma 3.1.13. $(M(\Sigma), \tau^\omega)$ *is a locally convex, Hausdorff, topological real vector space.* $P(\Sigma)$ *together with the induced relative topology is a Polish space.*

Proof. $M(\Sigma)$ can be made into a real vector space if we endow it with pointwise summation. In view of Theorem 6.0.2, we identify a subspace $\mathcal{Y} \subseteq M(\Sigma)'$ with

$$\mathcal{Y} = \left\{\hat{\phi} : M(\Sigma) \to \mathbb{R} \mid \hat{\phi}(\nu) = \int_\Sigma \phi(t) d\nu(t), \phi \in C_b(\Sigma), \nu \in M(\Sigma)\right\}$$

and note that $\mathcal{Y}$ is separating. Hence, τ^ω (which is precisely the $\mathcal{Y}$-topology) makes $M(\Sigma)$ to a vector space with the desired properties, as well as $\mathcal{Y} = M(\Sigma)^*$. If we restrict on $P(\Sigma)$ together with the induced topology, we receive the well known weak topology. $P(\Sigma)$ then becomes a Polish space (see Remark 3.3). $\square$

We denote by $\mathcal{B}^w$ the Borel σ-algebra generated by the weak topology.
Now the weak LDP for empirical measures follows as an immediate consequence of Theorem 3.1.12. Here we have $\mathcal{X} = M(\Sigma)$ and $\mathcal{X}^* = M(\Sigma)^* = \mathcal{Y}$, where $\mathcal{Y}$ is the set from Lemma 3.1.13. The sequence $\{X_i\}_{i \in \mathbb{N}}$ from Theorem 3.1.12 is the sequence of random Dirac-measures $\delta_{Y_1}, \delta_{Y_2}, \ldots$, where the sequence $\{Y_i\}_{i \in \mathbb{N}}$ is iid and distributed according to $\mu \in P(\Sigma)$. Moreover, the cumulant generating function is $\Lambda : M(\Sigma)^* \to (-\infty, \infty]$

$$\begin{aligned} \hat{\phi} \mapsto \log \mathbb{E}\big[\exp(\langle \hat{\phi}, \delta_{Y_1} \rangle)\big] &= \log \mathbb{E}\big[\exp(\phi(Y_1))\big] && (3.1.12) \\ &= \log \int_\Sigma \exp(\langle \phi, y \rangle) d\mu(y), && (3.1.13) \end{aligned}$$

where $\phi \in C_b(\Sigma)$ represents the functional $\hat{\phi} \in M(\Sigma)^*$. Hence, we use the (sloppy) notation $\Lambda(\phi)$ instead of $\Lambda(\hat{\phi})$. Now we can deduce the following corollary using Theorem 3.1.12.

Corollary. The sequence empirical measures $\{L_n^{\mathbf{Y}}\}_{n \in \mathbb{N}}$ satisfies a weak LDP in $P(\Sigma)$ with the convex rate function

$$\Lambda^*(\nu) = \sup_{\phi \in C_b(\Sigma)} \{\langle \phi, \nu \rangle - \Lambda(\phi)\},$$

for $\nu \in P(\Sigma)$.

Using the following lemma we can show that we have indeed a full LDP.

Lemma 3.1.14. *The laws of $L_n^{\mathbf{Y}}$ are exponentially tight.*

Proof. Using Theorem 6.0.10 implies: $\exists \Gamma_\ell \subseteq \Sigma$, such that

$$\mu(\Gamma_\ell^c) \leq e^{-2\ell}(e^\ell - 1)$$

for $\ell \in \mathbb{N}$. The set of measures

$$K^\ell = \{\nu : \nu(\Gamma_\ell) \geq 1 - 1/\ell\}$$

is closed by the Portmanteau Theorem 6.0.12. To see this assume $\{\nu_n\}_{n\in\mathbb{N}} \subseteq K^\ell$ with $\nu_n \to \nu$ weakly. Then

$$1 - 1/\ell \leq \limsup_{n\to\infty} \nu_n(\Gamma_\ell) \leq \nu(\Gamma_\ell)$$

since Γ_ℓ is compact, hence closed. Now define for $L \in \mathbb{N}$

$$K_L = \bigcap_{\ell=L}^{\infty} K^\ell.$$

K_L is tight (see Definition 6.0.7), since for $\nu \in K_L$ we have $\nu \in K^\ell, \forall \ell \geq L$, and hence for any $\ell \in \mathbb{N}$ there exists a compact Γ_ℓ such that $\nu(\Gamma_\ell) \geq 1 - 1/\ell$. Using Prohorov's theorem 6.0.11, we see that the closed set K_L is compact. Applying the exponential Chebyscheff inequality we get the following upper bound

$$\begin{aligned}
\mathbb{P}[L^{\mathbf{Y}_n} \notin K^\ell] &= \mathbb{P}[L^{\mathbf{Y}_n}(\Gamma_\ell) < 1 - 1/\ell] = \mathbb{P}[L^{\mathbf{Y}_n}(\Gamma_\ell^c) > 1/\ell] \\
&\leq \mathbb{E}\left[e^{2n\ell^2(L^{\mathbf{Y}_n}(\Gamma_\ell) - 1/\ell)}\right] \leq e^{-2n\ell}\mathbb{E}\left[e^{2n\ell^2 L^{\mathbf{Y}_n}(\Gamma_\ell)}\right] \\
&= e^{-2n\ell}\mathbb{E}\left[\exp(2\ell^2 \mathbb{1}_{\Gamma_\ell^c}(Y_1))\right]^n \\
&= e^{-2n\ell}\left(\mu(\Gamma_\ell) + e^{2\ell^2}\mu(\Gamma_\ell^c)\right)^n \leq e^{-n\ell}.
\end{aligned}$$

For K_L it follows

$$\mathbb{P}[L^{\mathbf{Y}_n} \notin K_L] \leq \sum_{\ell=L}^{\infty} \mathbb{P}[L^{\mathbf{Y}_n} \notin K^\ell] \leq \sum_{\ell=L}^{\infty} e^{-n\ell} \leq 2e^{-nL},$$

implying that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}[L^{\mathbf{Y}_n} \notin K_L] \leq -L.$$

□

In the last step we take a closer look at the Legendre transform $\Lambda^* : M(\Sigma) \to [0, \infty]$. In the following calculation we assume that the densities $d\mu/d\nu$ and $d\nu/d\mu = \big(d\mu/d\nu\big)^{-1}$ exist, where $\nu \in P(\Sigma)$ and $\mu \in P(\Sigma)$ is the distribution of Y_1.

$$\begin{aligned}
\Lambda^*(\nu) &= \sup_{\phi \in C_b(\Sigma)} \big\{ \langle \phi, \nu \rangle - \Lambda(\phi) \big\} \\
&= \sup_{\phi \in C_b(\Sigma)} \big\{ \langle \phi, \nu \rangle - \log \int_\Sigma \exp(\langle \phi, y \rangle) d\mu(y) \big\} \\
&= \sup_{\phi \in C_b(\Sigma)} \big\{ \langle \phi, \nu \rangle - \log \int_\Sigma \exp(\langle \phi, y \rangle) \frac{d\mu}{d\nu}(y) d\nu(y) \big\} \\
&\geq \sup_{\phi \in C_b(\Sigma)} \big\{ \langle \phi, \nu \rangle - \int_\Sigma \langle \phi, y \rangle + \log \Big(\frac{d\mu}{d\nu}(y) \Big) d\nu(y) \big\} \\
&= \sup_{\phi \in C_b(\Sigma)} \big\{ \langle \phi, \nu \rangle - \langle \phi, \nu \rangle + \int_\Sigma \log \Big(\frac{d\nu}{d\mu}(y) \Big) d\nu(y) \big\} \\
&= \int_\Sigma \frac{d\nu}{d\mu}(y) \log \Big(\frac{d\nu}{d\mu}(y) \Big) d\mu(y).
\end{aligned}$$

This calculation motivates the following definition.

Definition 3.1.10. *For a probability measure $\nu \in P(\Sigma)$ we define the relative entropy with respect to $\mu \in P(\Sigma)$ as*

$$H(\nu|\mu) := \begin{cases} \int_\Sigma f(x) \log f(x) d\mu(x) & \text{, if } f(x) = \frac{d\nu}{d\mu}(x) \text{ exists} \\ \infty & \text{,otherwise.} \end{cases}$$

$\frac{d\nu}{d\mu}$ denotes the Radon-Nikodym derivative of ν with respect to μ. The case when μ is the Lebesgue measure is treated separately and leads to a very similar notion, which we will just call the entropy. Assume a measure $\nu \in P(\Sigma)$ with density $d\nu/dx$. Then we define

$$h(\nu) := -\int_\Sigma \log \Big(\frac{d\nu}{dx}(x) \Big) d\nu(x).$$

We have seen that $\Lambda^*(\cdot) \geq H(\cdot|\mu)$. Indeed, we have equality here.

Lemma 3.1.15. *(Lemma 6.2.13 in [9])*
The identity $H(\cdot|\mu) = \Lambda^*$ holds over $P(\Sigma)$, where μ is the distribution of the Y_i's in 3.1.10 and Λ^* is the Legendre transform of Λ defined in 3.1.12

The previous lemma delivers the last step of our program and we are now able to deduce the following version of Sanov´s theorem.

Theorem 3.1.16. *(Sanov, Theorem 6.2.10 in [9])*
The empirical measures $\{L_n^Y\}_{n\in\mathbb{N}}$ satisfy the LDP in $P(\Sigma)$ equipped with the weak topology τ^ω with the good, convex rate function $H(\cdot|\mu)$.

Remark. In [9] Theorem 3.1.16 is stated in a more general version, where the LDP holds in the finer „τ-topology ". The latter is generated by sets $U_{\phi,x,\delta}$ with measurable ϕ instead of $\phi \in C_b(\Sigma)$.

We will now discuss some properties of the entropy, since H plays an important role in the main part of this thesis.

Application 3. Assume μ is the standard normal distribution. Then we have the relation

$$H(\nu|\mu) = -h(\nu) + \frac{1}{2}\log(2\pi) + \frac{m_2(\nu)}{2},$$

where $m_2(\nu)$ denotes the second moment of $\nu \in P(\Sigma)$ with density $\frac{d\nu}{dx}$.

Proof. First note that, due to the Radon-Nikodym theorem 6.0.7, $d\nu/d\mu$ exists as well as

$$\frac{dx}{d\mu} = \sqrt{2\pi}\exp\left(\frac{x^2}{2}\right).$$

Then we get

$$\begin{aligned} H(\nu|\mu) &= \int_\Sigma \log\left(\frac{d\nu}{d\mu}(x)\right)d\nu(x) \\ &= \int_\Sigma \log\left(\frac{d\nu}{dx}(x)\right)d\nu(x) + \int_\Sigma \log\left(\frac{dx}{d\mu}(x)\right)d\nu(x) \\ &= -h(\nu) + \frac{1}{2}\log(2\pi) + \frac{1}{2}\int_\Sigma x^2 d\nu(x) \\ &= -h(\nu) + \frac{1}{2}\log(2\pi) + \frac{m_2(\nu)}{2}. \end{aligned}$$

□

Lemma 3.1.17. *For $\mu, \nu \in P(\mathbb{R})$ we have $H(\mu|\nu) \geq 0$.*

Proof. We prove this statement if $d\mu/d\nu$ and $d\nu/d\mu = (d\mu/d\nu)^{-1}$ both exist. In this case the non-negativity of H is an application of Jensen's inequality. Indeed, we have

$$
\begin{aligned}
H(\mu|\nu) &= \int_{\mathbb{R}} \frac{d\mu}{d\nu}(x) \log\left(\frac{d\mu}{d\nu}(x)\right) d\nu(x) \\
&= \int_{\mathbb{R}} \log\left(\frac{d\mu}{d\nu}(x)\right) d\mu(x) \\
&= -\int_{\mathbb{R}} \log\left(\frac{d\nu}{d\mu}(x)\right) d\mu(x) \\
&\geq -\log\left(\int_{\mathbb{R}} \frac{d\nu}{d\mu}(x) d\mu(x)\right) \\
&= 0.
\end{aligned}
$$

□

Lemma 3.1.18. *We consider two distributions $\mu_1, \mu_2 \in P(\mathbb{R})$ and some constant $c > 0$. Then we have*

$$H(\mu_1(\cdot \times c)|\mu_2) = H(\mu_1|\mu_2(\cdot \times c^{-1})),$$

provided the respective Radon-Nikodym derivatives exist.

Proof. Denote by $\mu_1^c(\cdot) := \mu_1(\cdot \times c)$ and $\mu_2^{c^{-1}}(\cdot) := \mu_2(\cdot \times c^{-1})$ and consider

$$
\begin{aligned}
H(\mu_1^c|\mu_2) &= \int_{\mathbb{R}} \log\left(\frac{d\mu_1^c}{d\mu_2}(x)\right) d\mu_1^c(x) \\
&= \int_{\mathbb{R}} \log\left(\frac{d\mu_1^c}{d\mu_2}(c^{-1}x)\right) d\mu_1(x) \\
&\overset{!}{=} \int_{\mathbb{R}} \log\left(\frac{d\mu_1}{d\mu_2^{c^{-1}}}(x)\right) d\mu_1(x).
\end{aligned}
$$

The last equality can be proven by taking a closer look at $(d\mu_1^c/d\mu_2)(c^{-1}x)$. We use the defining property of a density, for a Borel-set $A \subseteq \mathbb{R}$ we have

$$
\begin{aligned}
\int_A \frac{d\mu_1^c}{d\mu_2}(c^{-1}x) d\mu_2^{c^{-1}}(x) &= \int_{c^{-1}A} \frac{d\mu_1^c}{d\mu_2}(x) d\mu_2(x) \\
&= \mu_1^c(c^{-1}A) \\
&= \mu_1(A).
\end{aligned}
$$

In the first equality we used the transformation rule for probability measures (see Theorem 6.0.8 in the Appendix). Hence, $(d\mu_1^c/d\mu_2)(c^{-1}x) = (d\mu_1/d\mu_2^{c^{-1}})(x)$, establishing the claimed equality. □

When working with the entropy h, one might be interested in finding a maximal "value" of h, given some constraints. The next result can be found in [7, Section 12]. Consider the following problem:
For some $m \in \mathbb{N}$, we are given functions $\{r_i\}_{i\in\{1,\dots,m\}}$ with $r_i : \mathbb{R} \to \mathbb{R}$ and real numbers $\{\mu_i\}_{i\in\{1,\dots,m\}}$. Then we want to find one density function f, which satisfies

$$\int_{\mathbb{R}} f(x) r_i(x) dx = \mu_i, \qquad \text{for } 1 \le i \le m \tag{3.1.14}$$

and maximizes h. Questions like this arise quite naturally when dealing with mathematical models containing entropy, for example in certain fields of Physics, but we will stay within pure mathematics. In order to shorten things a little bit, we are temporarily inconsistent with notation: we denote $h(f)$ for some density f instead of writing $h(\nu)$, where ν has the Lebesgue density f. Indeed the case where a given distribution ν has no density is not interesting, so in any case we only need to consider the situation when a density exists.

Theorem 3.1.19. *(Theorem 12.1.1 in [7])*
Assume there are constants $\lambda_0, \dots, \lambda_m \in \mathbb{R}$ such that $f_\lambda : \mathbb{R} \to \mathbb{R}$ with

$$f_\lambda(x) := e^{\lambda_0 + \sum_{j=1}^m \lambda_j r_j(x)} \tag{3.1.15}$$

fulfils the constraints 3.1.14. Then f_λ maximizes the entropy over all densities satisfying 3.1.14.

Proof. Let g be another density obeying 3.1.14. Then

$$\begin{aligned}
h(g) &= -\int_{\mathbb{R}} g(x) \log g(x) dx \\
&= -\int_{\mathbb{R}} g(x) \log\left(\frac{g(x)}{f_\lambda(x)} f_\lambda(x)\right) dx \\
&= -H(g|f_\lambda) - \int_{\mathbb{R}} g(x) \log f_\lambda(x) dx \\
&\le -\int_{\mathbb{R}} g(x) \log f_\lambda(x) dx \\
&= -\int_{\mathbb{R}} g(x)(\lambda_0 + \sum_{j=1}^m \lambda_j r_j(x)) dx \\
&\overset{3.1.14}{=} -\int_{\mathbb{R}} f_\lambda(x)(\lambda_0 + \sum_{j=1}^m \lambda_j r_j(x)) dx \\
&= h(f_\lambda).
\end{aligned}$$

□

Application 4. Assume we are interested in distributions with mean 0 and variance $\sigma^2 > 0$. This can be embedded in our setting via the two constraint functions $r_1(x) = x$ with $\mu_1 = 0$ and $r_2(x) = x^2$ with $\mu_2 = \sigma^2$. Then

$$f_\lambda(x) = \frac{1}{\sqrt{2\pi}} \exp\left(- \frac{x^2}{2\sigma^2} \right)$$

exists and corresponds to the $\mathcal{N}(0, \sigma^2)$ distribution. We can also compute the entropy evaluated at f_λ, namely

$$\begin{aligned} h(f_\lambda) &= - \int_{\mathbb{R}} f_\lambda(x) \log f_\lambda(x) dx \\ &= \frac{1}{2} \log(2\pi\sigma^2 e). \end{aligned}$$

This concludes our short excursion about the entropy function. The next classical result from large deviation theory is the useful theorem of Gärtner-Ellis, allowing to establish LDPs for non iid sequences.

Definition 3.1.11. *Consider a convex function $\Lambda : \mathbb{R}^k \to (-\infty, \infty]$ and define $\mathcal{D}_\Lambda := \{\lambda \in \mathbb{R}^k \mid \Lambda(\lambda) < \infty\}$. Λ is essentially smooth, if*

(a) $\mathring{\mathcal{D}}_\Lambda$ is non-empty.

(b) Λ is differentiable throughout $\mathring{\mathcal{D}}_\Lambda$.

(c) Λ is steep in that $\lim_{n\to\infty} ||\nabla\Lambda(\lambda_n)||_2 = \infty$ whenever $\{\lambda_n\}_{n\in\mathbb{N}}$ is a sequence in $\mathring{\mathcal{D}}_\Lambda$ converging to a boundary point of $\mathring{\mathcal{D}}_\Lambda$.

Theorem 3.1.20. *(Version of Gärtner-Ellis theorem, Theorem 2.3.6 in [9])*
Let $\{Z_n\}_{n\in\mathbb{N}}$ be a sequence of $\mathbb{R}^k$-valued random variables and for $\lambda \in \mathbb{R}^k$ define $\Lambda_n(\lambda) := \log \mathbb{E}[\exp(\langle \lambda, Z_n \rangle)]$. If for each $\lambda \in \mathbb{R}^k$ the limit

$$\Lambda(\lambda) = \lim_{n\to\infty} \frac{1}{n} \Lambda_n(n\lambda) \tag{3.1.16}$$

exists as an extended real number with $0 \in \mathring{\mathcal{D}}_\Lambda$ and if we further assume that Λ is essentially smooth and lower semi-continuous, then

(a) For each closed set F,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) \le - \inf_{x\in F} \Lambda^*(x).$$

(b) For each open set G

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) \geq -\inf_{x\in G} \Lambda^*(x).$$

In other words, an LDP holds with speed n and GRF Λ^*.

Remark. The theorem is still valid if we replace the speed n by any sequence $\{s_n\}_{n\in\mathbb{N}}$ with $s_n \to \infty$.

Application 5. Consider a k-dimensional standard normal distributed random vector $(\zeta_1, ..., \zeta_k)$ and some sequence $\{s_n\}_{n\in\mathbb{N}}$ with $s_n \to \infty$. We want to establish an LDP for the random variables

$$Z_n := \frac{(\zeta_1, ..., \zeta_k)}{\sqrt{s_n}}, \qquad n \in \mathbb{N}$$

by using the previous theorem. We therefore check the limit condition for the cumulant generating function,

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{s_n} \log\big(\mathbb{E}[\exp(s_n\langle\lambda, Z_n\rangle)]\big) &= \lim_{n\to\infty} \frac{1}{s_n} \log\big(\mathbb{E}[\exp(\sqrt{s_n}\langle(\zeta_1, ..., \zeta_k), \lambda\rangle)]\big) \\
&= \lim_{n\to\infty} \frac{1}{s_n} \log\Big(\exp\Big(\frac{s_n\|\lambda\|_2^2}{2}\Big)\Big) \\
&= \frac{\|\lambda\|_2^2}{2} \\
&=: \Lambda(\lambda).
\end{aligned}$$

Where in the second equality, we used that the moment generating function $\gamma : \mathbb{R}^k \to \mathbb{R}$ of a k-dimensional standard normal distributed random variable is

$$\gamma(\lambda) = \exp\Big(\frac{\|\lambda\|_2^2}{2}\Big).$$

Λ is obviously essentially smooth and invariant under the Legendre transform. The sequence $\{Z_n\}_{n\in\mathbb{N}}$ thus satisfies the LDP with GRF $\lambda \mapsto \|\lambda\|_2^2/2$ and speed s_n.

Application 6. Consider an iid sequence of standard normal distributed random variables $\zeta_1, \zeta_2, ...$ and define

$$Z_n := \frac{\|(\zeta_1, ..., \zeta_n)\|_2^2}{n}, \qquad n \in \mathbb{N}.$$

The limit condition for Gärtner-Ellis can be verified easily. For $t \in (-\infty, 1/2)$,

$$\lim_{n\to\infty} \frac{1}{n}\Lambda_n(nt) = \lim_{n\to\infty} \frac{1}{n} \log\left(\mathbb{E}[\exp(t\zeta_1^2)]^n\right) = -\frac{1}{2}\log(1-2t).$$

For the last step we used that ζ_1^2 is χ_1^2 distributed. For $t \geq 1/2$ the limit exists as extended real number and is equal to infinity. The essential smoothness follows directly, thus $\{Z_n\}_{n\in\mathbb{N}}$ satisfies the LDP with speed n and GRF

$$\Lambda^*(x) = \sup_{t\in(-\infty,1/2)} \left(xt + \frac{1}{2}\log(1-2t)\right) = \frac{1}{2}\left(x - \log x - 1\right)$$

for $x > 0$.

Application 7. Now consider the similar sequence

$$Z_n := \frac{\|(\zeta_1, \ldots, \zeta_{k_n})\|_2^2}{n}, \qquad n \in \mathbb{N},$$

where $\{k_n\}_{n\in\mathbb{N}}$ is a sequence with $k_n/n \to 0$, but $k_n \to \infty$. Then, for $t < 1/2$,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n}\Lambda_n(nt) &= \lim_{n\to\infty} \frac{1}{n} \log\left(\mathbb{E}[\exp(t\zeta_1^2)]^{k_n}\right) \\ &= 0.\end{aligned}$$

If $t \geq 1/2$, then the moment generation function of ζ_1^2 is infinity. The resulting Λ is essentially smooth and the Legendre transform can be calculated. Indeed, we get for $x > 0$

$$\Lambda^*(x) = \sup_{t\in\mathbb{R}} \left(xt - \Lambda(t)\right) = \frac{x}{2}.$$

Thus $\{Z_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ at speed n and GRF Λ^*.

3.2 Haar measure on the Stiefel manifold

Since we will consider projections of high dimensional random vectors on lower dimensional subspaces, we need to introduce something like a „uniform measure" on the set of k-dimensional subspaces of $\mathbb{R}^n$. This task can be realised via a short excursion to Haar measures on compact metric spaces (see Definition 6.0.4 in the Appendix), the desired concept will then be a special case. The following is based on [23, Chapter 1].

We consider a compact metric space (M, d) and let G be a group whose members act as isometries on M, i.e. for $g \in G$ and $t, s \in M$, $d(gs, gt) = d(s, t)$.

Theorem 3.2.1. *Let (M, d) be a compact metric space, G be a group acting on M by isometries. Then there exists a regular Borel-measure μ, which is G invariant, i.e.*

$$\mu(gA) = \mu(A) \text{ for all Borel sets } A \text{ and } g \in G.$$

Proof. For all $\epsilon > 0$, let N_ϵ be a minimal ϵ-net in M, i.e. $\bigcup_{x \in N_\epsilon} B(x, \epsilon) = M$, where $n_\epsilon = |N_\epsilon|$ is minimal with respect to this property and $B(x, \epsilon) = \{y \in M \mid d(x, y) \leq \epsilon\}$ denotes the closed ball. Note that we can find a finite covering due to the assumption of M being compact.
Define $C(M)$ as the space of continous functions on M (remember that $C(M)$ can be equipped with the supremum norm, since M is compact). Then we consider the mapping $\mu_\epsilon : C(M) \longrightarrow \mathbb{R}$ with

$$\mu_\epsilon(f) = \frac{1}{n_\epsilon} \sum_{x \in N_\epsilon} f(x).$$

In addition of being linear, $\mu_\epsilon(1) = 1$, μ_ϵ is positive, i.e. if $f \geq 0$, then $\mu_\epsilon(f) \geq 0$, and also for the operator norm

$$||\mu_\epsilon|| := \sup_{f \in C(M): ||f||_\infty \leq 1} |\mu_\epsilon(f)| \leq 1.$$

The family $\{\mu_\epsilon\}_{\epsilon > 0}$ is uniformly bounded in norm and thus by the theorem of Banach-Alaoglu 6.0.3 compact in the weak* topology. Using the property that a compact metric space is separable gives us a subsequence $\epsilon_i \to 0$ and a linear functional $\mu \in C(M)^*$ with

$$\mu_{\epsilon_i}(f) \to \mu(f).$$

The limiting process inherits the properties $f \geq 0 \implies \mu(f) \geq 0$ and $\mu(1) = 1$. By the Riesz-Markov representation theorem 6.0.4 μ can be represented by a unique probability measure which we (inconsistently) also denote by the letter μ. Hence, for $f \in C(M)$

$$\mu(f) = \int_M f(x) d\mu(x).$$

Now we need to show that this construction is independent of the sequence of minimal ϵ nets, since they need not be unique. Assume another family of minimal nets $\{N'_\epsilon\}_{\epsilon > 0}$ as well as a convergent subsequence $\mu'_{\epsilon_i} \to \mu'_\epsilon$ obtained by the previous procedure. We

use the Hall marriage theorem 6.0.1 for the bipartition sets N_ϵ and N'_ϵ with the relation $x \sim y$ for $(x,y) \in N_\epsilon \times N'_\epsilon$ iff $B(x,\epsilon) \cap B(y,\epsilon) \neq \emptyset$. Now assume $|A| > |K(A)|$ for some $A \subseteq N_\epsilon$ and choose some $z \in M$. We know there exists some $x \in N_\epsilon$ with $z \in B(x,\epsilon)$.
1. case: $x \in N_\epsilon \backslash A$. Then trivially $x \in N_\epsilon \backslash A \cup K(A)$.
2. case: $x \in A$. Then there exists a $y \in N'_\epsilon$ with $z \in B(y,\epsilon)$ and thus $B(x,\epsilon) \cap B(y,\epsilon) \neq \emptyset$ implying $y \in K(A)$. In either case we find a point in $x \in N_\epsilon \backslash A \cup K(A)$ contradicting the minimality of N_ϵ. Using Theorem 6.0.1 shows the existence of a bijective mapping $\psi : N_\epsilon \to N'_\epsilon$ with $x \sim \psi(x)$, in other words $d(x,\psi(x)) \leq 2\epsilon$. Then

$$|\mu_\epsilon(f) - \mu'_\epsilon(f)| \leq \frac{1}{n_\epsilon} \sum_{x \in N_\epsilon} |f(x) - f(\psi(x))| \leq \sup_{d(a,b) \leq 2\epsilon} |f(a) - f(b)| \xrightarrow{\epsilon \to 0} 0.$$

The limit on the right-hand is 0 since a continuous function on a compact set is uniformly continuous.
It remains to show that μ is invariant under G. Take $g \in G$, then for a minimal ϵ net N_ϵ also gN_ϵ is a minimal ϵ cover. Using that the construction of μ is independent of the used ϵ covering shows

$$\mu(f) = \lim_{\epsilon_i \to 0} \mu_{\epsilon_i}(f) = \lim_{\epsilon'_i \to 0} \mu'_{\epsilon'_i}(f \circ g) = \mu(f \circ g).$$

□

Remark. Under the additional assumption $GM = M$ one can show that the measure μ is unique up to a constant factor (e.g. Theorem 1.3 of [23]).

Application 8. Let $n \in \mathbb{N}$ and $1 \leq k \leq n$. The set of matrices

$$V_{n,k} = \{A \in \mathbb{R}^{n \times k} \mid A^T A = I_k\}$$

is called Stiefel manifold. We can equip $V_{n,k}$ with the metric

$$d(A,B) = \Big(\sum_{i=1}^{k} \|a_i - b_i\|_2^2\Big)^{\frac{1}{2}},$$

where a_i and b_i denote the columns of A and B respectively. $V_{n,k}$ is a closed and bounded subset of $\mathbb{R}^{n \times k}$ and hence compact. The orthogonal group $V_{n,n}$ acts on $V_{k,n}$ by isometries, as an elementary calculation shows. We can therefore apply Theorem 3.2.1 to show the existence of a (unique) invariant distribution μ^* on $V_{n,k}$. We can then construct a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $A_{n,k} : \Omega \to V_{n,k}$ (see

Theorem 6.0.16) such that for any Borel set $B \subseteq V_{n,k}$,

$$\mathbb{P}\big[A_{n,k} \in B\big] = \mu^*(B).$$

3.3 The Wasserstein spaces

In the main part of this thesis we will work with sets of probability measures on $\mathbb{R}$, thus we need to introduce a suitable topology on such spaces. For the general definition of the so called q-Wasserstein spaces we consider a Polish space $(\mathcal{X}, d)$ and will later restrict to the case $\mathcal{X} = \mathbb{R}$. The following is based on [28].

Definition 3.3.1. *(q-Wasserstein space)*
Let $q \geq 1$ and denote by $P(\mathcal{X})$ the set of probability measures on the Polish space $(\mathcal{X}, d)$. Then we define the q-Wasserstein space as

$$P_q(\mathcal{X}) := \left\{\mu \in P(\mathcal{X}) \mid \int_{\mathcal{X}} d(x_0, x)^q d\mu(x) < \infty\right\}$$

for some $x_0 \in \mathcal{X}$.

Remark. The point $x_0 \in \mathcal{X}$ makes no difference here, since for any $y_0 \in \mathcal{X}$ Minkowski's inequality implies

$$\begin{aligned}\left[\int_{\mathcal{X}} d(y_0, x)^q d\mu(x)\right]^q &\leq \left[\int_{\mathcal{X}} d(y_0, x_0)^q d\mu(x)\right]^q + \left[\int_{\mathcal{X}} d(x_0, x)^q d\mu(x)\right]^q \\ &= d(y_0, x_0) + \left[\int_{\mathcal{X}} d(x_0, x)^q d\mu(x)\right]^q.\end{aligned}$$

If the integral with x_0 is finite, then also the integral over $d(y_0, .)^q$.

Remark. The Wasserstein spaces are nested in an anti-lexicographical way, i.e. for $p \leq q$ we have $P_q(\mathcal{X}) \subseteq P_p(\mathcal{X})$. This can be verified directly. Take $\mu \in P_q(\mathcal{X})$, then

$$\begin{aligned}\int_{\mathcal{X}} d(x_0, x)^p d\mu(x) &= \int_{d(x_0,.)\leq 1} d(x_0, x)^p d\mu(x) + \int_{d(x_0,.)>1} d(x_0, x)^p d\mu(x) \\ &\leq 1 + \int_{\mathcal{X}} d(x_0, x)^q d\mu(x) < \infty\end{aligned}$$

and so $\mu \in P_p(\mathcal{X})$.

It is possible to define a natural metric on $P_q(\mathcal{X})$, which then induces the q-Wasserstein topology.

Definition 3.3.2. *For $\mu,\nu \in P_q(\mathcal{X})$ we define the distance*

$$W_q(\mu,\nu) = \Big(\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} d(x,y)^q d\pi(x,y)\Big)^{1/q},$$

where $\Pi(\mu,\nu)$ is the set of all „couplings" of (μ,ν), this means $\pi \in \Pi(\mu,\nu)$ iff $\pi \in P(\mathcal{X}\times\mathcal{X})$ and μ,ν are the marginals of π.

Remark. Using the inequality $d(x,y)^q \leq 2^q(d(x_0,x)^q + d(x_0,y)^q)$ implies

$$\int_{\mathcal{X}\times\mathcal{X}} d(x,y)^q d\pi(x,y) \leq 2^q \int_{\mathcal{X}} d(x_0,x)^q d\mu(x) + 2^q \int_{\mathcal{X}} d(x_0,y)^q d\nu(y).$$

Hence, W_q is well defined on $P_q(\mathcal{X})$.

Within $P_q(\mathcal{X})$ we can strengthen the concept of weak convergence a little bit and this will give us a nice property of W_q.

Definition 3.3.3. *(Weak convergence in P_q)*
Let $(\mathcal{X},d)$ be a Polish space, and $q \in [1,\infty)$. Let $\{\mu_k\}_{k\in\mathbb{N}}$ be a sequence of measures in $P_q(\mathcal{X})$ and let μ be another element in $P_q(\mathcal{X})$. Then $\{\mu_k\}_{k\in\mathbb{N}}$ is said to converge weakly in $P_q(\mathcal{X})$ if for some (and then any) $x_0 \in \mathcal{X}$, as $k \to \infty$,

$$\mu_k \text{ converges weakly to } \mu \text{ and } \int d(x_0,x)^q d\mu_k(x) \longrightarrow \int d(x_0,x)^q d\mu(x).$$

Remark. There are several equivalent formulations of weak convergence in $P_q(\mathcal{X})$.

At this point one might be interested in the relation between the Wasserstein topology and the weak topology.

Definition 3.3.4. *For $\mu,\nu \in P(\mathcal{X})$ we define the „bounded Lipschitz metric".*

$$d_{bL}(\mu,\nu) = \sup_{\phi\in C_{Lip}(\mathcal{X})} \Big\{ \int_{\mathcal{X}} \phi(x)d\mu(x) - \int_{\mathcal{X}} \phi(x)d\nu(x)\Big\},$$

where $C_{Lip}(\mathcal{X}) := \{\phi : \mathcal{X} \to \mathbb{R} \mid \phi$ Lipschitz continuous with $||\phi||_\infty + ||\phi||_{Lip} \leq 1\}$ and $||\phi||_{Lip}$ denotes the Lipschitz constant. d_{bL} is the metric induced by the characterisation of weak convergence using the set from Remark 6 from the Appendix.

Theorem 5.9 in [28] gives an alternative representation of W_1 via a supremum over Lipschitz continuous functions, namely

$$W_1(\nu,\mu) = \sup_{||\phi||_{Lip}\leq 1} \Big\{ \int_{\mathcal{X}} \phi(x)d\mu(x) - \int_{\mathcal{X}} \phi(x)d\nu(x)\Big\}.$$

This immediately leads to the inequality $d_{bL} \leq W_1$. Further, for $p \leq q$ we can use Hölder's inequality to show that $W_p \leq W_q$. For $\tilde{p} = q/p$ and $1/\tilde{q} = 1 - 1/\tilde{p}$, we have

$$\begin{aligned} W_p(\mu,\nu) &= \Big(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} d(x,y)^p d\pi(x,y) \Big)^{1/p} \\ &\leq \inf_{\pi \in \Pi(\mu,\nu)} \Big(\int_{\mathcal{X}\times\mathcal{X}} d(x,y)^{p\tilde{p}} d\pi(x,y) \Big)^{1/p\tilde{p}} \Big(\int_{\mathcal{X}\times\mathcal{X}} 1^{\tilde{q}} d\pi(x,y) \Big)^{1/p\tilde{q}} \\ &= W_q(\mu,\nu). \end{aligned}$$

In total we get the (not surprising) inequality

$$d_{bL}(\cdot,\cdot) \leq W_q(\cdot,\cdot),\ \forall q \geq 1.$$

The next theorem shows that W_q metrizes a certain convergence type, which quite naturally arises within the q-Wasserstein space.

Theorem 3.3.1. *(Theorem 6.8 in [28])*
Let $(\mathcal{X},d)$ be a Polish space, and $q \in [1,\infty)$. Then the Wasserstein distance W_q metrizes weak convergence in $P_q(\mathcal{X})$. In other words, if $\{\mu_k\}_{k\in\mathbb{N}}$ is a sequence of measures in $P_q(\mathcal{X})$ and μ is another measure in $P_q(\mathcal{X})$, then the statements

$$\mu_k \text{ converges weakly in } P_q(\mathcal{X}) \text{ to } \mu$$

and

$$W_q(\mu_k,\mu) \longrightarrow 0$$

are equivalent.

Finally it is possible to show that the q-Wasserstein topology is finer than the classical weak topology on $P_q(\mathcal{X})$. For a distance d on $P_q(\mathcal{X})$ we denote by $B^d(\nu,\epsilon)$ the ball of radius ϵ around $\nu \in P_q(\mathcal{X})$ with respect to d. Then, since $d_{bL} \leq W_q$ it follows that $B^{W_q}(\nu,\epsilon) \subseteq B^{d_{bL}}(\nu,\epsilon)$ for all $\epsilon > 0$. Hence, we can apply Hausdorff's criterion 6.0.6, where we use the set of open W_q balls as base of the q-Wasserstein topology and the system of open d_{bL} balls as base of neighbourhoods for the weak topology.

The next theorem shows that $P_q(\mathcal{X})$ inherits nice properties of its underlying space $\mathcal{X}$.

Theorem 3.3.2. *(Theorem 6.16 in [28])*
Let $\mathcal{X}$ be a complete separable metric space and $q \in [1,\infty)$. Then the Wasserstein space $P_q(\mathcal{X})$,

metrized by the Wasserstein distance W_q, is also a complete metric space. In short: The Wasserstein space over a Polish space is itself a Polish space. Moreover, any probability measure can be approximated by a sequence of probability measures with finite support.

Remark. Suppose we are given a bounded metric $\tilde{d}$ (e.g. $\tilde{d} = d/(1+d)$). Then $\tilde{P}_q(\mathcal{X}) = \left\{\mu \in P(\mathcal{X}) \mid \int_{\mathcal{X}} \tilde{d}(x_0, x)^q d\mu(x) < \infty\right\} = P(\mathcal{X})$ for all $q \geq 1$ and hence $P(\mathcal{X})$ equipped with the notion of weak convergence is a Polish space.

We now restrict to the situation $\mathcal{X} = \mathbb{R}$, where the q-Wasserstein space becomes

$$P_q(\mathbb{R}) = \left\{\nu \in P(\mathbb{R}) \;\Big|\; \int_{\mathbb{R}} |x|^q d\nu(x) < \infty\right\}.$$

We introduce the q-th moment map and a certain subset of $P(\mathbb{R})$

Definition 3.3.5. *For $\nu \in P_q(\mathbb{R})$ the p-th moment map is given by*

$$m_q(\nu) = \int_{\mathbb{R}} |x|^p d\nu(x).$$

Further, we consider the set

$$K_{2,j} = \left\{\nu \in P(\mathbb{R}) \mid m_2(\nu) \leq j\right\}.$$

Lemma 3.3.3. *The mapping $m_q : P(\mathbb{R}) \to [0, \infty]$ for $q \geq 1$ is lower-semicontinuous with respect to the weak topology on $P(\mathbb{R})$.*

Proof. Fix some $\alpha > 0$ and consider the level set $\psi_{m_q}(\alpha) = \{\nu \in P(\mathbb{R}) \mid m_q(\nu) \leq \alpha\}$. Take a sequence $\{\nu_n\}_{n \in \mathbb{N}} \subseteq \psi_{m_q}(\alpha)$ with $\nu_n \to \nu$ weakly, for some $\nu \in P(\mathbb{R})$. Then we have to show that $\nu \in \psi_{m_q}(\alpha)$. Let $M \in \mathbb{N}$ and consider the mapping $e_M : \mathbb{R} \to [0,1]$ with

$$e_M(x) := \begin{cases} 0, & \text{if } x \in [-M-1, M+1]^c \\ M+1-|x|, & \text{if } x \in [-M-1, -M] \cup [M, M+1] \\ 1, & \text{if } x \in [-M, M]. \end{cases}$$

For all $M \in \mathbb{N}$ we have that e_M is continuous and has compact support. Hence $x \mapsto |x|^q e_M(x)$ is continuous and bounded. Weak convergence of the sequence $\{\nu_n\}_{n \in \mathbb{N}}$ yields

$$\int_{\mathbb{R}} |x|^q e_M(x) d\nu(x) = \lim_{n \to \infty} \int_{\mathbb{R}} |x|^q e_M(x) d\nu_n(x) \leq \alpha.$$

Then by monotone convergence we get (note that $e_M \le e_{M+1}$ pointwise and $e_M \uparrow 1$)

$$\int_{\mathbb{R}} |x|^q d\nu(x) = \lim_{M\to\infty} \int_{\mathbb{R}} |x|^q e_M(x) d\nu(x) \le \alpha.$$

Thus $\nu \in \psi_{m_q}(\alpha)$. □

Lemma 3.3.4. *(Proposition 7.1.5 in [3])*
Let $q \ge 1$, a set $K \subseteq P_q(\mathbb{R})$ is relatively compact (i.e the W_q-closure of K is compact) iff it is q-uniformly integrable and tight. In particular, for a given sequence $\{\mu_n\}_{n\in\mathbb{N}} \subseteq P_q(\mathbb{R})$ and some $\mu \in P_q(\mathbb{R})$ we have

$$\lim_{n\to\infty} W_q(\mu_n, \mu) = 0 \iff \begin{cases} \{\mu_n\}_{n\in\mathbb{N}} \text{ converges weakly to } \mu \\ \{\mu_n\}_{n\in\mathbb{N}} \text{ has uniformly integrable } q\text{-moments}. \end{cases}$$

Lemma 3.3.5. *(Lemma 3.14 in [19])*
Fix $j \in \mathbb{N}$. The set $K_{2,j} \subseteq P_2(\mathbb{R})$ is compact with respect to the weak topology on $P(\mathbb{R})$. Moreover, for any $q < 2$, the set $K_{2,j}$ is compact with respect to q-Wasserstein topology. In addition $K_{2,j}$ is convex and non-empty.

Proof. $K_{2,j}$ is non-empty, since there are distributions with second moment less or equal j. If we consider $K_{2,j}$ as subset of the real vector space of finite signed measures (endowed with pointwise summation), then $K_{2,j}$ is convex. For $\nu \in K_{2,j}$ and $M \in \mathbb{N}$ we have

$$\nu([-M,M]^c) = \int_{[-M,M]^c} d\nu(x) \le \int_{[-M,M]^c} \frac{|x|^p}{M^p} d\nu(x) \le \frac{j}{M^p}.$$

Thus, the set $K_{2,j}$ is tight (see Definition 6.0.7) and by Prohorov's theorem 6.0.11, $K_{2,j}$ is weakly precompact (i.e. the weak closure of $K_{2,j}$ is compact in the weak topology). Since $K_{2,j}$ is a level set of the lower-semicontinuous mapping $\nu \mapsto m_2(\nu)$ (see Lemma 3.3.3), we have that $K_{2,j}$ is weakly closed. This implies that $K_{2,j}$ is compact in the weak topology.
Fix some $q < 2$ and note that

$$\sup_{\nu\in K_{2,j}} m_q(\nu) = \sup_{\nu\in K_{2,j}} \int_{\mathbb{R}} |x|^q d\nu(x) \le \sup_{\nu\in K_{2,j}} \left\{1 + \int_{[-1,1]^c} |x|^2 d\nu(x)\right\} \le 1 + j.$$

Thus, $K_{2,j}$ has uniformly integrable q-th moments. Lemma 3.3.4 implies that $K_{2,j}$ is relatively compact in the q-Wasserstein topology. To verify compactness, we consider a sequence $\{\nu_n\}_{n\in\mathbb{N}} \subseteq K_{2,j}$ with $\nu_n \xrightarrow{W_q} \nu \in P_q(\mathbb{R})$. We want to show that $\nu \in K_{2,j}$. Since $\{\nu_n\}_{n\in\mathbb{N}} \subseteq K_{2,j}$, we have that

$$\sup_{n\in\mathbb{N}} \int_{\mathbb{R}} |x|^2 d\nu_n(x) \leq j.$$

Moreover $\nu_n \to \nu$ weakly, since $\nu_n \xrightarrow{W_q} \nu$ in the stronger q-Wasserstein sense. By Lemma 3.3.4 we have that $\nu_n \xrightarrow{W_2} \nu$ and hence

$$\int_{\mathbb{R}} |x|^2 d\nu(x) = \lim_{n\to\infty} \int_{\mathbb{R}} |x|^2 d\nu_n(x) \leq j.$$

It follows that $\nu \in K_{2,j}$. $\square$

Remark. The previous lemma shows that the mapping $m_2 : P(\mathbb{R}) \to [0, \infty]$ is a GRF with respect to the weak topology.

4 Main results

In the main section we will elaborate the major results from [18], where we are interested in large deviations for random projections of high dimensional random vectors. Formally, this can be realized by a matrix vector multiplication $A_{n,k}^T X^{(n)}$, where $A_{n,k} \in V_{n,k}$ is drawn with respect to the Haar measure on the Stiefel manifold $V_{n,k}$ and independent of $X^{(n)} \in \mathbb{R}^n$. The dimension k of the subspace will sometimes be constant and sometimes tend to infinity as n grows. We assume that $A_{n,k}$ and $X^{(n)}$ are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ independent of n, which is possible by the theorem of Ionescu-Tulcea (see Theorem 6.0.15 in the Appendix).

Within the previous framework we state 4 assumptions, A, A*, B, C, which are then used to establish several LDPs including $\{X^{(n)}\}_{n\in\mathbb{N}}$, a sequence of random vectors, where each $X^{(n)}$ takes values in $\mathbb{R}^n$.

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

Remark. Later we will see several examples of random vectors satisfying Assumption A. In particular, the situation $s_n = n$ seems to occur naturally, this case will therefore be referred to as Assumption A*.

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

As a direct consequence of Assumption A one can derive a weak law of large numbers.

Lemma 4.0.1. *Let Assumption A hold. Moreover, we additionally assume J_X has a unique minimizer $m \in \mathbb{R}_+$. Then there exists a $c > 0$ and a sequence of positive real numbers $\{\delta_l\}_{l\in\mathbb{N}}$ with $\delta_l \downarrow 0$ such that*

$$\sqrt{s_l}\inf\left\{J_X(x) : |x - m| \geq \delta_l\right\} \geq c, \tag{4.0.1}$$

where $\{s_l\}_{l\in\mathbb{N}}$ is the speed of the LDP from Assumption A. Moreover, for $\varepsilon_l := \max\{\delta_l, 2\exp(-c\sqrt{\delta_l})\}$, *we have the asymptotic thin-shell bound*

$$\forall l \in \mathbb{N} : \exists N_l \in \mathbb{N} : \forall n \geq N_l : \mathbb{P}\left[\left|\frac{\|X^{(n)}\|_2}{\sqrt{n}} - m\right| \geq \varepsilon_l\right] \leq \varepsilon_l. \tag{4.0.2}$$

In particular, we get the following weak limit

$$\frac{\|X^{(n)}\|_2}{\sqrt{n}} \xrightarrow{\mathbb{P}} m,$$

where $\xrightarrow{\mathbb{P}}$ *denotes convergence in probability.*

Proof. Let $\{s_l\}_{l\in\mathbb{N}}$ be the speed of the LDP from Assumption A and choose $\delta_1 \in \mathbb{R}$ such that $0 < \delta_1 < m$. Since J_X is a GRF, we can attain the following infimum

$$c := \inf\left\{J_X(x) : |x - m| \geq \delta_1\right\}$$

at some $y \in (m - \delta_1, m + \delta_1)^c$ as shown in Lemma 3.1.1. Since m is the unique minimizer of J_X (with $J_X(m) = 0$), we know that $c > 0$. For $l \in \mathbb{N}$ with $l \geq 2$, we define

$$\delta_l' := \inf\left\{\delta > 0 : \inf\left\{J_X(x) : |x - m| \geq \delta\right\}\sqrt{s_l} \geq c\right\}.$$

Then, by construction we have that $\delta_{l+1}' \leq \delta_l'$ for all $l \in \mathbb{N}$ (since $\{s_l\}_{l\in\mathbb{N}}$ is monotone increasing). Moreover, $\delta_l' \downarrow 0$ for $l \to \infty$. If not, then there exists $\epsilon > 0$ with $\lim_{l\to\infty} \delta_l \geq \epsilon > 0$. Hence,

$$\inf\left\{J_X(x) : |x - m| > \frac{\epsilon}{2}\right\}\sqrt{s_l} < c, \qquad \forall l \in \mathbb{N}.$$

Which is only possible if the infimum is equal to zero, contradicting the uniqueness of m. Define $\delta_l := \delta_l' + \frac{1}{l}$ for $l \geq 2$. It follows that $\{\delta_l\}_{l\in\mathbb{N}}$ is a sequence of positive real numbers, monotone decreasing for $l \geq 2$ and $\delta_l \downarrow 0$ as $l \to \infty$. Moreover, we have that 4.0.1 holds. To see this, we assume there exists $l_0 \in \mathbb{N}$ with

$$\sqrt{s_{l_0}} \inf\left\{J_X(x) : |x - m| \geq \delta_{l_0}\right\} < c.$$

Then, by construction $\delta_{l_0}' \geq \delta_{l_0} = \delta_{l_0}' + \frac{1}{l_0}$, which is not possible. This gives us the first claim of our lemma.

Fix $l \in \mathbb{N}$ and consider the large deviation upper bound for $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ with the closed set $(m-\varepsilon_l, m+\varepsilon_l)^c$, i.e.

$$\limsup_{n\to\infty} \frac{1}{s_n} \log \mathbb{P}\Big[\frac{\|X^{(n)}\|_2}{\sqrt{n}} \in (m-\varepsilon_l, m+\varepsilon_l)^c\Big] \leq -\inf_{x\in(m-\varepsilon_l,m+\varepsilon_l)^c} J_X(x). \tag{4.0.3}$$

We can use the definition of $\varepsilon_l = \max\{\delta_l, 2\exp(-c\sqrt{\delta_l})\} \geq \delta_l$, to bound the infimum in 4.0.3,

$$-\inf_{x\in(m-\varepsilon_l,m+\varepsilon_l)^c} J_X(x) \leq -\inf_{x\in(m-\delta_l,m+\delta_l)^c} J_X(x) \leq -\frac{c}{\sqrt{s_l}}.$$

The second inequality follows from the definition of δ_l. Now we distinguish two cases, first assume there exists $N_l' \in \mathbb{N}$ such that

$$\sup_{n\geq N_l'} \frac{1}{s_n} \log \mathbb{P}\Big[\frac{\|X^{(n)}\|_2}{\sqrt{n}} \in (m-\varepsilon_l, m+\varepsilon_l)^c\Big] \leq -\frac{c}{\sqrt{s_l}}.$$

Rearranging the terms yields

$$\mathbb{P}\Big[\frac{\|X^{(n)}\|_2}{\sqrt{n}} \in (m-\varepsilon_l, m+\varepsilon_l)^c\Big] \leq \exp\Big(-\frac{cs_n}{\sqrt{s_l}}\Big),$$

for all $n \geq N_l'$. In particular, we find $N_l \in \mathbb{N}$ with $N_l \geq N_l'$ such that

$$\exp\Big(-\frac{cs_n}{\sqrt{s_l}}\Big) \leq 2\exp(-c\sqrt{s_l}) \leq \varepsilon_l,$$

for all $n \geq N_l$. Thus, in this case, we have that

$$\begin{aligned}\mathbb{P}\Big[\Big|\frac{\|X^{(n)}\|_2}{\sqrt{n}} - m\Big| \geq \varepsilon_l\Big] &= \mathbb{P}\Big[\frac{\|X^{(n)}\|_2}{\sqrt{n}} \in (m-\varepsilon_l, m+\varepsilon_l)^c\Big] \\ &\leq \exp\Big(-\frac{cs_n}{\sqrt{s_l}}\Big) \\ &\leq \varepsilon_l.\end{aligned}$$

Now we assume that for all $N_l' \in \mathbb{N}$ we have

$$\sup_{n\geq N_l'} \frac{1}{s_n} \log \mathbb{P}\Big[\frac{\|X^{(n)}\|_2}{\sqrt{n}} \in (m-\varepsilon_l, m+\varepsilon_l)^c\Big] > -\frac{c}{\sqrt{s_l}}.$$

Together with the large deviation upper bound for $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ this can only happen if

$$\lim_{n\to\infty} \frac{1}{s_n} \log \mathbb{P}\Big[\frac{\|X^{(n)}\|_2}{\sqrt{n}} \in (m - \varepsilon_l, m + \varepsilon_l)^c\Big] = -\frac{c}{\sqrt{\delta_l}}.$$

Since the right-hand side is negative and $s_n \to \infty$, the probability in the log tends to zero, i.e.

$$\mathbb{P}\Big[\frac{\|X^{(n)}\|_2}{\sqrt{n}} \in (m - \varepsilon_l, m + \varepsilon_l)^c\Big] \xrightarrow{n\to\infty} 0.$$

In particular, we find $N_l \in \mathbb{N}$ such that for all $n \in \mathbb{N}$ with $n \geq N_l$

$$\mathbb{P}\Big[\frac{\|X^{(n)}\|_2}{\sqrt{n}} \in (m - \varepsilon_l, m + \varepsilon_l)^c\Big] \leq \varepsilon_l.$$

□

Remark. The sequence $\{\varepsilon_l\}_{l\in\mathbb{N}}$ converges to zero as $l \to \infty$. Hence, 4.0.2 is an asymptotic version of the „concentration hypothesis“ from [4].

The following assumption is a slight modification of Assumption A*, allowing to consider more types of sequences $\{X^{(n)}\}_{n\in\mathbb{N}}$.

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

Lemma 4.0.2. *If we assume the modification of Assumption B with* $\lim_{n\to\infty} s_n/n = r \in (0, \infty)$ *and GRF* $J_X^{(r)}$*, then this modified Assumption B is equivalent to A*, i.e. it holds with GRF* $J_X^{(r)}$ *iff Assumption A* holds with GRF* $J_X(x) = rJ_X^{(r)}(\sqrt{r}x)$.

Proof. We assume A* and show the LDP for the modified Assumption B with GRF $\frac{1}{r}J_X(\frac{1}{\sqrt{r}}x)$.

A direct calculation shows that the sequence $\{\sqrt{r}\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies the LDP with speed s_n and GRF $\frac{1}{r}J_X(\frac{1}{\sqrt{r}}x)$. Now we need to show that $\{\sqrt{s_n}\|X^{(n)}\|_2/n\}_{n\in\mathbb{N}}$ is asymptotically equivalent to $\{\sqrt{r}\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ with speed s_n, i.e. for $\delta > 0$

$$\limsup_{n\to\infty} \frac{1}{s_n} \log\left(\mathbb{P}\left[\left|\sqrt{r}\frac{\|X^{(n)}\|_2}{\sqrt{n}} - \frac{\sqrt{s_n}\|X^{(n)}\|_2}{n}\right| > \delta\right]\right) = -\infty. \tag{4.0.4}$$

First we note that in the lim sup, $|\sqrt{r} - \sqrt{\frac{s_n}{n}}|$ can be made arbitrarily small, in other words

$$\limsup_{n\to\infty} \frac{1}{s_n} \log\left(\mathbb{P}\left[\frac{\|X^{(n)}\|_2}{\sqrt{n}}\left|\sqrt{r} - \sqrt{\frac{s_n}{n}}\right| > \delta\right]\right) \leq \limsup_{n\to\infty} \frac{1}{s_n} \log\left(\mathbb{P}\left[\frac{\|X^{(n)}\|_2}{\sqrt{n}} \geq M\right]\right)$$

for every $M > 0$. Using the LDP of Assumption A* and the asymptotic equivalence of s_n and n, we can bound the right-hand side of the inequality,

$$\limsup_{n\to\infty} \frac{1}{s_n} \log\left(\mathbb{P}\left[\frac{\|X^{(n)}\|_2}{\sqrt{n}} \geq M\right]\right) \leq -\frac{1}{r} \inf_{x\geq M} J_X(x).$$

This implies that the lim sup in (4.0.4) can be bounded from above by $-\inf_{x\geq M} J_X(x)/r$ for any $M > 0$. Assume $\inf_{x\geq M} J_X(x)$ is uniformly bounded from above by a constant $c \geq 0$, i.e. for all $M > 0$

$$\inf_{x\geq M} J_X(x) \leq c.$$

Then the level set $\psi_{J_X}(c) = \{x \in \mathbb{R} \mid J_X(x) \leq c\}$ is compact, since J_X is a GRF. Hence, there exists an $a^* \in \psi_{J_X}(c)$ with $a^* = \sup\{x : x \in \psi_{J_X}(c)\}$. By assumption, $\inf_{x\geq a^*+1} J_X(x) \leq c$, but $J_X(x) > c$ for all $x \geq a^* + 1$. Otherwise $x \in \psi_{J_X}(c)$, which is not possible, since $x \geq a^* + 1 > a*$. Thus, on the one hand we have

$$\inf_{x\geq a^*+1} J_X(x) \leq c,$$

on the other hand we got

$$J_X(x) > c,$$

for all $x \geq a^* + 1$. This is only possible if $\lim_{x\to\infty} J_X(x) = c$, a contradiction, since all level sets are bounded. Which implies that

$$\lim_{M\to\infty} \inf_{x\geq M} J_X(x) = \infty.$$

□

Finally we are introducing Assumption C, which allows to use a relation between the speed s_n of the LDP and the dimension k_n of the subspace we are projecting on.

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

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

Lemma 4.0.3. *For $r \in (0, \infty)$, Assumption C holds with speed s_n and GRF $J_X^{(r)}$ iff it also holds with $r' = 1, s_n' = k_n$ and GRF $J_X^{(1)}(x) = rJ_X^{(r)}(\sqrt{r}x)$.*

Remark. The proof of Lemma 4.0.3 is very similar to the one of Lemma 4.0.2.

We start with two lemmas, which we will need several times in the following sections. Denote by $A_{n,k}$ a random projection matrix on the Stiefel manifold $V_{n,k}$ drawn uniformly with respect to the Haar probability measure on it.

Lemma 4.0.4. *(Lemma 4.1 in [18])*
Fix $k, n \in \mathbb{N}$ such that $k \le n$. Then

$$A_{n,k}(1,\cdot) \overset{(d)}{=} \frac{(\zeta_1,\ldots,\zeta_k)}{\|(\zeta_1,\ldots,\zeta_n)\|_2},$$

where $\{\zeta_j\}_{j\in\mathbb{N}}$ is an iid sequence of standard normal distributed random variables and $A_{n,k}(1,\cdot)$ denotes the first row of $A_{n,k}$.

Proof. Let O_n be an orthogonal $n \times n$ matrix sampled from the normalized Haar measure. We define

$$A'_{n,k} := O_n^T I_{n,k},$$

where $I_{n,k} \in \mathbb{R}^{n\times k}$ is a matrix with ones on the diagonal and zeros else. Since O_n is translation invariant we see that for any $n \times n$ deterministic orthogonal matrix E_n,

$$\begin{aligned} O_n \overset{(d)}{=} O_n E_n^T &\implies O_n^T \overset{(d)}{=} E_n O_n^T \\ &\implies O_n^T I_{n,k} \overset{(d)}{=} E_n O_n^T I_{n,k}. \end{aligned}$$

This shows $A'_{n,k} \overset{(d)}{=} E_n A'_{n,k}$ and hence $A'_{n,k} \overset{(d)}{=} A_{n,k}$. Since each row of O_n is uniformly distributed on the sphere, the result follows because the normalized n-multivariate standard normal distribution is itself uniformly distributed on the sphere. $\square$

Lemma 4.0.5. *(Lemma 4.2 in [18])*
Fix $n \in \mathbb{N}$ and $k \le n$. Suppose $X^{(n)}$ is an n-dimensional random vector independent of $A_{n,k}$. Then

$$\left(A_{n,k}^T \frac{X^{(n)}}{\|X^{(n)}\|_2}, \|X^{(n)}\|_2\right) \overset{(d)}{=} \left(A_{n,k}^T e_1, \|X^{(n)}\|_2\right), \tag{4.0.5}$$

where $e_1 := (1,\ldots,0)^T \in \mathbb{R}^n$ denotes the first unit vector.

Proof. We prove the claim by direct computation. Let $B \subseteq \mathbb{R}^{k+1}$ be a Borel set and let O_n be an $n \times n$ random „Householder transformation" matrix (see Theorem 6.0.20 in the Appendix) with

$$O_n \frac{X^{(n)}}{\|X^{(n)}\|_2} = e_1 = (1,0,\ldots,0)^T \in \mathbb{R}^n.$$

We can achieve this by defining

$$O_n := I_n - 2v_n v_n^T \in \mathbb{R}^{n\times n},$$

where v_n is an $\mathbb{R}^n$-valued random vector with

$$v_n := \frac{\frac{X^{(n)}}{\|X^{(n)}\|_2} - e_1}{\left\|\frac{X^{(n)}}{\|X^{(n)}\|_2} - e_1\right\|_2}.$$

By construction O_n is an orthogonal matrix and only depends on $X^{(n)}$. Then we get

$$\begin{aligned}
\mathbb{P}\left[\left(A_{n,k}^T \frac{X^{(n)}}{\|X^{(n)}\|_2}, \|X^{(n)}\|_2\right) \in B\right] &= \mathbb{P}\left[\left(A_{n,k}^T O_n^T e_1, \|X^{(n)}\|_2\right) \in B\right] \\
&= \mathbb{E}\left[\mathbb{P}\left[\left(A_{n,k}^T O_n^T e_1, \|X^{(n)}\|_2\right) \in B \middle| X^{(n)}\right]\right] \\
&= \mathbb{E}\left[\mathbb{P}\left[\left(A_{n,k}^T O_n^T e_1, \|x\|_2\right) \in B\right]\Big|_{x=X^{(n)}}\right] \\
&= \mathbb{E}\left[\mathbb{P}\left[\left(A_{n,k}^T e_1, \|x\|_2\right) \in B\right]\Big|_{x=X^{(n)}}\right] \\
&= \mathbb{P}\left[\left(A_{n,k}^T e_1, \|X^{(n)}\|_2\right) \in B\right],
\end{aligned}$$

where we used in the second equality from below that O_n is constant given $X^{(n)}$ (see Theorem 6.0.14 in the Appendix) and $O_n A_{n,k} \overset{(d)}{=} A_{n,k}$ in this case. □

Remark. As elaborated in [15], one can further show the result that $A_{n,k}^T \frac{X^{(n)}}{\|X^{(n)}\|_2}$ is independent of $X^{(n)}$.

Definition 4.0.1. *We now consider the situation when $k = k_n$ grows in n and denote the empirical measure of the top row of $\sqrt{n}A_{n,k_n}$ by*

$$\hat{\mu}_{\mathbf{A}}^n := \frac{1}{k_n}\sum_{j=1}^{k_n} \delta_{\sqrt{n}A_{n,k_n}(1,j)}.$$

As suggested by the previous lemma, we want to establish a relation to the empirical measure of the random projection

$$L^n := \frac{1}{k_n}\sum_{j=1}^{k_n} \delta_{(A_{n,k_n}^T X^{(n)})_j}.$$

Lemma 4.0.6. *(Lemma 4.3 in [18])*
For $n \in \mathbb{N}$, let $X^{(n)}$ be independent of A_{n,k_n}. Then we have

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

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

Proof. It follows from Lemma 4.0.5 that

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

For the continuity of the mapping we recall the fact that weak convergence of measures is the same as convergence in distribution for associated random variables. Let $\{\nu_n\}_{n\in\mathbb{N}} \subseteq P(\mathbb{R})$ and $\{c_n\}_{n\in\mathbb{N}} \subseteq (0, \infty)$ such that

$$\begin{aligned} \nu_n &\xrightarrow{w} \nu \in P(\mathbb{R}) \\ c_n &\longrightarrow c \in (0, \infty) \end{aligned}$$

as well as $X_n \sim \nu_n$ for all $n \in \mathbb{N}$. Then Slutsky's theorem 6.0.13 implies

$$\lim_{n\to\infty} c_n X_n \overset{(d)}{=} cX,$$

where X is distributed according ν. In terms of distributions this means

$$\nu_n(\cdot \times c_n^{-1}) \xrightarrow{w} \nu(\cdot \times c^{-1}).$$

This argument shows the continuity of the defined mapping since $P(\mathbb{R})$ is a Polish space (see Remark 3.3). □

Lemma 4.0.7. *(Lemma 4.4 in [18])*
Let $q \geq 1$ and suppose $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$, the empirical measure from Definition 4.0.1, satisfies an LDP in $P_q(\mathbb{R})$ and GRF $\mathbb{I}_1 : P_q(\mathbb{R}) \to [0, \infty]$. Let $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfy an LDP at speed s_n with GRF $\mathbb{I}_2 : \mathbb{R}_+ \to [0, \infty]$. Then, $\{L^n\}_{n\in\mathbb{N}}$ statisfies an LDP in $P_q(\mathbb{R})$ at speed s_n with

GRF $\mathbb{I} : P_q(\mathbb{R}) \to [0,\infty]$ *defined by*

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

Proof. A_{n,k_n} and $X^{(n)}$ are independent, hence $\{(\hat{\mu}_A^n, \|X^{(n)}\|_2/\sqrt{n})\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF

$$P_q(\mathbb{R}) \times \mathbb{R}_+ \ni (\nu, c) \mapsto \mathbb{I}_1(\nu) + \mathbb{I}_2(c).$$

We need to show that the mapping $F : P_q(\mathbb{R}) \times \mathbb{R}_+ \to P_q(\mathbb{R})$ with

$$(\nu, c) \mapsto \nu(\cdot \times c^{-1})$$

is continuous. Take a sequence $\{(\nu_n, c_n)\}_{n\in\mathbb{N}} \subseteq P_q(\mathbb{R}) \times \mathbb{R}_+$ with $\nu_n \overset{W_q}{\longrightarrow} \nu \in P_q(\mathbb{R})$ and $c_n \overset{n\to\infty}{\longrightarrow} c > 0$. Since $\nu_n \longrightarrow \nu$ weakly, we can use Lemma 4.0.6 to show that $F(\nu_n, c_n) \longrightarrow F(\nu, c)$ weakly. For the q-th moment we have

$$\begin{aligned} m_q(F(\nu_n, c_n)) &= \int_{\mathbb{R}} |x|^q d\nu_n(x \times c_n^{-1}) \\ &= c_n^q \int_{\mathbb{R}} |x|^q d\nu_n(x) \overset{n\to\infty}{\longrightarrow} c \int_{\mathbb{R}} |x|^q d\nu(x) = m_q(F(\nu, c)). \end{aligned}$$

Thus, $F(\nu_n, c_n) \overset{W_q}{\longrightarrow} F(\nu, c)$ and hence F is continuous. The claim then follows by applying the contraction principle. □

Lemma 4.0.8. *(Lemma 4.5 in [18])*
Let $\{\zeta\}_{j\in\mathbb{N}}$ *be an iid sequence of standard normal distributed random variables, denote* $\zeta^{(n)} = (\zeta_1, ..., \zeta_n)$ *and consider the sequence*

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

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

$$\chi_{\gamma_1}(\nu) := \begin{cases} 0, & \textit{if } \nu = \gamma_1 \\ \infty, & \textit{else.} \end{cases}$$

Proof. The first assertion follows directly by applying Sanov's theorem 3.1.16. The second claim is a consequence of Remark 3.1.2. $\square$

The goal is to establish LDPs for three different asymptotic regimes for k. We therefore distinguish the following cases.

Definition 4.0.2. *Given a sequence $\{k_n\}_{n\in\mathbb{N}}$, we say:*

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

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

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

4.1 LDP in the constant regime

In the previous chapter, we identified the distribution of the first row $A_{n,k}(1,.)$ of the projection matrix. The next lemma establishes a corresponding LDP

Lemma 4.1.1. *Let $\{\zeta_j\}_{j\in\mathbb{N}}$ be an iid sequence of standard normal distributed random variables. Then*

$$Z_n := \frac{(\zeta_1, \dots, \zeta_k)}{||(\zeta_1, \dots, \zeta_n)||_2}$$

satisfies the LDP with speed n and GRF $J_k : \mathbb{R}^k \to [0, \infty]$, $J_k(y) = -\frac{1}{2}\log(1 - ||y||_2^2)$.

Proof. We can represent our sequence as

$$Z_n = F\left(\frac{1}{\sqrt{n}}(\zeta_1, \dots, \zeta_k), \frac{1}{n}\sum_{j=k+1}^{n} \zeta_j^2\right),$$

where $F : \mathbb{R}^k \times \mathbb{R}_+ \to \mathbb{R}^k$ with

$$(x_1, x_2) \mapsto \frac{x_1}{\sqrt{||x_1||_2^2 + x_2}}$$

is a continuous mapping. Then we can use Application 5 for $(\zeta_1, \dots, \zeta_k)/\sqrt{n}$ and Application 6 for $\sum_{j=k+1}^{n} \zeta_j^2/n$ to establish an LDP for $\{Z_n\}_{n\in\mathbb{N}}$ in $\mathbb{R}^k$ at speed n via the contraction principle. For $||y||_2 \leq 1$ the corresponding GRF is given by

$$\begin{aligned} J_k(y) &= \inf\left\{ \frac{||x_1||_2^2}{2} + \frac{1}{2}(x_2 - \log x_2 - 1) : y = \frac{x_1}{\sqrt{||x_1||_2^2 + x_2}}, x_1 \in \mathbb{R}^k, x_2 > 0 \right\} \\ &= \frac{1}{2}\inf\left\{ \frac{||x_1||_2^2}{||y||_2^2} - \log\left(\frac{||x_1||_2^2(1 - ||y||_2^2)}{||y||_2^2} \right) - 1 : y = \frac{x_1}{\sqrt{||x_1||_2^2 + x_2}}, x_1 \in \mathbb{R}^k, x_2 > 0 \right\} \\ &= -\frac{1}{2}\log(1 - ||y||_2^2) + \frac{1}{2}\inf\left\{ \frac{||x_1||_2^2}{||y||_2^2} - \log\left(\frac{||x_1||_2^2}{||y||_2^2} \right) - 1 : y = \frac{x_1}{\sqrt{||x_1||_2^2 + x_2}}, x_1 \in \mathbb{R}^k, x_2 > 0 \right\} \\ &= -\frac{1}{2}\log(1 - ||y||_2^2). \end{aligned}$$

The infimum vanishes, since $t - \log t - 1$ has a global minimum at $t = 1$ and by choosing $x_2 = 1 - ||x_1||_2, x_1 = y$ we can achieve this value. □

The next lemma provides an adaptation of the contraction principle, showing an LDP for certain dependent random variables.

Lemma 4.1.2. *Suppose $\{U_n\}_{n\in\mathbb{N}}$, $\{V_n\}_{n\in\mathbb{N}}$ and $\{W_n\}_{n\in\mathbb{N}}$ satisfy LDPs in $\mathbb{R}$ at speeds α_n, β_n and γ_n with rate functions J_U, J_V and J_W, respectively. Let $\alpha_n = \beta_n \ll \gamma_n$. Assume $\{U_n\}_{n\in\mathbb{N}}$ is independent of $\{V_n\}_{n\in\mathbb{N}}$ and J_W has a unique minimizer m. Then $\{(U_n, V_n, W_n)\}_{n\in\mathbb{N}}$ is asymptotically equivalent to $\{(U_n, V_n, m)\}_{n\in\mathbb{N}}$ and satisfies an LDP at speed α_n with GRF*

$$J(u, v, w) := \begin{cases} J_U(u) + J_V(v), & w = m \\ \infty, & \textit{otherwise}. \end{cases} \tag{4.1.1}$$

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

Proof. For $n \in \mathbb{N}$ define $\mathcal{F}_n := (U_n, V_n, W_n)$ and $\tilde{\mathcal{F}}_n := (U_n, V_n, m)$. Since U_n and V_n are independent and $\alpha_n = \beta_n$, we can use Lemma 3.1.6 to show that $\{(U_n, V_n)\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^2$ and GRF $(u, v) \mapsto J_U(u) + J_V(v)$. Another application of Lemma 3.1.6, where we contract the LDP of $\{(U_n, V_n)\}_{n\in\mathbb{N}}$ with the trivial LDP of the constant sequence m yields an LDP for $\{\tilde{\mathcal{F}}_n\}_{n\in\mathbb{N}}$ at speed α_n with GRF J given in Equation (4.1.1). In order to show asymptotic equivalence of $\{\tilde{\mathcal{F}}_n\}_{n\in\mathbb{N}}$ and $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$, fix $\epsilon > 0$ and consider

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{\alpha_n}\log\mathbb{P}\big[\|\mathcal{F}_n-\tilde{\mathcal{F}}_n\|_2>\epsilon\big]&=\limsup_{n\to\infty}\frac{1}{\alpha_n}\log\mathbb{P}\big[|W_n-m|>\epsilon\big]\\&\le\limsup_{n\to\infty}\frac{\gamma_n}{\alpha_n}\frac{1}{\gamma_n}\log\mathbb{P}\big[W_n\in(m-\epsilon/2,m+\epsilon/2)^c\big]\\&\le -M\inf_{x\in(m-\epsilon/2,m+\epsilon/2)^c}J_W(x)\end{aligned}$$

for any $M>0$, since $\gamma_n\gg\alpha_n$. Because m is the unique minimizer of J_W and a GRF takes its infimum over closed sets, the infimum in the lowest line is positive, implying that the limsup in the first line has to be $-\infty$. This proves the exponential equivalence of $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ and $\{\tilde{\mathcal{F}}_n\}_{n\in\mathbb{N}}$, hence they satisfy the same LDP.
The second assertion follows by application of the contraction principle using the continuous mapping $F:\mathbb{R}^3\to\mathbb{R}$ with $(u,v,w)\mapsto vw$. Then, $\{F(\tilde{\mathcal{F}}_n)\}_{n\in\mathbb{N}}$ satisfies the LDP with GRF $J':\mathbb{R}\to[0,\infty]$ given by

$$\begin{aligned}J'(y)&=\inf\{J(u,v,w)\mid(u,v,w)\in\mathbb{R}^3:y=vw\}\\&=\inf\{J_U(u)+J_V(v)\mid(u,v,w)\in\mathbb{R}^3:v=y/m,m=w\}\\&=J_V(y/m).\end{aligned}$$

In the last step we used that the global infimum of J_U is zero and that $v=y/m$ is constant. □

We are now able to state the first main result, establishing an LDP for projections of high dimensional random vectors on subspaces with fixed dimension.

Theorem 4.1.3. *(Theorem 2.6 in [18])*
Suppose $\{k_n\}_{n\in\mathbb{N}}$ is constant at $k\in\mathbb{N}$, and that either Assumption A or Assumption B holds, with sequence $\{s_n\}_{n\in\mathbb{N}}$ and GRF J_X. Then $\{n^{-1/2}A_{n,k}^TX^{(n)}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^k$ at speed s_n, with GRF $I_{AX,k}:\mathbb{R}^k\to[0,\infty]$ defined by*

$$I_{AX,k}(x):=\begin{cases}\inf_{0<c<1}\left\{J_X\big(\frac{\|x\|_2}{c}\big)-\frac{1}{2}\log(1-c^2)\right\}, & \textit{if Assumption A* holds}\\ \inf_{c>0}\left\{J_X\big(\frac{\|x\|_2}{c}\big)+\frac{c^2}{2}\right\}, & \textit{if Assumption B holds.}\end{cases}$$

Proof. Case 1. Suppose Assumption A* holds with GRF J_X:

By Lemma 4.0.5 we have

$$\frac{1}{\sqrt{n}} A_{n,k}^T X^{(n)} \overset{(d)}{=} A_{n,k}(1,.) \frac{||X^{(n)}||_2}{\sqrt{n}}.$$

$A_{n,k}(1,.)$ and $||X^{(n)}||_2$ are independent by assumption, hence we can use the contraction principle to combine the respective LDPs with GRF

$$I_{AX,k}(x) = \inf_{y \in \mathbb{R}^k, z \in \mathbb{R}} \left\{ -\frac{1}{2} \log(1 - ||y||_2^2) + J_X(z) : x = zy, ||y||_2 \le 1, z \ge 0 \right\}.$$

Since J_X is infinity on the negative reals, we can restrict on positive values for z. With $y = x/z$ the condition $||y||_2 \le 1$ becomes $||x||_2 \le z$. Then

$$I_{AX,k}(x) = \inf_{||x||_2 \ge z} \left\{ -\frac{1}{2} \log\left(1 - \frac{||x||_2^2}{z^2}\right) + J_X(z) \right\} = \inf_{||x||_2 > z} \left\{ -\frac{1}{2} \log\left(1 - \frac{||x||_2^2}{z^2}\right) + J_X(z) \right\}.$$

Using the parametrisation $c = ||x||_2/z$ we obtain the desired form of the rate function.

Case 2. Suppose Assumption B holds with sequence $\{s_n\}_{n \in \mathbb{N}}$ and GRF J_X. Using again Lemma 4.0.5 and Lemma 4.0.4 yields

$$\frac{1}{\sqrt{n}} A_{n,k}^T X^{(n)} \overset{(d)}{=} A_{n,k}(1,.) \frac{||X^{(n)}||_2}{\sqrt{n}} \overset{(d)}{=} \frac{\zeta^{(k)}/\sqrt{s_n}}{||\zeta^{(n)}||_2/\sqrt{n}} \frac{\sqrt{s_n}||X^{(n)}||_2}{n}, \tag{4.1.2}$$

where we have the product of the three entries of the vector

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

As shown in Application (5), $\{\zeta^{(k)}/\sqrt{s_n}\}_{n \in \mathbb{N}}$ satisfies the LDP with speed s_n. For the second component one can use Application (6) and the contraction principle for the continuous mapping $F : \mathbb{R}_+ \to \mathbb{R}_+, x \mapsto x^{-1/2}$ to receive an LDP for $\{\sqrt{n}/||\zeta^{(n)}||_2\}_{n \in \mathbb{N}}$ with speed $n \gg s_n$. Since $\zeta_i \sim \mathcal{N}(0,1)$ for $i \in \mathbb{N}$, we have that

$$\frac{1}{n} \sum_{i=1}^{n} \zeta_i^2 \longrightarrow 1$$

almost surely by the strong law of large numbers. Thus, $\sqrt{n}/||\zeta^{(n)}||_2 \to 1$ almost surely. The third component of R_n satisfies the LDP with GRF J_X by Assumption B. Hence, we can apply Lemma 4.1.1 to see that $\{R_n\}_{n\in\mathbb{N}}$ satisfies the LDP at speed s_n and GRF

$$J(u,v,w) := \begin{cases} \frac{1}{2}\|u\|_2^2 + J_X(w), & v = 1 \\ \infty, & \text{otherwise.} \end{cases}$$

Applying the contraction principle to the sequence $\{R_n\}_{n\in\mathbb{N}}$ and the continuous function $F : \mathbb{R}^k \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}, (u,v,w) \mapsto uvw$ gives the LDP for (4.1.2) with speed s_n and GRF

$$\begin{aligned} I_{AX,k}(x) &= \inf\left\{\frac{1}{2}\|u\|_2^2 + J_X(w) : u \in \mathbb{R}^k, w \in \mathbb{R} : v = 1, x = uvw\right\} \\ &= \inf\left\{\frac{1}{2}\|u\|_2^2 + J_X(w) : u \in \mathbb{R}^k, w > 0, x = uw\right\} \\ &= \inf\left\{\frac{1}{2}c^2 + J_X(\|x\|_2/c) : c > 0\right\}. \end{aligned}$$

We used that J_X is infinite on the negative real numbers, due to the positivity of the norm and substituted $c = \|u\|_2$ in the last step. □

As a direct consequence we get the LDP for scaled q-norms. Similar results are much more work in the sublinear and linear regimes. We define

$$Y_{q,k}^n := n^{-1/q}\|A_{n,k}X^{(n)}\|_q, \tag{4.1.3}$$

for $q > 0$.

Corollary. Suppose $\{X^{(n)}\}_{n\in\mathbb{N}}$, $\{s_n\}_{n\in\mathbb{N}}$ and $I_{AX,k}$ are as in Theorem 4.1.3. Then $\{n^{1/q-1/2}Y_{q,k}^n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF

$$\mathbb{J}_{Y_{q,k}}(x) := \inf_{z\in\mathbb{R}^k}\left\{I_{AX,k}(z) : x = \|z\|_q\right\}.$$

Proof. The result follows immediately by applying the contraction principle to the LDP from Theorem 4.1.3 and the continuous mapping $x \mapsto \|x\|_q$. □

4.2 LDP in the sublinear regime

The central idea is to consider the empirical measure of the projection, in the case of non-constant dimension. We recall the definitions of $\hat{\mu}_{\mathbf{A}}^n$ and L^n in 4.0.1. Before we start with the theory in this subsection, we think a little bit about the quantity L^n.

What can we expect to happen? So, first of all, L^n is an empirical measure consisting of k_n summands. Thus, it is plausible that Sanov's theorem will play a major role and that the relative entropy H will appear. Given the form of L^n one would expect that the „Sanov type" LDP has speed k_n. On the other hand L^n contains our sequence $\{X^{(n)}\}_{n\in\mathbb{N}}$, where we assume that e.g. $\{\|X^{(n)}\|/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP at rate s_n with GRF J_X. In total, we would expect that L^n somehow combines those two LDPs to an LDP at rate $s_n \wedge k_n$ and some rate function depending on the respective GRFs H and J_X. It is not surprising that, depending on the relation of s_n and k_n, sometimes the LDP with GRF J_X will dominate (and the LDP with GRF H will degenerate) and sometimes the other way around. In the case of $s_n = k_n$ both of those LDPs will equally influence the behaviour of L^n.

In the following we denote by γ_σ the normal distribution with mean 0 and variance σ^2.

Lemma 4.2.1. *(Proposition 4.6 in [18])*
Fix $q < 2$. Suppose $\{k_n\}_{n\in\mathbb{N}}$ grows sublinearly. Then, $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ satisfies an LDP in $P_q(\mathbb{R})$ at speed k_n with GRF $H(\cdot|\gamma_1) : P_q(\mathbb{R}) \to [0,\infty]$.

Proof. Previous results have shown that $A_{n,k_n}(1,.) \overset{(d)}{=} (\zeta_1, ..., \zeta_{k_n})/\|\zeta^{(n)}\|_2$ for iid standard normal ζ_j. This implies

$$\hat{\mu}^n_{\mathbf{A}} = \frac{1}{k_n}\sum_{j=1}^{k_n} \delta_{\sqrt{n}A_{n,k_n}(1,j)} \overset{(d)}{=} \frac{1}{k_n}\sum_{j=1}^{k_n} \delta_{\sqrt{n}\zeta_j/\|\zeta^{(n)}\|_2} =: \tilde{\nu}_n.$$

If we are able to show the asymptotic exponential equivalence of $\{\tilde{\nu}_n\}_{n\in\mathbb{N}}$ and

$$\nu_n = \frac{1}{k_n}\sum_{j=1}^{k_n} \delta_{\zeta_j}, \qquad n \in \mathbb{N}$$

at speed k_n, we can then use Lemma 4.0.8 to consequently establish an LDP for $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$.
Let d_{bL} be the bounded Lipschitz metric, which induces weak convergence on $P(\mathbb{R})$, let $BL(\mathbb{R})$ denote the space of Lipschitz continuous functions mapping from $\mathbb{R}$ to $\mathbb{R}$

with constant 1 and define $z_n := \sqrt{n}/\|\zeta^{(n)}\|_2$. Then

$$
\begin{aligned}
d_{bL}(\nu_n,\tilde{\nu}_n) &= \sup_{f\in BL(\mathbb{R})}\left|\int_{\mathbb{R}} f(x)d\nu_n(x) - \int_{\mathbb{R}} f(x)d\tilde{\nu}_n(x)\right| \\
&\le \sup_{f\in BL(\mathbb{R})}\frac{1}{k_n}\sum_{j=1}^{k_n}\left|f(\zeta_j)-f(z_n\zeta_j)\right| \le \frac{1}{k_n}\left|1-z_n\right|\sum_{j=1}^{k_n}\left|\zeta_j\right|.
\end{aligned}
$$

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

$$
\begin{aligned}
\mathbb{P}\left[d_{bL}(\nu_n,\tilde{\nu}_n)>\delta\right] &\le \mathbb{P}\left[\left|1-z_n\right|\frac{1}{k_n}\sum_{j=1}^{k_n}\left|\zeta_j\right|>\delta\right] \\
&= \mathbb{P}\left[\left|1-z_n\right|\frac{1}{k_n}\sum_{j=1}^{k_n}\left|\zeta_j\right|>\delta, \left|1-z_n\right|>\epsilon\right] + \mathbb{P}\left[\left|1-z_n\right|\frac{1}{k_n}\sum_{j=1}^{k_n}\left|\zeta_j\right|>\delta, \left|1-z_n\right|\le\epsilon\right] \\
&\le \mathbb{P}\left[\left|1-z_n\right|>\epsilon\right] + \mathbb{P}\left[\frac{1}{k_n}\sum_{j=1}^{k_n}\left|\zeta_j\right|>\frac{\delta}{\epsilon}\right] =: b_n + a_n.
\end{aligned}
$$

As shown in Application 2, we have

$$
\limsup_{n\to\infty}\frac{1}{k_n}\log\mathbb{P}\left[\frac{1}{k_n}\sum_{j=1}^{k_n}\left|\zeta_j\right|>\frac{\delta}{\epsilon}\right] \le -\Lambda^*(\delta/\epsilon)
$$

for a convex and superlinear growing function Λ^*. Using Application 6 and the contraction principle (similar to the proof of Theorem 4.1.3) implies that $\{z_n\}_{n\in\mathbb{N}}$ satisfies an LDP with speed n. Since $n \gg k_n$, we have that

$$
\lim_{n\to\infty}\frac{b_n}{a_n} = 0.
$$

Thus

$$
\begin{aligned}
\limsup_{n\to\infty}\frac{1}{k_n}\log\mathbb{P}\left[d_{bL}(\nu_n,\tilde{\nu}_n)>\delta\right] &\le \limsup_{n\to\infty}\frac{1}{k_n}\log(a_n+b_n) \\
&\le \limsup_{n\to\infty}\frac{1}{k_n}\log a_n + \limsup_{n\to\infty}\frac{1}{k_n}\log\left(1+\frac{b_n}{a_n}\right) \\
&\le -\Lambda^*(\delta/\epsilon) \xrightarrow{\epsilon\to 0} -\infty.
\end{aligned}
$$

Hence, $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ satisfies the LDP in $P(\mathbb{R})$ at speed k_n, equipped with the weak topology. Now we can use Corollary 3.1.2, derived from the inverse contraction principle, to establish the LDP for the q-Wasserstein topology. For this we need to show the exponential tightness of $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$. Recall the set $K_{2,j}$ from Lemma 3.3.5 and define $j_n := \lfloor n/k_n \rfloor + 1$. Then

$$\mathbb{P}\left[\hat{\mu}^n_{\mathbf{A}} \in K_{2,j_n}\right] = \mathbb{P}\left[\int_{\mathbb{R}} x^2 d\hat{\mu}^n_{\mathbf{A}}(x) \le j_n\right] = \mathbb{P}\left[\frac{1}{k_n}\sum_{j=1}^{k_n}\int_{\mathbb{R}} x^2 d\delta_{\sqrt{n}A_{n,k_n}(1,j)}(x) \le j_n\right]$$
$$= \mathbb{P}\left[\frac{n}{k_n}||A_{n,k_n}(1,.)||_2^2 \le j_n\right] = 1.$$

In the last step we used that $A_{n,k_n}(1,\cdot) \overset{(d)}{=} (\zeta_1, ...\zeta_{k_n})/||(\zeta_1,...,\zeta_n)||_2$ for iid standard normal ζ_j and hence $||A_{n,k_n}||_2 \le 1$. K_{2,j_n} is compact with respect to the q-Wasserstein topology for $q < 2$. Thus, $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ is exponentially tight, since for $\alpha > 0$

$$\limsup_{n\to\infty} \frac{1}{k_n} \log \mathbb{P}\left[\hat{\mu}^n_{\mathbf{A}} \in K^c_{2,j_n}\right] = -\infty < -\alpha.$$

Hence, we can apply the corollary to establish an LDP on $P_q(\mathbb{R})$ for $q < 2$ with GRF $H(\cdot|\gamma_1)$. □

Now we are ready for the the next theorem, where the LDP is shown for the sequence of empirical measures $\{L^n\}_{n\in\mathbb{N}}$.

Theorem 4.2.2. *(LDP in the sublinear case, Theorem 2.8 in [18])*

Suppose $\{k_n\}_{n\in\mathbb{N}}$ grows sublinearly and Asumption A holds with associated speed s_n and GRF J_X. Also, suppose that J_X has a unique minimum at $m > 0$. Let H be the relative entropy functional. Then, for every $q < 2$,

1. *If $s_n \gg k_n$, $\{L^n\}_{n\in\mathbb{N}}$ satisfies an LDP in $P_q(\mathbb{R})$ at speed k_n, with GRF $\mathbb{I}_{L,k_n} : P_q(\mathbb{R}) \to [0,\infty]$, defined by*
$$\mathbb{I}_{L,k_n}(\mu) := H(\mu|\gamma_1).$$

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

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

Proof. We will prove each claim separately using different previous results. We start with 1. where $s_n \gg k_n$:
By Proposition 4.2.1 we have that $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ satisfies an LDP with speed k_n and GRF $H(\cdot|\gamma_1)$ and by Assumption A we know that $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP with speed s_n and some GRF J_X. Using Remark 3.1.2 gives us an LDP for $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ with speed $k_n \ll s_n$ and GRF χ_m. By Lemma 4.0.7 we get that $\{L^n\}_{n\in\mathbb{N}}$ satisfies an LDP with GRF

$$\begin{aligned}
\mathbb{I}_{L,k_n}(\mu) &= \inf_{\nu\in P_q(\mathbb{R}),c>0} \left\{H(\nu|\gamma_1) + \chi_m(c) : \mu(\cdot) = \nu(\cdot \times c^{-1})\right\} \\
&= H(\mu(\cdot \times m)|\gamma_1) \\
&= H(\mu|\gamma_1(\cdot \times m^{-1})) \\
&= H(\mu|\gamma_m),
\end{aligned}$$

where the second last equality is shown in Lemma 3.1.18.
For 2., where $k_n = s_n$, we again use Lemma 4.2.1 and the LDP for $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ from Assumption A, while due to $k_n = s_n$, the rate function does not degenerate. Using Lemma 4.0.7 gives us an LDP for $\{L^n\}_{n\in\mathbb{N}}$ at speed s_n and GRF

$$\begin{aligned}
\mathbb{I}_{L,k_n}(\mu) &= \inf_{\nu\in P_q(\mathbb{R}),c>0} \left\{H(\nu|\gamma_1) + J_X(c) : \mu(\cdot) = \nu(\cdot \times c^{-1})\right\} \\
&= \inf_{c>0} \left\{H(\mu|\gamma_c) + J_X(c)\right\},
\end{aligned}$$

where the transformation in the infimum is similar to the previous case.
For 3., where $s_n \ll k_n$, we basically make the same observations as in 1., but in this case the LDP for $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ degenerates to an LDP with speed s_n and GRF χ_{γ_1}. We deduce an LDP for $\{L^n\}_{n\in\mathbb{N}}$ with speed s_n and GRF

$$\begin{aligned}
\mathbb{I}_{L,k_n}(\mu) &= \inf_{\nu\in P_q(\mathbb{R}),c>0} \left\{\chi_{\gamma_1}(\nu) + J_X(c) : \mu(\cdot) = \nu(\cdot \times c^{-1})\right\} \\
&= \inf_{c>0} \left\{\chi_{\gamma_1}(\mu(\cdot \times c)) + J_X(c)\right\} \\
&= \begin{cases} J_X(y), & \text{if } \exists y > 0 : \mu = \gamma_y \\ \infty, & \text{otherwise.} \end{cases}
\end{aligned}$$

□

The following results treat the asymptotic behaviour of scaled q-norms. We therefore recall the sequence

$$Y^n_{q,k_n} = n^{-1/q} \|A^T_{n,k_n} X^{(n)}\|_q$$

and we will need the following lemma.

Lemma 4.2.3. *(Lemma 4.2 in [1])*

Let $\zeta_1, \zeta_2, \ldots$ be an iid sequence of standard normal distributed random variables and consider

$$V_n := \frac{\|(\zeta_1, \ldots, \zeta_{k_n})\|_2}{\|(\zeta_1, \ldots, \zeta_n)\|_2}.$$

Then the sequence of random variables $\{V_n\}_{n\in\mathbb{N}}$ satisfies an LDP with speed n and rate function

$$\mathbb{I}_V(y) := \begin{cases} -\frac{1}{2}\log(1-y^2), & \text{if } y \in [0,1) \\ \infty, & \text{otherwise} . \end{cases}$$

Proof. The proof is not done as in [1], we provide a more elementary approach.
We can use the representation

$$V_n = F\Big(\frac{1}{n}\|(\zeta_1, \ldots, \zeta_{k_n})\|_2^2, \frac{1}{n}\|(\zeta_{k_n+1}, \ldots, \zeta_n)\|_2^2\Big), \tag{4.2.1}$$

where $F : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+, (x_1, x_2) \mapsto \sqrt{\frac{x_1}{x_1+x_2}}$. The sequences on the right-hand side of Equation 4.2.1 are independent and satisfy respective LDPs in $\mathbb{R}$ at speed n and GRFs given in Application 6 and Application 7 (To be fully precise the LDP for $\{\|(\zeta_{k_n+1}, \ldots, \zeta_n)\|_2^2/n\}_{n\in\mathbb{N}}$ can be shown analogously as in 7). For $y \le 1$ the GRF for the sequence in Equation 4.2.1 is then given by

$$\begin{aligned} \mathbb{I}_V(y) &= \inf\Big\{\frac{x_1}{2} + \frac{1}{2}\big(x_2 - \log x_2 - 1\big) : y = \sqrt{\frac{x_1}{x_1+x_2}}\Big\} \\ &= -\frac{1}{2}\log(1-y^2). \end{aligned}$$

The calculation in the infimum is essentially the same as in Lemma 4.1.1. □

The next definition introduces the p-generalized Gaussian distribution, a natural generalization of the normal distribution.

Definition 4.2.1. *For $p \in [1, \infty)$, let f_p be the density of the p-generalized normal distribution, i.e.*

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

where Γ denotes the Gamma function.

We can provide an LDP including such random variables, which will occur in the proof of the next theorem.

Lemma 4.2.4. *(Lemma 3.2 in [18])*
Let $p \in [1,2]$, let $\{k_n\}_{n\in\mathbb{N}}$ satisfy $k_n \to \infty$ as $n \to \infty$. Then, given $\{s_n\}_{n\in\mathbb{N}}$ such that $k_n/s_n \to 0$ as $n \to \infty$, the sequence $\{(s_n)^{-1}\sum_{j=1}^{k_n}(\xi_i^{(p)})^2\}_{n\in\mathbb{N}}$ satisfies an LDP with speed $s_n^{p/2}$ and GRF $\mathfrak{I}_{\xi,p} : \mathbb{R} \to [0,\infty]$ defined by

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

We can now state the next theorem treating the asymptotical behaviour of scaled 2-norms.

Theorem 4.2.5. *(Theorem 2.10 in [18])*
Suppose k_n grows sublinearly.
1. *If Assumption A* holds with GRF J_X, then $\{Y_{2,k_n}^n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ at speed n with GRF $\mathbb{J}_{Y_{2,k_n}} : \mathbb{R} \to [0,\infty]$, defined by*

$$\mathbb{J}_{Y_{2,k_n}}(x) := \inf_{c\in[0,1)} \left\{ -\frac{1}{2}\log(1-c^2) + J_X\left(\frac{x}{c}\right)\right\}.$$

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

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

Proof. We recall some of the previous results in order to establish a different representation of Y_{2,k_n}^n. By Lemma 4.0.4 and Lemma 4.0.5 we know

$$A_{n,k_n}^T \frac{X^{(n)}}{||X^{(n)}||_2} \overset{(d)}{=} A_{n,k_n}(1,\cdot) \overset{(d)}{=} \frac{\zeta^{(k_n)}}{||\zeta^{(n)}||_2}$$

for an iid standard normal distributed sequence $\zeta_1, \zeta_2, \ldots$ and $\zeta^{(k)} = (\zeta_1, \ldots, \zeta_k)$. Combining these identities and the fact that $A^T_{n,k_n} \frac{X^{(n)}}{\|X^{(n)}\|_2}$ is independent of $X^{(n)}$ leads to

$$Y^n_{2,k_n} = \frac{\|A^T_{n,k_n} \frac{X^{(n)}}{\|X^{(n)}\|_2}\|_2 \|X^{(n)}\|_2}{\sqrt{n}} \overset{(d)}{=} \frac{\|\zeta^{(k_n)}\|_2}{\|\zeta^{(n)}\|_2} \frac{\|X^{(n)}\|_2}{\sqrt{n}} = \frac{\|\zeta^{(k_n)}\|_2/\sqrt{k_n}}{\|\zeta^{(n)}\|_2/\sqrt{n}} \frac{\sqrt{k_n}\|X^{(n)}\|_2}{n}. \quad (4.2.2)$$

Situation 1: We assume Assumption A*. Then the claim follows immediately by Lemma 4.2.3 and the contraction principle with continuous function $F : \mathbb{R}^2 \to \mathbb{R}, (\zeta, x) \mapsto \zeta x$. We obtain the GRF $\mathbb{J}_{Y_{2,k_n}} : \mathbb{R}_+ \to [0, \infty]$ given by

$$\mathbb{J}_{Y_{2,k_n}}(x) = \inf_{c \in [0,1), x = cy} \left\{ -\frac{1}{2}\log(1-c^2) + J_X(y) \right\} = \inf_{c \in [0,1)} \left\{ -\frac{1}{2}\log(1-c^2) + J_X\left(\frac{x}{c}\right) \right\}.$$

Situation 2: We assume Assumption C. Then we use the representation in Lemma 4.2.3 and Lemma 4.1.2 to derive LDPs in three different cases for the parameter r from Assumption C. But first we consider the sequences $\{\|\zeta^{(n)}\|_2/n\}_{n\in\mathbb{N}}$ and $\{\|\zeta^{(k_n)}\|_2/k_n\}_{n\in\mathbb{N}}$. Using Application 5 and the contraction principle with the continuous function $x \mapsto \sqrt{x}$ allows to derive LDPs with speed n and k_n respectively. The corresponding GRF $\mathbb{I} : \mathbb{R}_+ \to [0, \infty]$ follows directly and has a closed representation

$$\mathbb{I}(x) = \inf_{c = x^2} \left[\frac{1}{2}\big(c - \log c - 1\big) \right] = \frac{x^2 - 1}{2} - \log x.$$

Case 1: $r = 0$, implying $s_n \ll k_n \ll n$. We want to apply Lemma 4.1.2 on the sequence from Equation 4.2.2, hence we define $U_n := \sqrt{k_n}\|X^{(n)}\|_2/n$, $V_n := \sqrt{k_n}/\|\zeta^{(k_n)}\|_2$ and $W_n := \|\zeta^{(n)}\|_2/n$. $\{U_n\}_{n\in\mathbb{N}}$ satisfies an LDP with speed s_n and GRF $J^{(0)}_X$ by assumption. As shown before, $\{V_n\}_{n\in\mathbb{N}}$ and $\{W_n\}_{n\in\mathbb{N}}$ satisfy LDPs with speed k_n and n respectively. Thus $\{V_n\}_{n\in\mathbb{N}}$ satisfies an LDP with speed s_n and degenerated GRF χ_1, where $\{W_n\}_{n\in\mathbb{N}}$ satisfies an LDP with speed k_n and the same GRF. We are therefore in the required setting and deduce that $\{(U_n, V_n, W_n)\}_{n\in\mathbb{N}}$ satisfies an LDP with speed s_n and GRF $J : \mathbb{R}^3 \to [0, \infty]$ given by

$$J(u, v, w) = \begin{cases} J^{(0)}_X(u) + \chi_1(v) & \text{, if w=m} \\ \infty & \text{else.} \end{cases}$$

We can use the contraction principle for the continuous mapping $(u, v, w) \mapsto uvw$ to derive an LDP for the sequence in Equation 4.2.2 with speed s_n and GRF $\mathbb{J}_{Y_{2,k_n}} : \mathbb{R}_+ \to [0, \infty]$ given by

$$\mathbb{J}_{Y_{2,k_n}}(x) = \inf_{x = uv} \left[J^{(0)}_X(u) + \chi_1(v) \right] = J^{(0)}_X(x).$$

Case 2: $r \in (0, \infty)$. As we mentioned in Remark 4.0.3, we get an LDP with speed k_n and GRF $rJ_X^{(r)}(\sqrt{r}x)$. We again work with the representation from Equation 4.2.2 and use Lemma 4.1.2 to derive an LDP with GRF $\mathbb{J}_{Y_{2,k_n}} : \mathbb{R}_+ \to [0, \infty]$

$$\mathbb{J}_{Y_{2,k_n}}(x) = \inf_{x=u/v} \left[\frac{u^2-1}{2} - \log u + rJ_X^{(r)}(\sqrt{r}v) \right] = \inf_{u>0} \left[\frac{u^2-1}{2} - \log u + rJ_X^{(r)}\left(\frac{\sqrt{r}u}{x} \right) \right].$$

Case 3: $r = \infty$. The same argument as done to receive Equation 4.2.2 leads to the representation

$$Y_{2,k_n} \overset{(d)}{=} \frac{\|\zeta^{(k_n)}\|_2/\sqrt{s_n}}{\|\zeta^{(n)}\|_2/\sqrt{n}} \frac{\sqrt{s_n}\|X^{(n)}\|_2}{n}.$$

We can use Lemma 4.2.4 and the contraction principle for the continuous mapping $t \mapsto \sqrt{t}$ to get an LDP for $\{\|\zeta^{(k_n)}\|_2/\sqrt{s_n}\}_{n\in\mathbb{N}}$ with speed s_n and GRF $\mathbb{I} : \mathbb{R}_+ \to [0, \infty]$ given by

$$\mathbb{I}(x) = \inf_{y=x^2} \left[\frac{y}{2} \right] = \frac{x^2}{2}.$$

Now we are in the situation of Lemma 4.1.2 with two sequences satisfying an LDP at speed s_n and one sequence with speed n, where $s_n \ll n$ by assumption. Further, $\|\zeta^{(n)}\|_2/\sqrt{n} \to 1$ almost surely and hence get the GRF $\mathbb{J}_{Y_{2,k_n}} : \mathbb{R}_+ \to [0, \infty]$, where

$$\mathbb{J}_{Y_{2,k_n}}(x) = \inf_{x=uv/w, u>0} \left[\frac{u^2}{2} + J_X^{(\infty)}(v) + \chi_1(w) \right] = \inf_{u>0} \left[\frac{u^2}{2} + J_X^{(\infty)}\left(\frac{x}{u} \right) \right],$$

which concludes the proof.

In the last section of this chapter we want to find an LDP for q-norms with a different scaling. A simple rearrangement, which we will see in a moment, suggests that the scaling $k_n^{-1/q}$ is meaningful. We therefore consider

$$\tilde{Y}_{q,k_n}^n(x) := k_n^{-1/q} \|A_{n,k_n}^T X^{(n)}\|_q. \tag{4.2.3}$$

Similar to Representation 4.2.2 in the previous theorem, we use Lemma 4.0.4 and

Lemma 4.0.5 to derive a new quantity, namely

$$\begin{aligned} k_n^{-1/q}\|A_{n,k_n}^T X^{(n)}\|_q &= k_n^{-1/q}\left\|A_{n,k_n}^T \frac{X^{(n)}}{\|X^{(n)}\|_2}\right\|_q \|X^{(n)}\|_2 \\ &\overset{(d)}{=} k_n^{-1/q}\|A_{n,k_n}(1,.)\|_q \|X^{(n)}\|_2 \\ &\overset{(d)}{=} k_n^{-1/q}\frac{\|\zeta^{(k_n)}\|_q}{\|\zeta^{(n)}\|_2}\|X^{(n)}\|_2 \\ &= \frac{k_n^{-1/q}\|\zeta^{(k_n)}\|_q}{\|\zeta^{(n)}\|_2 n^{-1/2}} n^{-1/2}\|X^{(n)}\|_2. \end{aligned}$$

As it turns out, $k_n^{-1/q}$ is just the right scaling to get an interesting LDP for $\{\|\zeta^{(k_n)}\|_q\}_{n\in\mathbb{N}}$. For this we introduce two quantities, we start with the cumulant generating function of $|X|^q$, where $X \sim \mathcal{N}(0,1)$, i.e.

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

for $q \in [1,2)$ and $t \in \mathbb{R}$ or for $q = 2$ and $t < 1/2$. We will also encounter the q-th moment of $|X|$, we hence define

$$M_q := \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}|x|^q \exp\left(-\frac{1}{2}x^2\right)dx = \frac{2^{q/2}}{\sqrt{\pi}}\Gamma\left(\frac{q+1}{2}\right). \tag{4.2.4}$$

Theorem 4.2.6. *(Theorem 2.11 in [18])*
Fix $q \in [1,2]$, suppose $1 \ll k_n \ll n$ and Assumption A holds with speed s_n and GRF J_X, which additionally has a unique minimum at $m > 0$. Then $\{\tilde{Y}^n_{q,k_n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $s_n \wedge k_n$ with rate function $\mathbb{J}_{\tilde{Y}_{2,k_n}} : \mathbb{R} \to [0,\infty]$ defined by, for $x \geq 0$

$$\mathbb{J}_{\tilde{Y}_{2,k_n}}(x) := \begin{cases} \Lambda^*(x^q/m^q), & \text{if } s_n \gg k_n, \\ \inf_{c>0}\left\{\Lambda^*(c^q) + J_X(x/c)\right\}, & \text{if } s_n = k_n, \\ J_X(x/M_q^{1/q}), & \text{if } s_n \ll k_n. \end{cases}$$

Proof. We will work with Representation 4.2.3, thus we consider the sequence of vectors

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

By Cramér's theorem we can easily establish an LDP for

$$\frac{|\zeta_1|^q + \ldots + |\zeta_n|^q}{n}$$

with speed n and rate function Λ_q^*. Hence, $\{\|\zeta^{(k_n)}\|_q^q / k_n\}_{n\in\mathbb{N}}$ satisfies an LDP with speed k_n and the same rate function. We can apply the contraction principle to get an LDP for $\{f(\|\zeta^{(k_n)}\|_q^q/k_n)\}_{n\in\mathbb{N}}$ with $f : \mathbb{R}_+ \to \mathbb{R}_+, x \mapsto x^{1/q}$. As rate function we get $J_{q,\zeta}(x) := \Lambda_q^*(x^q)$, $x \geq 0$. Now we distinguish three cases, depending on the speed s_n relative to k_n:

Case 1: $k_n \ll s_n$. We apply Lemma 4.1.2 and observe that $\{R_n\}_{n\in\mathbb{N}}$ is exponentially equivalent with speed k_n to $\{(\|\zeta^{(k_n)}\|_q/k_n^{1/q}, 1, m)\}_{n\in\mathbb{N}}$. $\{\tilde{Y}_{2,k_n}\}_{n\in\mathbb{N}}$ hence satisfies an LDP at speed k_n and rate function $J_{\tilde{Y}_{2,k_n}} : \mathbb{R}_+ \to [0, \infty]$ given by

$$J_{\tilde{Y}_{2,k_n}}(x) = \inf_{x=uw/v} \Big\{ J_{\zeta,q}(u) + \chi_1(v) + \chi_m(w) \Big\} = J_{\zeta,q}(x/m) = \Lambda_q^*((x/m)^q).$$

Case 2: $k_n = s_n$. Again Lemma 4.1.2 provides an LDP for $\{R_n\}_{n\in\mathbb{N}}$ with speed k_n. As rate function we get $J_{\tilde{Y}_{2,k_n}} : \mathbb{R}_+ \to [0, \infty]$

$$J_{\tilde{Y}_{2,k_n}}(x) = \inf_{x=uv/w} \Big\{ J_{\zeta,q}(u) + \chi_1(v) + J_X(w) \Big\} = \inf_{u>0} \Big\{ J_{\zeta,q}(u) + J_X(x/u) \Big\}.$$

Case 3: $s_n \ll k_n$. By the strong law of large numbers, we have the almost sure convergence

$$\frac{\|\zeta^{(k_n)}\|_q^q}{k_n} \longrightarrow M_q,$$

and hence $\frac{\|\zeta^{(k_n)}\|_q}{k_n^{1/q}} \to M_q^{1/q}$ almost surely. By Lemma 4.1.2, $\{R_n\}_{n\in\mathbb{N}}$ is exponentially equivalent to $\{(M_q^{1/q}, 1, \|X^{(n)}\|_2/\sqrt{n})\}_{n\in\mathbb{N}}$ at speed s_n, the resulting rate function $\mathbb{J}_{\tilde{Y}_{2,k_n}} : \mathbb{R}_+ \to [0, \infty]$ for $\tilde{Y}_{2,k_n}$ is therefore

$$\mathbb{J}_{\tilde{Y}_{2,k_n}}(x) = \inf_{x=uw/v} \Big\{ \chi_{M_q^{1/q}}(u) + J_X(v) + \chi_1(v) \Big\} = J_X(x/M_q^{1/q}).$$

□

4.3 LDP in the linear regime

In this section we consider the situation when k_n grows linearly with rate $\lambda \in (0,1]$. As in the sublinear case, we establish an LDP for the sequence of empirical measures $\{L^{(n)}\}_{n\in\mathbb{N}}$, as well as an LDP for the sequence of properly scaled norms. We start by analysing the behaviour of $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$. Therefore, we need a few auxiliary results starting with an application of the approximate contraction principle.

Lemma 4.3.1. *(Corollary A.2. in [18])*
Let Σ be a Polish space and $\mathfrak{X}$ be a separable Banach space. Suppose that $\{k_n\}_{n\in\mathbb{N}}$ grows linearly with rate $\lambda \in (0,1]$, each $\mathscr{L}_n$ is the empirical measure of k_n iid Σ-valued random variables $\eta_1, ..., \eta_{k_n}$ with common distribution μ (that does not depend on n), and for continuous $W : \Sigma \to \mathbb{R}$, define

$$\hat{\Lambda}(W) := \log \mathbb{E}\left[e^{\lambda^{-1}W(\eta_1)}\right].$$

Also let $c : \Sigma \to \mathfrak{X}$ be a continuous map such that 0 lies in the interior $\mathscr{D}^o$ of the set

$$\mathscr{D} := \{\alpha \in \mathfrak{X}^* : \hat{\Lambda}(\langle \alpha, c(\cdot)\rangle) < \infty\},$$

and let $\mathscr{C}_n := \int_\Sigma c(x) d\mathscr{L}_n(x)$. Then $\{(\mathscr{L}_n, \mathscr{C}_n)\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n and GRF $\mathbb{I} : P(\mathbb{R}) \times \mathbb{R} \to [0,\infty]$ given by

$$\mathbb{I}(\mu, s) := \begin{cases} \mathbb{I}_0(\mu) + F(s - \int_\Sigma c(x) d\mu(x)), & \text{if } \mathbb{I}_0(\mu) < \infty \\ \infty & \text{else,} \end{cases}$$

where $\mathbb{I}_0(\nu) = \lambda H(\nu|\mu)$ and $F(x) = \sup_{\alpha\in\mathscr{D}^o} \langle \alpha, x\rangle$.

Proof. We want to apply Theorem 3.1.9, therefore we need to check the assumptions. At first we will only work with $\hat{\Lambda}$, later we will establish a relation between $\bar{\Lambda}$ from Theorem 3.1.9 and $\hat{\Lambda}$. Let us start with the sequence $\mathscr{L}_n = \frac{1}{k_n}\sum_{j=1}^{k_n} \delta_{\eta_j}, n \in \mathbb{N}$ which satisfies an LDP at speed k_n and GRF $H(\cdot|\mu)$ by Sanov's theorem 3.1.16 on the space $P(\Sigma)$. Since $k_n/n \to \lambda$, one can directly deduce an LDP with speed n and GRF $\mathbb{I}(\cdot) = \lambda H(\cdot|\mu)$. In order to establish an LDP for

$$(\mathscr{L}_n, \mathscr{C}_n) = \frac{1}{k_n}\sum_{i=1}^{k_n} (\delta_{\eta_j}, c(\eta_j)), \qquad n \in \mathbb{N}, \tag{4.3.1}$$

we recall the general version of Cramér's theorem 3.1.12 and Remark 3.1.3 and note that the quantity from Equation 4.3.1 is the empirical mean of iid random variables.

By assumption, 0 lies in the interior of $\mathcal{D}$ and hence Corollary 3.1.3 applies and we get an LDP for the random objects in 4.3.1 on the space $\mathfrak{X} \times P(\Sigma)$ at speed k_n with convex GRF. Again, since $k_n/n \to \lambda$, $\{(\mathscr{L}_n, \mathscr{C}_n)\}_{n\in\mathbb{N}}$ of 4.3.1 satisfies an LDP with speed n. The fact that $0 \in \mathcal{D}^o$ implies the first assumption from Theorem 3.1.9. For Assumption 3 we take a sequence $\{W_n\}_{n\in\mathbb{N}}$ with $W_n := V_n + \langle \alpha_n, c(\cdot)\rangle$, where $\alpha_n \in \mathcal{D}^o$ and V_n is continuous and bounded. By construction we have that $\alpha_1 \in \mathcal{D}^o$ as well as $W_n \downarrow W_\infty$. Since $W_1 \in \mathcal{D}^o$, we can apply dominated convergence to get

$$\lim_{n\to\infty} \hat{\Lambda}(W_n) = \hat{\Lambda}(W_\infty).$$

We thus have verified all conditions of Theorem 3.1.9, after showing that $\bar{\Lambda} = \lambda\hat{\Lambda}$. We show both inequalities separately, consider

$$\begin{aligned}
\lambda^{-1}\bar{\Lambda}(W) &\geq \lambda^{-1} \lim_{R\to\infty} \bar{\Lambda}(W \wedge R) \\
&= \lambda^{-1} \lim_{R\to\infty} \sup_{r>0} \Lambda_r(W \wedge R) \\
&= \lambda^{-1} \lim_{R\to\infty} \Lambda_\infty(W \wedge R) \\
&= \lambda^{-1} \lim_{R\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\big[e^{n\int_\Sigma W(x)\wedge R d\mathscr{L}_n(x)}\big] \\
&= \lambda^{-1} \lim_{R\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}\big[\prod_{j=1}^{k_n} e^{n/k_n (W(\eta_j)\wedge R)}\big] \\
&= \lambda^{-1} \lim_{R\to\infty} \limsup_{n\to\infty} \frac{k_n}{n} \log \mathbb{E}\big[e^{n/k_n (W(\eta_1)\wedge R)}\big] \\
&= \hat{\Lambda}(W).
\end{aligned}$$

The other inequality follows directly from the trivial relation $\Lambda_\infty \geq \bar{\Lambda}$. Then

$$\begin{aligned}
\hat{\Lambda}(W) &= \lambda^{-1}\Lambda_\infty(W) \\
&\geq \lambda^{-1}\bar{\Lambda}(W).
\end{aligned}$$

It thus follows that the set $\mathcal{D}$ from the assumption and the corresponding set from Theorem 3.1.9 coincide. □

Lemma 4.3.2. *(Lemma 4.8 in [18])*

Given the iid sequence $\{\zeta_j\}_{j\in\mathbb{N}}$ with common law γ_1. Then the sequence

$$\left(\frac{1}{k_n}\sum_{j=1}^{k_n}\delta_{\zeta_j}, \frac{1}{n-k_n}\sum_{j=k_n+1}^{n}\delta_{\zeta_j}, \frac{1}{n}\sum_{j=1}^{n}\zeta_j^2\right), \quad n\in\mathbb{N} \tag{4.3.2}$$

satisfies an LDP in $[P(\mathbb{R})^2\times\mathbb{R}_+]$ at speed n with GRF $\mathbb{I}_1$ defined by

$$\mathbb{I}_1(\mu,\nu,s) := \begin{cases} \lambda H(\mu|\gamma_1) + (1-\lambda)H(\nu|\gamma_1) + \frac{1}{2}[s-\lambda m_2(\mu) - (1-\lambda)m_2(\nu)], \\ \text{if } \lambda m_2(\mu) + (1-\lambda)m_2(\nu) \le s, \\ \infty \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise.} \end{cases}$$

Proof. We want to apply Lemma 4.3.1 for the specific situation given. In the introduced notation we have

- $\Sigma = \mathbb{R}$
- $\mathfrak{X} = \mathbb{R} = \mathfrak{X}^*$
- $c:\mathbb{R}\to\mathbb{R}, x\mapsto x^2$
- For $n\in\mathbb{N}$, let $\mathscr{L}_n^{(1)} = \frac{1}{k_n}\sum_{j=1}^{k_n}\delta_{\zeta_j}$ and $\mathscr{L}_n^{(2)} = \frac{1}{n-k_n}\sum_{j=k_n+1}^{n}\delta_{\zeta_j}$
- Let $\mathscr{C}_n^{(i)} = \int_{\mathbb{R}} c(x)d\mathscr{L}_n^{(i)}(x)$ for $i=1,2$.

We start with the sequence $\{(\mathscr{L}_n^{(1)},\mathscr{C}_n^{(1)})\}_{n\in\mathbb{N}} \subseteq P(\mathbb{R})\times\mathbb{R}$, where we can calculate the domain of $\hat{\Lambda}$ explicitly, namely

$$\mathscr{D} = \left\{\alpha\in\mathbb{R} \mid \log\mathbb{E}\left[e^{\lambda^{-1}\alpha\zeta_1^2}\right] < \infty\right\} = \left(-\infty, \frac{\lambda}{2}\right).$$

We used that the domain of the moment generating function of $\zeta_1^2 \sim \chi_1^2$ equals $\left(-\infty,\frac{1}{2}\right)$. We can derive that $\mathscr{D} = \mathscr{D}^o$ as well as $0\in\mathscr{D}^o$, hence all assumptions of Lemma 4.3.1 are fulfilled. Further, we can provide a closed representation of the function $F:\mathbb{R}\to[0,\infty]$, where we have

$$F(x) = \sup_{t<\lambda/2}[tx] = \begin{cases} \frac{\lambda}{2}x & ,\text{if } x\ge 0 \\ \infty & ,\text{else }. \end{cases}$$

By Sanov's theorem 3.1.16, we get an LDP for $\{\mathscr{L}_n^{(1)}\}_{n\in\mathbb{N}}$ with GRF $\lambda H(\cdot|\gamma_1)$. This is used in Lemma 4.3.1 to establish an LDP for $\{(\mathscr{L}_n^{(1)}, \mathscr{C}_n^{(1)})\}_{n\in\mathbb{N}}$ with GRF

$$\mathbb{J}_1(\mu,t) = \begin{cases} \lambda H(\mu|\gamma_1) + \frac{\lambda}{2}\big[t - m_2(\mu)\big] & ,\text{if } m_2(\mu) \leq t \\ \infty & ,\text{else.} \end{cases}$$

For $\{(\mathscr{L}_n^{(2)}, \mathscr{C}_n^{(2)})\}_{n\in\mathbb{N}}$ we can apply the exact same argument, the only difference here is the speed $\tilde{k}_n = n - k_n$, thus the constant λ is replaced by $1-\lambda$. We hence establish an LDP with speed n and GRF

$$\mathbb{J}_2(\nu,s) = \begin{cases} (1-\lambda) H(\nu|\gamma_1) + \frac{1-\lambda}{2}\big[s - m_2(\nu)\big] & ,\text{if } m_2(\nu) \leq s \\ \infty & ,\text{else.} \end{cases}$$

Now we consider the continuous mapping

$$\begin{aligned} F : [P(\mathbb{R}) \times \mathbb{R}]^2 &\to P(\mathbb{R})^2 \times \mathbb{R} \\ (\mu,t,\nu,s) &\mapsto (\mu,\nu,\lambda t + (1-\lambda)s) \end{aligned}$$

and apply the contraction principle to the sequence $\{(\mathscr{L}_n^{(1)}, \mathscr{C}_n^{(1)}, \mathscr{L}_n^{(2)}, \mathscr{C}_n^{(2)})\}_{n\in\mathbb{N}}$ to obtain an LDP for

$$(\mathscr{L}_n^{(1)}, \mathscr{L}_n^{(2)}, \lambda\mathscr{C}_n^{(1)} + (1-\lambda)\mathscr{C}_n^{(2)}) \tag{4.3.3}$$

with speed n and GRF $I : P(\mathbb{R})^2 \times \mathbb{R}_+ \to [0,\infty]$ given by

$$\begin{aligned} I(\mu,\nu,u) &= \inf_{\lambda t + (1-\lambda)s = u} \big[\mathbb{J}_1(\mu,t) + \mathbb{J}_2(\nu,s)\big] \\ &= \lambda H(\mu|\gamma_1) + (1-\lambda)H(\nu|\gamma_1) \\ &+ \frac{1}{2} \inf_{\lambda t + (1-\lambda)s = u} \big[\lambda t + (1-\lambda)s - \lambda m_2(\mu) - (1-\lambda) m_2(\nu)\big] \\ &= \lambda H(\mu|\gamma_1) + (1-\lambda)H(\nu|\gamma_1) + \frac{1}{2}\big[u - \lambda m_2(\mu) - (1-\lambda)m_2(\nu)\big] \\ &= \mathbb{I}_1(\mu,\nu,u). \end{aligned}$$

Next we show the exponential equivalence of the sequences $\{a_n\}_{n\in\mathbb{N}}$, $\{b_n\}_{n\in\mathbb{N}}$ defined by $a_n := \frac{1}{n}\sum_{j=1}^n \zeta_j^2$ and $b_n := \frac{\lambda}{k_n}\sum_{j=1}^{k_n} \zeta_j^2 + \frac{1-\lambda}{n-k_n}\sum_{j=k_n+1}^{n} \zeta_j^2$. In order to do so, fix $\epsilon > 0$ and denote by $\{Y_k\}_{k\in\mathbb{N}}$ a sequence of independent χ_k^2 distributed random variables. Then we consider

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big[|a_n - b_n| > \epsilon\big] \\
=&\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Bigg[\bigg|\frac{Y_{k_n}}{k_n} - \frac{Y_{n-k_n}}{n-k_n}\bigg| > \epsilon\Big|\lambda - \frac{k_n}{n}\Big|^{-1}\Bigg] \\
\le&\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Bigg[\frac{Y_{k_n}}{k_n} + \frac{Y_{n-k_n}}{n-k_n} > \epsilon\Big|\lambda - \frac{k_n}{n}\Big|^{-1}\Bigg] \\
\le&\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Bigg[\frac{Y_n}{n} > \epsilon\Big|\lambda - \frac{k_n}{n}\Big|^{-1}\Bigg] \\
\le& - \inf_{x>M} \frac{1}{2}\big(x - \log x - 1\big) \overset{M\to\infty}{\longrightarrow} -\infty,
\end{aligned}$$

where the last inequality comes from the LDP established in Application 6 and holds for any big $M > 0$, since $\epsilon\big|\lambda - \frac{k_n}{n}\big|^{-1}$ exceeds any given bound for sufficiently large n. Optimizing in M shows the desired equivalence with speed n. Thus, the LDP for the random object 4.3.3 is equivalent to the LDP we are interested in. □

Before we come to the next major result, we discuss an interesting property of the relative entropy H.

Proposition 1. Denote by γ_s the normal distribution with mean 0 and variance $s > 0$, let $\mu \in P(\mathbb{R})$. Then we have the following relation

$$H(\mu|\gamma_s) = H(\mu|\gamma_1) + \frac{1}{2}\log s + \frac{1}{2}\left(\frac{1}{s} - 1\right)m_2(\mu).$$

Remark. Consider the case where the Radon-Nikodym derivative $\frac{d\mu}{d\gamma_s}$ exists. Then we have the representation

$$H(\mu|\gamma_s) = \int_{\mathbb{R}} \frac{d\mu}{d\gamma_s}(x)\log\left(\frac{d\mu}{d\gamma_s}(x)\right)d\gamma_s(x) = \int_{\mathbb{R}} \log\left(\frac{d\mu}{d\gamma_s}(x)\right)d\mu(x).$$

Proof. (Proof of Proposition 1)
The case where $\frac{d\mu}{d\gamma_s}$ does not exist is not interesting, since then the claim in Proposition 1 trivially holds. In the other case we can use the previous remark to get

$$\begin{aligned}
H(\mu|\gamma_s) = \int_{\mathbb{R}} \log\left(\frac{d\mu}{d\gamma_s}(x)\right)d\mu(x) = &\int_{\mathbb{R}} \log\left(\frac{d\mu}{d\gamma_1}(x)\right) + \log\left(\frac{d\gamma_1}{d\gamma_s}(x)\right)d\mu(x) \\
=&H(\mu|\gamma_1) + \int_{\mathbb{R}} \log\left(\frac{d\gamma_1}{d\gamma_s}(x)\right)d\mu(x).
\end{aligned}$$

Now we can calculate the Radon-Nikodym derivative $\frac{\gamma_1}{\gamma_s}$ using its definition. For any Borel set $A \subseteq \mathbb{R}$, we have

$$\begin{aligned}\gamma_1(A) &= \int_A \frac{d\gamma_1}{d\gamma_s}(x)d\gamma_s(x)\\ &= \int_A \frac{d\gamma_1}{d\gamma_s}(x)\frac{1}{\sqrt{2\pi s}}\exp\Big(-\frac{x^2}{2s}\Big)dx\\ &\overset{!}{=} \frac{1}{\sqrt{2\pi}}\int_A \exp\Big(-\frac{x^2}{2}\Big)dx.\end{aligned}$$

The third equality allows us to find a representation of $\frac{d\gamma_1}{d\gamma_s}$, namely

$$\frac{d\gamma_1}{d\gamma_s} = \sqrt{s}\exp\Big(\frac{1}{2}x^2\Big(\frac{1}{s}-1\Big)\Big).$$

Taking the logarithm and integration gives us the claimed expression. □

Lemma 4.3.3. *(Lemma 4.9 in [18])*
The sequence of pairs of measures

$$\Bigg(\frac{1}{k_n}\sum_{j=1}^{k_n}\delta_{\sqrt{n}\zeta_j/\|\zeta^{(n)}\|_2}, \frac{1}{n-k_n}\sum_{j=k_n+1}^{n}\delta_{\sqrt{n}\zeta_j/\|\zeta^{(n)}\|_2}\Bigg), \quad n \in \mathbb{N} \tag{4.3.4}$$

satisfies an LDP in $P(\mathbb{R})^2$ at speed n and GRF $\mathbb{I}_2 : P(\mathbb{R})^2 \to [0,\infty]$ defined by

$$\begin{aligned}\mathbb{I}_2(\mu,\nu) :=& \mathbb{I}_1(\mu,\nu,1)\\ =& \begin{cases}\lambda H(\mu|\gamma_1) + (1-\lambda)H(\nu|\gamma_1) + \frac{1}{2}\big(1-\lambda m_2(\mu)-(1-\lambda)m_2(\nu)\big),\\ \textit{if } \lambda m_2(\mu)+(1-\lambda)m_2(\nu) \le 1,\\ \infty, \qquad \textit{otherwise.}\end{cases}\end{aligned}$$

Proof. First we consider the mapping

$$\begin{aligned}F : P(\mathbb{R})^2 \times \mathbb{R}_+ &\longrightarrow P(\mathbb{R})^2\\ (\mu,\nu,s) &\mapsto (\mu(\cdot \times s^{1/2}), \nu(\cdot \times s^{1/2}))\end{aligned}$$

and note that it is continuous. In order to see this, we remember the similar function defined in Lemma 4.0.6, where we showed continuity using Slutsky's theorem. Here one applies essentially the same argument. We can thus use F to contract the LDP from Lemma 4.3.2 to receive an LDP for the quantity in 4.3.4 with speed n and GRF

$$\begin{aligned}(\mu,\nu)\mapsto &\inf_{\bar\mu,\bar\nu\in P(\mathbb{R}),s\in\mathbb{R}_+}\left\{\mathbb{I}_1(\bar\mu,\bar\nu,s):\mu=\bar\mu(\cdot\times s^{1/2}),\nu=\bar\nu(\cdot\times s^{1/2})\right\}\\ &=\inf_{s\in\mathbb{R}_+}\left\{\lambda H(\mu(\cdot\times s^{-1/2})|\gamma_1)+(1-\lambda)H(\nu(\cdot\times s^{-1/2})|\gamma_1)\right.\\ &+\frac{1}{2}\left[s-\lambda m_2(\mu(\cdot\times s^{-1/2}))-(1-\lambda)m_2(\nu(\cdot\times s^{-1/2}))\right]:\\ &\left. s\geq\lambda m_2(\mu(\cdot\times s^{-1/2}))+(1-\lambda)m_2(\nu(\cdot\times s))\right\}.\end{aligned}\tag{4.3.5}$$

Now we use that for some constant $c\in\mathbb{R}_+$, $H(\mu_1(\cdot\times c)|\mu_2)=H(\mu_1|\mu_2(\cdot\times c^{-1}))$ and note

$$m_2(\mu(\cdot\times s^{-1/2}))=\int_{\mathbb{R}}x^2d\mu(x\times s^{-1/2})=\int_{\mathbb{R}}sy^2d\mu(y)=sm_2(\mu).$$

Using these observations, we can continue to simplify our GRF and get that 4.3.5 is equal to

$$\begin{aligned}&\inf_{s\in\mathbb{R}_+}\left\{\lambda H(\mu|\gamma_1(\cdot\times s^{1/2}))+(1-\lambda)H(\nu|\gamma_1(\cdot\times s^{1/2}))\right.\\ &\left.+\frac{s}{2}\left[1-\lambda m_2(\mu)-(1-\lambda)m_2(\nu)\right]:1\geq\lambda m_2(\mu)+(1-\lambda)m_2(\nu)\right\}.\end{aligned}$$

Now we can use Proposition 1, here $\gamma_1(\cdot\times s^{1/2})=\gamma_{s^{-1}}$, hence

$$H(\mu|\gamma_1(\cdot\times s^{1/2}))=H(\mu|\gamma_1)-\frac{1}{2}\log s-\frac{1-s}{2}m_2(\mu).$$

Thus the GRF from 4.3.5 equals to

$$\begin{aligned}\lambda H(\mu|\gamma_1)+(1-\lambda)H(\nu|\gamma_1)+&\inf_{s\in\mathbb{R}_+}\left\{-\lambda\left(\frac{1}{2}\log s+\frac{s-1}{2}m_2(\mu)\right)-(1-\lambda)\left(\frac{1}{2}\log s+\frac{1-s}{2}m_2(\nu)\right)\right.\\ &\left.+\frac{s}{2}\left[1-\lambda m_2(\mu)-(1-\lambda)m_2(\nu)\right]:1\geq\lambda m_2(\mu)+(1-\lambda)m_2(\nu)\right\}\\ &=\mathbb{I}_2(\mu,\nu)-\frac{1}{2}+\frac{1}{2}\inf_{s\in\mathbb{R}_+}\left\{s-\log s\right\}\\ &=\mathbb{I}_2(\mu,\nu).\end{aligned}$$

□

Now we collected everything we need in order to analyse the behaviour of $\hat\mu^n_{\mathbf{A}}$, which is done in the next lemma.

Lemma 4.3.4. *(Proposition 4.7 in [18])*

Fix $q < 2$. Suppose k_n grows linearly with rate $\lambda \in (0,1]$. Then, the sequence $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ satisfies an LDP in $P_q(\mathbb{R})$ at speed n and GRF $\mathcal{H}_\lambda : P(\mathbb{R}) \to [0,\infty]$, where for $\nu \in P(\mathbb{R})$

$$\mathcal{H}_\lambda(\nu) := \begin{cases} \lambda h(\nu) + \frac{\lambda}{2}\log(2\pi e) + \frac{1-\lambda}{2}\log\left(\frac{1-\lambda}{1-\lambda m_2(\nu)}\right) & , \text{if } m_2(\nu) < 1/\lambda \\ \infty & , \text{else.} \end{cases} \tag{4.3.6}$$

In case of $\lambda = 1$, the third term containing the logarithm is set to 0.

Proof. Instead of proving the stronger LDP for the q-Wasserstein topology on $P_q(\mathbb{R})$, we can use the same argument as in Lemma 4.2.1 to simplify the problem and show only the LDP w.r.t the weak topology on $P(\mathbb{R})$. Further, we use the established representation of $\hat{\mu}^n_{\mathbf{A}}$ via iid normal distributed ζ_j, namely

$$\hat{\mu}^n_{\mathbf{A}} = \frac{1}{k_n}\sum_{j=1}^{k_n} \delta_{\sqrt{n}\zeta_j/\|\zeta^{(n)}\|_2}.$$

We remember that this representation of $\hat{\mu}^n_{\mathbf{A}}$ equals to the first component of the analysed vector in Lemma 4.3.3. Next we note that a different representation of the GRF $\mathbb{I}_2$ from Lemma 4.3.3, using Application 3, is possible. We substitute $H(\cdot|\gamma_1) = -h(\cdot) + 1/2\log(2\pi) + 1/2 m_2(\cdot)$. Then, for $\lambda m_2(\mu) + (1-\lambda)m_2(\nu) \le 1$,

$$\begin{aligned} \mathbb{I}_2(\mu,\nu) :=& -\lambda h(\mu) + \frac{\lambda}{2}\log(2\pi) + \frac{\lambda}{2}m_2(\mu) \\ & - (1-\lambda)h(\nu) + \frac{1-\lambda}{2}\log(2\pi) + \frac{1-\lambda}{2}m_2(\nu) \\ & + \frac{1}{2}(1 - \lambda m_2(\mu) - (1-\lambda)m_2(\nu)) \\ =& -\lambda h(\mu) - (1-\lambda)h(\nu) + \frac{1}{2}\log(2\pi e). \end{aligned}$$

We can hence apply the contraction principle to the LDP from 4.3.3 and project via the continuous mapping $F : P(\mathbb{R}) \times P(\mathbb{R}) \to P(\mathbb{R}), (\mu,\nu) \mapsto \mu$ on the first component. This establishes an LDP for $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ at speed n and GRF

$$\begin{aligned} \mu \mapsto \inf_{\nu\in P(\mathbb{R})} \mathbb{I}_2(\mu,\nu) &= -\lambda h(\mu) + \frac{1}{2}\log(2\pi e) \\ &\quad + \inf_{\nu\in P(\mathbb{R})}\{-(1-\lambda)h(\nu) : (1-\lambda)m_2(\nu) + \lambda m_2(\mu) \le 1\} \\ &= -(1-\lambda)\sup_{\nu\in P(\mathbb{R})}\left\{h(\nu) : m_2(\nu) \le \frac{1-\lambda m_2(\mu)}{1-\lambda}\right\} - \lambda h(\mu) + \frac{1}{2}\log(2\pi e). \end{aligned}$$

If $\lambda = 1$, then our GRF equals $\mathcal{H}_\lambda$. In case of $\lambda < 1$, we note that if $1 \le \lambda m_2(\mu)$, then $m_2(\nu) \le 0$ and hence ν is degenerated with $h(\nu) = -\infty$. In this case our rate function equals ∞. If $m_2(\mu) < 1/\lambda$, then the restriction on the second moment of ν is of the form $m_2(\nu) \le z$ for $z > 0$. We can use Application 4 and deduce that our GRF can be simplified to

$$\begin{aligned}\mu \mapsto -\lambda h(\mu) + \frac{1}{2}\log(2\pi e) - (1-\lambda)\frac{1}{2}\log\left(2\pi e\frac{1-\lambda m_2(\mu)}{1-\lambda}\right) \\ = -\lambda h(\mu) + \frac{\lambda}{2}\log(2\pi e) + \frac{(1-\lambda)}{2}\log\left(\frac{1-\lambda}{1-\lambda m_2(\mu)}\right) \\ = \mathcal{H}_\lambda(\mu).\end{aligned}$$

□

Now we can finally state the central theorem in this section, showing the desired LDP for the sequence of empirical measures $\{L^n\}_{n\in\mathbb{N}}$.

Theorem 4.3.5. *(Theorem 2.13. in [18])*
Fix $q < 2$. Suppose $\{k_n\}_{n\in\mathbb{N}}$ grows linearly with rate $\lambda \in (0,1]$ and Assumption A holds with sequence $\{s_n\}_{n\in\mathbb{N}}$ and GRF J_X. Then $\{L^n\}_{n\in\mathbb{N}}$ satisfies an LDP in $P_q(\mathbb{R})$ at speed s_n with GRF $\mathbb{I}_{L,\lambda} : P_q(\mathbb{R}) \to [0,\infty]$, where

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

$$\mathbb{I}_{L,\lambda}(\mu) = \begin{cases} \inf_{c\in\mathbb{R}_+}\left\{J_X(c) - \frac{1-\lambda}{2}\log\left(1 - \frac{\lambda m_2(\mu)}{c^2}\right) + \lambda\log c : m_2(\mu) < c^2/\lambda\right\} \\ \qquad -\lambda h(\mu) + \frac{\lambda}{2}\log(2\pi e) + \frac{1-\lambda}{2}\log(1-\lambda) \\ \infty, & \textit{otherwise}. \end{cases}$$

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

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

Proof. We distinguish the two growth regimes of the speed s_n in Assumption A, starting with the case $s_n = n$:
The idea is to apply Lemma 4.0.7 and rearrange and simplify the corresponding GRF as far as possible. The LDP for $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ established in Lemma 4.3.3 and Assumption

A on the sequence $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ leads to

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

We know that $\mathcal{H}_\lambda(\nu) = \infty$ for $m_2(\nu) \geq 1/\lambda$ and since

$$m_2(\mu) = \int_{\mathbb{R}} x^2 d\mu(x) = \int_{\mathbb{R}} x^2 d\nu(xc^{-1}) = c^2 m_2(\nu),$$

it follows that $\hat{\mathbb{I}}_{L,\lambda}(\mu) = \infty$, if $m_2(\mu) \geq c^2/\lambda$. Thus, in this case $\hat{\mathbb{I}}_{L,\lambda}(\mu) = \mathbb{I}_{L,\lambda}(\mu)$. Hence, we assume $m_2(\mu) < c^2/\lambda$. Then $m_2(\nu) < 1/\lambda$ and we get

$$\begin{aligned}
\mathcal{H}_\lambda(\nu) + J_X(c) &= \mathcal{H}_\lambda(\mu(\cdot \times c)) + J_X(c) \\
&= -\lambda h(\mu(\cdot \times c)) + \frac{\lambda}{2}\log(2\pi e) + \frac{1-\lambda}{2}\log\left(\frac{1-\lambda}{1-\lambda m_2(\mu(\cdot\times c))}\right) + J_X(c) \\
&= -\lambda h(\mu) + \lambda \log c + \frac{\lambda}{2}\log(2\pi e) + \frac{1-\lambda}{2}\log\left(\frac{1-\lambda}{1-\lambda/c^2 m_2(\mu)}\right) + J_X(c) \\
&= -\lambda h(\mu) - \frac{1-\lambda}{2}\log\left(1-\frac{\lambda m_2(\mu)}{c^2}\right) + \lambda\log c + \frac{\lambda}{2}\log(2\pi e) + J_X(c) \\
&+ \frac{(1-\lambda)}{2}\log(1-\lambda).
\end{aligned}$$

Taking the infimum over all $c > 0$ establishes the identity $\hat{\mathbb{I}}_{L,\lambda}(\mu) = \mathbb{I}_{L,\lambda}(\mu)$ in the case $m_2(\mu) < c^2/\lambda$.

Now we consider the situation when $s_n \ll n$: We remember Remark 3.1.2 which can be applied to $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ to establish an LDP with speed s_n and GRF

$$\chi_{\gamma_1}(\nu) = \begin{cases} 0, & \text{if } \nu = \gamma_1 \\ \infty, & \text{else.} \end{cases}$$

We get an LDP for L^n with speed s_n and GRF

$$\begin{aligned}
\mathbb{I}_{L,\lambda}(\mu) &:= \inf_{\nu\in P(\mathbb{R}),c\in\mathbb{R}_+} \{\chi_{\gamma_1}(\nu) + J_X(c)\} \\
&= \begin{cases} J_X(c), & \text{if } \mu = \gamma_c \\ \infty, & \text{else.} \end{cases}
\end{aligned}$$

$\square$

As in the previous sections we finally want to investigate the behaviour of properly scaled q-norms. Therefore, we recall the quantity $Y^n_{q,k}$ from Equation 4.1.3, i.e.

$$Y^n_{q,k_n} = n^{-1/q} \| A^T_{n,k_n} X^{(n)} \|_q. \tag{4.3.7}$$

Our goal is to establish LDPs for $\{Y^n_{q,k_n}\}_{n\in\mathbb{N}}$ for all values $q \in [1,2]$. We start with the case $q < 2$, where we can simply use the contraction principle, in the case $q = 2$ this does not work anymore, since the used mapping is not continuous in the 2-Wasserstein topology. Here we need a different technique, which in principle would work for all q, but as we will see, the representation of the corresponding GRF is very cumbersome compared to the received representation using the contraction principle. We start with a different representation of Y^n_{q,k_n}, using some of our previous results. We have

$$\begin{aligned} Y^n_{q,k_n} &= n^{-1/q} \| A^T_{n,k_n} X^{(n)} \|_q = n^{-1/q} \left\| A^T_{n,k_n} \frac{X^{(n)}}{\|X^{(n)}\|_2} \|X^{(n)}\|_2 \right\|_q \\ &\overset{(d)}{=} n^{-1/q} \left\| A^T_{n,k_n} e_1 \|X^{(n)}\|_2 \right\|_q \\ &\overset{(d)}{=} n^{-1/q} \frac{\|\zeta^{(k_n)}\|_q}{\|\zeta^{(n)}\|_2} \|X^{(n)}\|_2, \end{aligned} \tag{4.3.8}$$

where $\zeta^{(k)} = (\zeta_1, \dots, \zeta_k)$ for a sequence of iid standard normal distributed random variables $\{\zeta_n\}_{n\in\mathbb{N}}$. Our next goal is to find an LDP for the quotient of properly scaled norms. This is possible, but takes some effort and will lead to a rather complicated GRF. We start by introducing some notation. Fix $q \in [1,2]$ and define

$$\Lambda_{A,q}(t_1,t_2) := \log \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(t_1 |x|^q + \left(t_2 - \frac{1}{2} \right) x^2 \right) dx, \qquad (t_1,t_2) \in \mathbb{R}^2, \tag{4.3.9}$$

where we note that for $q \in [1,2)$, we have $\Lambda_{A,q}(t_1,t_2) < \infty$ when $t_1 \in \mathbb{R}$ and $t_2 < \frac{1}{2}$. In case of $q = 2$, we have $\Lambda_{A,q}(t_1,t_2) < \infty$ when $t_1 + t_2 < \frac{1}{2}$. Further, Λ_B will appear, with

$$\Lambda_B(t_3) := \log \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(\left(t_3 - \frac{1}{2} \right) x^2 \right) dx, \qquad t_3 \in \mathbb{R}. \tag{4.3.10}$$

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

$$\mathbf{J}_{q,\lambda}(z) := \inf_{(x_1,x_2,x_3)\in\mathbb{R}^3} \left\{ \lambda \Lambda^*_{A,q}\left(\frac{(x_1,x_2)}{\lambda} \right) + (1-\lambda) \Lambda^*_B\left(\frac{x_3}{1-\lambda} \right) : z = \frac{x_1^{1/q}}{(x_2+x_3)^{1/2}} \right\}. \tag{4.3.11}$$

Corollary. We consider the case $q = 2$. Then $\mathbf{J}_{q,\lambda}$ can be simplified to

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

Proof. It can be verified immediately that $\Lambda_{A,2}(t_1,t_2) = -\frac{1}{2}\log(1-2[t_1+t_2])$ for $t_1+t_2 < \frac{1}{2}$ as well as $\Lambda_B(t_3) = -\frac{1}{2}\log(1-2t_3)$ for $t_3 < \frac{1}{2}$. An elementary computation leads to the respective Legendre transforms, namely

$$\Lambda_B^*(z) = \frac{1}{2}[z - \log z - 1], \qquad z \in \mathbb{R}_+$$

$$\Lambda_{A,2}^*((z_1,z_2)) = \begin{cases} \infty & ,\text{if } z_1 \neq z_2 \\ \Lambda_B^*(z) & ,\text{if } z_1 = z_2 = z \end{cases} \qquad z_i \in \mathbb{R}_+.$$

Using these representations in the formula for $\mathbf{J}_{2,\lambda}$ gives us

$$\mathbf{J}_{2,\lambda}(z) = \inf_{(x_1,x_2,x_3)\in\mathbb{R}} \left\{ \frac{\lambda}{2}\left[\frac{x}{\lambda} - \log\left(\frac{x}{\lambda}\right) - 1\right] + \frac{1-\lambda}{2}\left[\frac{x_3}{1-\lambda} - \log\left(\frac{x_3}{1-\lambda}\right) - 1\right] : z = \sqrt{\frac{x}{x+x_3}}, x_1 = x_2 = x > 0, x_3 > 0 \right\}.$$

Now we can use the relation $x_3 = x(1/z^2 - 1)$ and after some elementary rearrangements we are left with

$$\mathbf{J}_{2,\lambda}(z) = \inf_{x>0}\left\{ -\frac{\lambda}{2}\log\left(\frac{x}{\lambda}\right) + \frac{x}{2z^2} - \frac{1-\lambda}{2}\log\left(\frac{x}{1-\lambda}\right)\right\} - \frac{1-\lambda}{2}\log\left(\frac{1}{z^2}-1\right) - \frac{1}{2}$$

for $z \in (0,1)$. Minimization over x leads to the desired function. □

The next lemma establishes an LDP for $\{\|\zeta^{(k_n)}\|_q / \|\zeta^{(n)}\|_2\}_{n\in\mathbb{N}}$, which we can then use to receive an LDP for $\{Y^n_{q,k_n}\}_{n\in\mathbb{N}}$.

Lemma 4.3.6. *(Lemma 6.2. in [18])*
Suppose $\{k_n\}_{n\in\mathbb{N}}$ grows linearly with rate $\lambda \in (0,1]$. For $q \in [1,2]$, the sequence

$$n^{1/2-1/q}\frac{\|\zeta^{(k_n)}\|_q}{\|\zeta^{(n)}\|_2}, \quad n \in \mathbb{N} \tag{4.3.12}$$

satisfies an LDP at speed n with GRF $\mathbf{J}_{q,\lambda}$.

Proof. Fix $q \in [1,2]$. For $n \in \mathbb{N}$, let

$$\mathbf{A}_{q,n} := \frac{1}{k_n} \sum_{i=1}^{k_n} (|\zeta_i|^q, \zeta_i^2) \quad \text{and} \quad \mathbf{B}_n := \frac{1}{n-k_n} \sum_{i=k_n+1}^{n} \zeta_i^2.$$

We note that $\Lambda_{A,q}$ is the cumulant generating function of $(|\zeta_1|^q, \zeta_1^2)$ and that Λ_B is the one of ζ_1^2. Due to Remark 3.1.3, $\{\mathbf{A}_{q,n}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^2$ with speed k_n and GRF $\Lambda^*_{A,q}$, since $0 \in \mathcal{D}^o_{\Lambda_{A,q}}$. The same argument applies to $\{\mathbf{B}_n\}_{n\in\mathbb{N}}$, where we obtain an LDP in $\mathbb{R}$ with speed $n-k_n$ and GRF Λ^*_B. Since $k_n/n \to \lambda$, we can establish an LDP for

$$\frac{1}{k_n} \sum_{i=1}^{k_n} (|\zeta_i|^q, \zeta_i^2), \qquad n \in \mathbb{N}$$

at speed n and GRF $\lambda\Lambda^*_{A,q}$. Further, we can use the contraction principle for the function $F : \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2, (x,y) \mapsto xy$ to establish an LDP for

$$F\Big(\frac{k_n}{n}, \mathbf{A}_{n,q}\Big) = \frac{1}{n} \sum_{i=1}^{k_n} (|\zeta_i|^q, \zeta_i^2)$$

with speed n and GRF

$$z \mapsto \inf_{z=xy} \big[\lambda\Lambda^*_{A,q}(y) + \chi_\lambda(x)\big] = \lambda\Lambda^*_{A,q}\Big(\frac{(z_1,z_2)}{\lambda}\Big),$$

where χ_λ is the GRF of the trivial LDP for k_n/n. Analogously one can show that also

$$\frac{1}{n} \sum_{i=k_n+1}^{n} \zeta_i^2, \qquad n \in \mathbb{N}$$

satisfies an LDP with speed n with the GRF

$$z \mapsto (1-\lambda)\Lambda^*_B\Big(\frac{z}{1-\lambda}\Big).$$

Hence, the sequence

$$\frac{1}{n}\Big(\sum_{i=1}^{k_n} (|\zeta_i|^q, \zeta_i^2), \sum_{i=k_n+1}^{n} \zeta_i^2\Big), \qquad n \in \mathbb{N}$$

satisfies an LDP at speed n and with GRF

$$(z_1,z_2,z_3) \mapsto \lambda\Lambda^*_{A,q}\Big(\frac{(z_1,z_2)}{\lambda}\Big) + (1-\lambda)\Lambda^*_B\Big(\frac{z_3}{1-\lambda}\Big).$$

Finally, another application of the contraction principle for the mapping

$$T : \mathbb{R}_+^3 \to \mathbb{R}_+$$
$$(z_1, z_2, z_3) \mapsto \frac{z_1^{1/q}}{(z_2 + z_3)^{1/2}}$$

yields an LDP for

$$T\bigg(\frac{1}{n}\sum_{i=1}^{k_n}(|\zeta_i|^q, \zeta_i^2), \frac{1}{n}\sum_{i=k_n+1}^{n}\zeta_i^2\bigg) = n^{1/2-1/q}\frac{\|\zeta^{(k_n)}\|_q}{\|\zeta^{(n)}\|_2}, \qquad n \in \mathbb{N}$$

at speed n and GRF $\mathbf{J}_{q,\lambda}$ from Equation 4.3.11. □

The LDP for $\{Y_{q,k_n}^n\}_{n\in\mathbb{N}}$ can now be derived directly using our representation via the quotient of standard normal distributed random vectors. In the following theorem we will first use a different approach, avoiding the GRF $\mathbf{J}_{q,\lambda}$.

Theorem 4.3.7. *(Theorem 2.14 in [18])*
Suppose $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption A, and $k_n \sim \lambda n$ for some $\lambda \in (0,1]$. Also, for $q \in [1,2]$ and $n, k_n \in \mathbb{N}$, with $k_n \leq n$, define $\{Y_{q,k_n}^n\}_{n\in\mathbb{N}}$ as in Equation 4.3.7. Then the sequence $\{Y_{q,k_n}^n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF $\mathbb{J}_{Y_{q,k_n,\lambda}}$, where for $x \in \mathbb{R}_+$,

$$\mathbb{J}_{Y_{q,k_n,\lambda}}(x) := \begin{cases} \inf_{\nu\in P(\mathbb{R}), c\in\mathbb{R}_+}\{\mathcal{H}_\lambda(\nu) + J_X(c) : \lambda m_q(\nu) = (x/c)^q\}, & \text{if } s_n = n, \\ J_X\Big(\frac{x}{(\lambda M_q)^{1/q}}\Big), & \text{if } s_n \ll n, \end{cases} \tag{4.3.13}$$

with $m_q(\nu)$ the q-th moment map and M_q the q-th absolute moment of a standard Gaussian random variable defined in Equation 4.2.4.

Proof. First we consider the case $q < 2$ and start by taking a closer look at Y_{q,k_n}^n. We observe that

$$\begin{aligned} n^{-1/q}\|A_{n,k_n}^T X^{(n)}\|_q &= \bigg(\frac{k_n}{n}\frac{1}{k_n}\sum_{j=1}^{k_n}|(A_{n,k_n}^T X^{(n)})_j|^q\bigg)^{1/q} \\ &= \bigg(\frac{k_n}{n}m_q(L^n)\bigg)^{1/q}. \end{aligned}$$

A very similar argument as in the previous lemma shows that $\{\frac{k_n}{n}m_q(L^n)\}_{n\in\mathbb{N}}$ satisfies the same LDP as $\{\lambda m_q(L^n)\}_{n\in\mathbb{N}}$. We can now use our LDP for the sequence $\{L^n\}_{n\in\mathbb{N}}$ for $q < 2$ and the fact that $\nu \mapsto m_q(\nu)$ is continuous in the q-Wassertein topology. Hence,

$\{Y^n_{q,k_n}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ at speed s_n and with GRF

$$\mathbb{J}_{Y_q,k_n,\lambda}(x) := \inf_{\mu\in P(\mathbb{R})} \big\{\mathbb{I}_{L,\lambda}(\mu) : (\lambda m_q(\mu))^{1/q} = x\big\},$$

where $\mathbb{I}_{L,\lambda}$ is the GRF from Theorem 4.3.5.
For $s_n = n$, the rate function equals

$$\begin{aligned}
\mathbb{J}_{Y_q,k_n,\lambda}(x) &= \inf_{\mu\in P(\mathbb{R})} \big\{\mathbb{I}_{L,\lambda}(\mu) : [\lambda m_q(\mu)]^{1/q} = x\big\} \\
&= \inf_{\mu\in P(\mathbb{R})} \Big\{ \inf_{\nu\in P(\mathbb{R}),c>0} \big\{\mathcal{H}_\lambda(\nu) + J_X(c) : \mu = \nu(\cdot\times c^{-1})\big\} : [\lambda m_q(\mu)]^{1/q} = x\Big\} \\
&= \inf_{\mu\in P(\mathbb{R}),c>0} \big\{\mathcal{H}_\lambda(\mu(\cdot\times c)) + J_X(c) : [\lambda m_q(\mu)]^{1/q} = x\big\} \\
&= \inf_{\mu\in P(\mathbb{R}),c>0} \big\{\mathcal{H}_\lambda(\mu) + J_X(c) : [\lambda m_q(\mu)]^{1/q} = x/c\big\}, \qquad x\in\mathbb{R}_+,
\end{aligned}$$

where we used in the last step that $m_q(\mu) = m_q(\mu[\cdot\times cc^{-1}]) = c^q m_q(\mu[\cdot\times c])$.
If $s_n \ll n$, then

$$\begin{aligned}
\mathbb{J}_{Y_q,k_n,\lambda}(x) &: = \inf_{\mu\in P(\mathbb{R})} \big\{\big[J_X(c) : \mu = \gamma_c\big] : [\lambda m_q(\mu)]^{1/q} = x\big\} \\
&= \inf_{c>0} \big\{J_X(c) : [\lambda m_q(\gamma_c)]^{1/q} = x\big\} \\
&= \inf_{c>0} \big\{J_X(c) : [\lambda c^q M_q]^{1/q} = x\big\} \\
&= J_X\Big(\frac{x}{(\lambda M_q)^{1/q}}\Big), \qquad x\in\mathbb{R}_+.
\end{aligned}$$

This establishes the desired LDP in both cases for s_n if $q<2$. An alternative way using Lemma 4.3.6 allows to show an LDP for all $q\in[1,2]$. Recall Representation 4.3.8, where we showed that

$$Y^n_{q,k_n} \overset{(d)}{=} n^{1/2-1/q} \frac{\|\zeta^{(k_n)}\|_q}{\|\zeta^n\|_2} \frac{\|X^{(n)}\|_2}{\sqrt{n}}, \qquad n\in\mathbb{N}.$$

If $s_n = n$, then the contraction principle with the function $F:\mathbb{R}^2\to[0,\infty]$, $(u,v)\mapsto uv$ can be applied directly and gives us an LDP at speed n and with GRF

$$\bar{\mathbb{J}}_{Y_q,k_n,\lambda}(x) := \inf_{y,z\in\mathbb{R}} \big\{\mathbf{J}_{q,\lambda}(y) + J_X(z) : yz = x\big\}.$$

Since the GRF is unique, we deduce that $\bar{\mathbb{J}}_{Y_q,k_n,\lambda} = \mathbb{J}_{Y_q,k_n,\lambda}$ for $q<2$ and $s_n = n$.

In the case of $s_n \ll n$ we still have an LDP for

$$n^{1/2-1/q}\frac{||\zeta^{(k_n)}||_q}{||\zeta^n||_2}, \qquad n \in \mathbb{N}$$

at speed s_n and with (degenerated) GRF χ_ξ (see Remark 3.1.2) with

$$n^{1/2-1/q}\frac{||\zeta^{(k_n)}||_q}{||\zeta^n||_2} = \frac{||\zeta^{(k_n)}||_q}{k_n^{1/q}}\frac{k_n^{1/q}}{n^{1/q}}\frac{n^{1/2}}{||\zeta^n||_2} \xrightarrow{a.s.} (M_q\lambda)^{1/q} := \xi.$$

The GRF for $\{Y^n_{q,k_n}\}_{n\in\mathbb{N}}$ is then given as

$$\begin{aligned}\bar{\mathbb{J}}_{Y_q,k_n,\lambda}(x) &= \inf_{y,z\in\mathbb{R}}\left\{\chi_\xi(z) + J_X(y) : x = yz\right\}\\ &= J_X\left(\frac{x}{(\lambda M_q)^{1/q}}\right).\end{aligned}$$

Hence, we established the identity $\bar{\mathbb{J}}_{Y_q,k_n,\lambda} = \mathbb{J}_{Y_q,k_n,\lambda}$ for all $q < 2$ and for $q = 2$ in the case $s_n \ll n$.

It remains to show that this equality also holds for $q = 2$ and $s_n = n$. If we are able to prove that

$$\mathbf{J}_{2,\lambda}(z) = \inf_{\mu\in P(\mathbb{R})}\left\{\mathcal{H}_\lambda(\mu) : z^2 = \lambda m_2(\mu)\right\} \tag{4.3.14}$$

we are finished, since then

$$\begin{aligned}\bar{\mathbb{J}}_{Y_q,k_n,\lambda}(x) &= \inf_{x=yz}\left\{\mathbf{J}_{2,\lambda}(z) + J_X(y)\right\}\\ &= \inf_{y>0}\left\{\mathbf{J}_{2,\lambda}\left(\frac{x}{y}\right) + J_X(y)\right\}\\ &= \inf_{y>0}\inf_{\mu\in P(\mathbb{R})}\left\{\mathcal{H}_\lambda(\mu) + J_X(y) : (x/y)^2 = \lambda m_2(\mu)\right\}\\ &= \mathbb{J}_{Y_q,k_n,\lambda}(x).\end{aligned}$$

We get for $z^2 < 1$ that

$$\begin{aligned}\inf_{\nu\in P(\mathbb{R}):z^2=\lambda m_2(\nu)}\mathcal{H}_\lambda(\nu) &= \inf_{\nu\in P(\mathbb{R}):z^2=\lambda m_2(\nu)}\left\{-\lambda h(\nu) + \frac{\lambda}{2}\log(2\pi e) + \frac{1-\lambda}{2}\log\left(\frac{1-\lambda}{1-\lambda m_2(\nu)}\right)\right\}\\ &= -\lambda\sup_{\nu\in P(\mathbb{R}):z^2=\lambda m_2(\nu)}h(\nu) + \frac{\lambda}{2}\log(2\pi e) + \frac{1-\lambda}{2}\log\left(\frac{1-\lambda}{1-z^2}\right).\end{aligned}$$

In order to show the equality in 4.3.14, we recall Application 4 about maximizing the relative entropy under a bounded second moment. The supremum is attained with value $\frac{1}{2}\log(2\pi e z^2/\lambda)$. Using this leads to

$$\inf_{\nu\in P(\mathbb{R}):z^2=\lambda m_2(\nu)} \mathcal{H}_\lambda(\nu) = -\frac{\lambda}{2}\log\left(\frac{2\pi e z^2}{\lambda}\right) + \frac{\lambda}{2}\log(2\pi e) + \frac{1-\lambda}{2}\log\left(\frac{1-\lambda}{1-z^2}\right)$$
$$= \bar{\mathbb{J}}_{Y_q,k_n,\lambda}(z), \qquad z \in (0,1).$$

For $z \notin (0,1)$ the equality holds with value infinity, since $\mathcal{H}_\lambda(\nu) = \infty$ whenever $m_2(\nu) \geq 1/\lambda$. $\square$

Remark. In [18] the authors elaborated a comment on the interesting observation that for $q < 2$ we were usually able to establish LDPs in all the asymptotic regimes, whereas the case $q = 2$ is more delicate. From the technical point of view this appears in terms of the set $K_{2,j}$ from Lemma 3.3.5, which is not compact in the q-Wasserstein topology for $q \geq 2$. The same obstacle becomes visible by looking at the representation

$$\mathbf{J}_{q,\lambda}(z) = \inf_{\nu\in P(\mathbb{R}):z^2=\lambda m_2(\nu)} \mathcal{H}_\lambda(\nu)$$

appearing in the proof of Theorem 4.3.7. The expression looks like we applied the contraction principle to the sequence of empirical measures $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$, even if this seems not possible for $q = 2$. To illustrate this phenomenon in a simpler setting, the authors considered a sequence of iid exponential distributed random variables, where essentially the same observation can be made: in this case the theory suggests that the LDP for empirical measures does not hold for $q = 1$, but it is still possible to show the LDP and a variational formula for the corresponding GRF. They suggested that these observations can be seen as a more general problem in large deviation theory.

5 Applications

In this section we want to verify some of our assumptions for different applications. We start with a „warm-up“ example containing product measures. We then conclude by studying certain Orlicz balls and Gibbs measures.

5.1 Product measures

Lemma 5.1.1. *(Lemma 3.1 in [9])*
Let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of iid real-valued random variables and let $X^{(n)} := (X_1, ..., X_n)$. Suppose that we have

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

for all $t \le \epsilon$ with $\epsilon > 0$. Let $\Lambda^* : \mathbb{R} \to [0, \infty]$ be the Legendre transform of Λ, i.e.

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

Then, $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption A with GRF $J_X : \mathbb{R} \to [0, \infty]$, where*

$$J_X(x) := \begin{cases} \Lambda^*(x^2), & \text{if } x \ge 0 \\ \infty & \text{else.} \end{cases}$$

Moreover, J_X has a unique minimizer $m := \sqrt{\mathbb{E}[X_1^2]}$.

Proof. Using Cramer's theorem 3.1.11, we can establish an LDP for

$$\frac{\|X^{(n)}\|_2^2}{n} = \frac{X_1^2 + ... + X_n^2}{n}, \qquad n \in \mathbb{N}$$

at speed n and with GRF Λ^*. We can then apply the contraction principle for the continuous mapping $F : \mathbb{R}_+ \to \mathbb{R}_+, x \mapsto \sqrt{x}$, where we get an LDP for $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$

at speed n and with GRF

$$J_X(x) = \inf_{y \in F^{-1}(\{x\})} \Lambda^*(y) = \begin{cases} \Lambda^*(x^2), & \text{if } x \geq 0 \\ \infty & \text{else.} \end{cases}$$

For the minimizer m, we have by the strong law of large numbers

$$\frac{\|X^{(n)}\|_2}{\sqrt{n}} = \sqrt{\frac{1}{n}\sum_{i=1}^n X_i^2} \overset{n\to\infty}{\longrightarrow} \sqrt{\mathbb{E}[X_1^2]} = m \quad \text{almost surely.}$$

This value m is indeed the unique minimizer of Λ^*. First, we recall that $\Lambda^* \geq 0$ and we note that for every $\delta > 0$

$$\mathbb{P}\Big[\frac{\|X^{(n)}\|_2}{\sqrt{n}} \in [m-\delta, m+\delta]\Big] \longrightarrow 1.$$

Then, using the large deviation upper bound, we have

$$0 = \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big[\frac{\|X^{(n)}\|_2}{\sqrt{n}} \in [m-\delta, m+\delta]\Big] \leq - \inf_{x\in[m-\delta,m+\delta]} \Lambda^*(x) \leq 0.$$

And hence, we get

$$0 = \lim_{\delta\to 0} \inf_{x\in[m-\delta,m+\delta]} \Lambda^*(x) \geq \Lambda^*(m),$$

where we used the lower semi-continuity of Λ^*. Hence, m is a minimizer of Λ^*. The uniqueness follows, since a Legendre transform of a cumulant generating function is strictly convex on the set where it is finite. □

5.2 Orlicz balls

Orlicz balls are a natural generalization of ℓ_p^n balls with nice properties like the "negative association" (as shown in Theorem 1.2 of [24]), a weaker property than independence, or the KLS (Kannan-Lovász-Simonovits) conjecture (under mild conditions concerning the growth of the involved Orlicz functions), see Theorem 1.1 and Theorem 2.4 of [21]. We start with some definitions and notation.

Definition 5.2.1. *We say V is an Orlicz function if $V : \mathbb{R} \to \mathbb{R}_+$ is convex and satisfies*

$V(0) = 0$ and $V(x) = V(-x)$ for $x \in \mathbb{R}$. Further, we say $V : \mathbb{R} \to \mathbb{R}_+$ grows superquadratic if

$$V(x)/x^2 \longrightarrow \infty, \quad \text{as } x \to \infty.$$

For such a V, i.e. a superquadratic Orlicz function, we define the symmetric Orlicz ball as

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

Now fix a sequence $V_1, V_2, \dots$ of such functions and denote the generalized Orlicz ball by

$$\mathbb{B}_{V_1,\dots,V_n}^n := \Big\{ x \in \mathbb{R}^n : \sum_{i=1}^n V_i(x_i) \le n \Big\}.$$

Remark. There are different definitions of Orlicz balls possible, for example one could replace the convexity assumption by assuming continuity and quasi-convexity (a function is called quasi-convex, if all level sets are convex).

Remark. For $p > 2$ it can be checked immediately that $V(x) = |x|^p$ is a superquadratic Orlicz function. We then have the relation

$$\begin{aligned}
\mathbb{B}_V^n &= \Big\{ x \in \mathbb{R}^n : \sum_{i=1}^n V(x_i) \le n \Big\} \\
&= n^{1/p} \Big\{ n^{-1/p} x \in \mathbb{R}^n : \sum_{i=1}^n V(x_i / n^{1/p}) \le 1 \Big\} \\
&= n^{1/p} \mathbb{B}_p^n,
\end{aligned}$$

where $\mathbb{B}_p^n$ denotes the ℓ_p ball in $\mathbb{R}^n$ with radius 1.

5.2.1 Symmetric Orlicz balls

We want to consider random variables distributed uniformly on symmetric Orlicz balls. We therefore introduce the distribution

$$\mu(A) := \frac{\lambda^n(\mathbb{B}_V^n \cap A)}{\lambda^n(\mathbb{B}_V^n)}, \qquad A \in \mathcal{B}(\mathbb{R}^n),$$

where λ^n is the n-dimensional Lebesgue measure. μ is a distribution on $\mathbb{R}^n$, we hence can construct a random variable $X^{(n)} \sim \mu$, which we call the uniform distribution on $\mathbb{B}_V^n$, denoted by $X^{(n)} \sim Unif(\mathbb{B}_V^n)$.

Remark. For this definition it is important that $\lambda^n(\mathbb{B}_V^n)$ is finite. This can be guaranteed, since $\mathbb{B}_V^n$ is compact and in particular bounded.

Our goal in this subsection is to verify Assumption A* for the sequence $\{X^{(n)}\}_{n\in\mathbb{N}}$ with $X^{(n)} \sim Unif(\mathbb{B}_V^n)$. This needs some preparation, we therefore introduce a few quantities.

Definition 5.2.2. *For $\nu \in P(\mathbb{R})$, we define the "V-th moment" mapping*

$$m_V(\nu) := \int_{\mathbb{R}} V(x) d\nu(x)$$

for a ν such that this integral is finite. Further for $b > 0$, define $\mu_{V,b} \in P(\mathbb{R})$ via

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

where $Z_{V,b} = \int_{\mathbb{R}} e^{-bV(x)} dx$ is the normalizing constant. Note that $Z_{V,b}$ is finite since $-b$ is negative and V grows superquadratic. Hence, $d_{\mu,b}$ is well-defined. For $s < 0$ and $t \in \mathbb{R}$, we will also work with the distribution $\nu_{s,t}$ given by

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

where $Z_{s,t} := \int_{\mathbb{R}} e^{sV(x)+tx^2} dx$. Then we can introduce the function $\mathcal{J} : \mathbb{R}_+^2 \to [0, \infty]$ with

$$\begin{aligned}(u, v) \mapsto & \sup_{s,t\in\mathbb{R}} \left\{ su + tv - \log \int_{\mathbb{R}} e^{sV(x)+tx^2} dx \right\} \\ &= \sup_{s<0,t\in\mathbb{R}} \left\{ su + tv - \log \int_{\mathbb{R}} e^{sV(x)+tx^2} dx \right\}.\end{aligned}$$

The equality holds, because the integral would be infinite for $s \geq 0$.

The next lemma collects several properties of $\mathcal{J}$.

Lemma 5.2.1. *(Lemma 3.12 in [18])*
Let $\mathcal{J}$ be as in the previous definition, then:

1. *For $v \in \mathbb{R}_+$, there exists a unique $(s, t) \in \mathbb{R}_- \times \mathbb{R}$ such that the supremum in the definition of $\mathcal{J}(1, v)$ is attained and is the unique solution to $m_V(\nu_{s,t}) = 1$ and $m_2(\nu_{s,t}) = v$.*

2. *There exists a unique $b^* > 0$ such that $m_V(\mu_{V,b^*}) = 1$ and $v \mapsto \mathcal{J}(1, v)$ is a convex*

function on $\mathbb{R}_+$ with minimizer $m = m_2(\mu_{V,b^*})$. Moreover,

$$\mathcal{J}(1,m) = \sup_{s<0}\left\{s - \log\int_{\mathbb{R}} e^{sV(x)}dx\right\}. \tag{5.2.1}$$

3. *For $v > m$, the supremum in the definition of $\mathcal{J}(1,v)$ is attained at $(s,t) \in \mathbb{R}_- \times \mathbb{R}_+$, while for $0 < v < m$, the supremum is attained at $(s,t) \in \mathbb{R}_- \times \mathbb{R}_-$.*
4. *For $v \in \mathbb{R}_+$, $\mathcal{J}(1,v) = -\max_{\nu\in P(\mathbb{R})}\{h(\nu) : m_V(\nu) = 1, m_2(\nu) = v\}$ and $\mathcal{J}(1,m) = -\max_{\nu\in P(\mathbb{R})}\{h(\nu) : m_V(\nu) = 1\}$.*
5. *$\partial_u \mathcal{J}(u,v) < 0$ for $u,v \in \mathbb{R}_+$.*

Proof. See Theorem 6.0.19 and the subsequent comment. □

Remark. Property 1 of Lemma 5.2.1 can be used to show that the supremum in the definition of $\mathcal{J}(u,v)$ is uniquely attained for some $(s,t) \in \mathbb{R}_- \times \mathbb{R}$, depending on (u,v). This can be verified directly

$$\begin{aligned}\mathcal{J}(u,v) &= \sup_{s,t\in\mathbb{R}}\left\{su + tv - \log\int_{\mathbb{R}} e^{sV(x)+tx^2}dx\right\}\\ &= \sup_{s,t\in\mathbb{R}}\left\{su + tv - \log\int_{\mathbb{R}} e^{suV(x)/u+tx^2}dx\right\}\\ &= \tilde{\mathcal{J}}(1,v),\end{aligned}$$

where $\tilde{\mathcal{J}}$ is defined in the same way as $\mathcal{J}$, but instead of V, the superquadratic Orlicz function V/u is used.

Theorem 5.2.2. *(Proposition 3.11. in [9])*
For $n \in \mathbb{N}$ suppose $X^{(n)} \sim Unif(\mathbb{B}_V^n)$. Then, there exists a unique $b^* > 0$ such that $m_V(\mu_{V,b^*}) = 1$ and $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption A with $J_X = J_{X,V}$, where*

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

Moreover, $J_{X,V}$ has a unique minimizer $m := \sqrt{m_2(\mu_{V,b*})}$.

Proof. We recall Assumption A*, where we need to show an LDP for $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ at speed n and with GRF $J_{X,V}$. We will start by analysing $\{\|X^{(n)}\|_2^2/n\}_{n\in\mathbb{N}}$. Then we can apply the contraction principle to the resulting LDP. Let us introduce the set

$$\mathbb{B}_{2,V}^n(A) := \mathbb{B}_V^n \cap \left\{x \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 \in nA\right\}.$$

To abbreviate expressions in what follows, we denote $|A| := \lambda^n(A)$. Using the previous set, we get

$$\begin{aligned}\mathbb{P}\Big[\frac{\|X^{(n)}\|_2^2}{n} \in A\Big] &= \mathbb{P}\Big[\sum_{i=1}^{n}(X_i^n)^2 \in nA\Big]\\ &= \frac{|\mathbb{B}_{2,V}^n(A)|}{|\mathbb{B}_V^n|},\end{aligned}$$

where the second equality follows from the definition of the distribution of $X^{(n)}$. We will now consider the numerator and the denominator of the right-hand expression separately. We start with our set $\mathbb{B}_{2,V}^n(A)$.
1. Step: Upper bound of the LDP. Fix a closed set $F \subseteq \mathbb{R}_+$, then we want to show

$$\limsup_{n\to\infty} \frac{1}{n} \log |\mathbb{B}_{2,V}^n[F]| \le -\inf_{x\in F} \mathcal{J}(1, x).$$

Part 2 of Lemma 5.2.1 gives us a unique constant $b^* > 0$ and an induced distribution $\mu_{V,b*}$ such that we can define $m := m_2(\mu_{V,b^*})$. Further set

$$\begin{aligned}\alpha_+ &:= \min\{x \mid x \in [m, \infty) \cap F\}\\ \alpha_- &:= \max\{x \mid x \in [0, m] \cap F\},\end{aligned}$$

where ∞ or $-\infty$ is possible. By definition we always have $\alpha_- \le m \le \alpha_+$, we hence distinguish two cases.
1. Case: If $\alpha_- < m < \alpha_+$, then we get the bound

$$\begin{aligned}|\mathbb{B}_{2,V}^n[F]| &= |\mathbb{B}_V^n \cap [x \in \mathbb{R}^n : \|x\|_2^2 \in nF]]|\\ &= |\mathbb{B}_V^n \cap [x \in \mathbb{R}^n : \|x\|_2^2 \in n[F \cap [m, \infty)]| + |\mathbb{B}_V^n \cap [x \in \mathbb{R}^n : \|x\|_2^2 \in n[F \cap [0, m]]|\\ &\le |\mathbb{B}_V^n \cap [x \in \mathbb{R}^n : \|x\|_2^2 \ge n\alpha_+| + |\mathbb{B}_V^n \cap [x \in \mathbb{R}^n : \|x\|_2^2 \le n\alpha_-]|\\ &= |\mathbb{B}_{2,V}^n[[\alpha_+, \infty)]| + |\mathbb{B}_{2,V}^n[[0, \alpha_-]]|.\end{aligned}$$

We analyse the first summand. Fix $s < 0$ and $t > 0$, then (since $s < 0$)

$$\begin{aligned}\mathbb{B}_{2,V}^n[[\alpha_+, \infty)] &= \Big\{x \in \mathbb{R}^n : \sum_{i=1}^{n} V(x_i) \le n, \|x\|_2^2 \ge n\alpha_+\Big\}\\ &= \Big\{x \in \mathbb{R}^n : \exp\Big(s\sum_{i=1}^{n} V(x_i)\Big) \ge \exp(sn), \exp\Big(t\|x\|_2^2\Big) \ge \exp(nt\alpha_+)\Big\}.\end{aligned}$$

Applying the Lebesgue measure on both sides yields

$$\begin{aligned}
|\mathbb{B}^n_{2,V}[[\alpha_+,\infty)]| &= \int_{\mathbb{B}^n_{2,V}[[\alpha_+,\infty)} 1\,dx \\
&\le \int_{\mathbb{B}^n_{2,V}[[\alpha_+,\infty)} \exp\Big(s\sum_{i=1}^n V(x_i) - sn + t\|x\|_2^2 - nt\alpha_+)dx \\
&\le e^{-sn-nt\alpha_+}\int_{\mathbb{R}^n}\exp\Big(s\sum_{i=1}^n V(x_i) + t\sum_{i=1}^n x_i^2\Big)dx \\
&= e^{-sn-nt\alpha_+}\Big\{\int_{\mathbb{R}}\exp\Big(sV(x)+tx^2\Big)dx\Big\}^n.
\end{aligned}$$

We can divide by n and apply the limsup, then we are left with

$$\limsup_{n\to\infty}\frac{1}{n}|\mathbb{B}^n_{2,V}[[\alpha_+,\infty)]| \le -s - t\alpha_+ + \int_{\mathbb{R}}\exp\Big(sV(x)+tx^2\Big)dx.$$

Now we can minimize over all $s<0$ and $t>0$, which leads to

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

The last equality used statement 3 of Lemma 5.2.1, since we are in the case $\alpha_+ > m$.

The argument for $|\mathbb{B}^n_{2,V}[[0,\alpha_-]]|$ works very similar, here we fix $s<0$ and $t<0$ and again use statement 3 of Lemma 5.2.1. We end up with

$$\limsup_{n\to\infty}\frac{1}{n}|\mathbb{B}^n_{2,V}[[0,\alpha_-]]| \le -\mathcal{J}(1,\alpha_-).$$

Combining both estimates gives us

$$\begin{aligned}
\limsup_{n\to\infty}\frac{1}{n}\log|\mathbb{B}^n_{2,V}[F]| &\le \limsup_{n\to\infty}\frac{1}{n}\log|\mathbb{B}^n_{2,V}[[\alpha_+,\infty)]| + \limsup_{n\to\infty}\frac{1}{n}\log|\mathbb{B}^n_{2,V}[[0,\alpha_-]]| \\
&= -\mathcal{J}(1,\alpha_-) - \mathcal{J}(1,\alpha_+) \\
&\le -\inf_{x\in F}\mathcal{J}(1,x).
\end{aligned}$$

For the last inequality we used that $\mathcal{J}(1,y)\ge 0$ for all y (by the second property of Lemma 5.2.1) and that $\alpha_+,\alpha_-\in F$.
2. Case: $\alpha_- = m$ or $\alpha_+ = m$, then by the definition of those quantities we have that $m\in F$.

Now we analyse the asymptotic behaviour of $|\mathbb{B}_V^n|$. Fix $s < 0$, then

$$\begin{aligned}
\frac{1}{n}\log|\mathbb{B}_V^n| &= \frac{1}{n}\log\int_{\{x\in\mathbb{R}^n:\sum_{i=1}^n V(x_i)\le n\}} dx \\
&\le \frac{1}{n}\log\int_{\{x\in\mathbb{R}^n:\sum_{i=1}^n V(x_i)\le n\}} \exp\left(s\sum_{i=1}^n V(x_i) - sn\right)dx \\
&\le -s + \log\int_{x\in\mathbb{R}} \exp\big(sV(x)\big)dx.
\end{aligned}$$

Now, since $\mathbb{B}_{2,V}^n[F] \subseteq \mathbb{B}_V^n$,

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

For the last equality we used property 2 of Lemma 5.2.1 and since $m \in F$ we get $\mathcal{J}(1,m) = \inf_{x\in F}\mathcal{J}(1,x)$.
2. Step: Lower bound of the LDP. Fix an open set $U \subseteq \mathbb{R}_+$, we claim

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

Since $\mathcal{J}$ is convex, $\mathcal{J}$ is also continuous in $\mathbb{R}_+\times\mathbb{R}_+$. Further, due to property 5 of Lemma 5.2.1, $\mathcal{J}$ is monotone decreasing in the first component and hence for a given $\epsilon > 0$, we can find $z \in U$ and $y \in (0,1)$, such that there exists $\delta > 0$ with $y \in (\delta, 1-\delta)$ and $(z-\delta, z+\delta) \in U$ and

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

Let (s,t) be the unique maximizer in the definition of $\mathcal{J}(y,z)$ (see the Remark after Lemma 5.2.1) and define

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

By construction A_δ^n is a subset of $\mathbb{B}_{2,V}^n[U]$, since for $x \in A_\delta^n$ the first condition implies $x \in \mathbb{B}_V^n$ and the second guarantees $x \in U$. Now we distinguish two cases for t, we start with the assumption $t < 0$. Then

$$
\begin{aligned}
|\mathbb{B}^n_{2,V}[U]| &\geq \int_{A^n_\delta} dx \\
&= \int_{A^n_\delta} (Z_{s,t})^n e^{-s\sum_{i=1}^n V(x_i) - tx_i^2} \prod_{i=1}^n \frac{1}{Z_{s,t}} e^{sV(x_i)+tx_i^2} dx \\
&= \exp(n(\log Z_{s,t} - s(y-\delta) - t(z-\delta))) \int_{A^n_\delta} \prod_{i=1}^n \frac{1}{Z_{s,t}} e^{sV(x_i)+tx_i^2} dx.
\end{aligned}
$$

If we assume $t \geq 0$, then we can have essentially the same bound, but instead of $(z-\delta)$ in the exponent, we get the factor $(z+\delta)$ in this case.

Now we analyse the integral. In order to do so, we consider an iid sequence $\{\Xi_i\}_{i\in\mathbb{N}}$ all distributed with respect to the density

$$
\frac{1}{Z_{s,t}} e^{sV(x)+tx^2}.
$$

Then by Lemma 5.2.1, $\mathbb{E}[V(\Xi_i)] = y$ and $\mathbb{E}[\Xi_i^2] = z$. With this notation we get

$$
\begin{aligned}
\int_{A^n_\delta} \prod_{i=1}^n \frac{1}{Z_{s,t}} e^{sV(x_i)+tx_i^2} dx &= \mathbb{P}\big[(\Xi_1, \ldots, \Xi_n) \in A^n_\delta\big] \\
&= \mathbb{P}\left[y - \delta < \frac{1}{n}\sum_{i=1}^n V(\Xi_i) < y + \delta, z - \delta < \frac{1}{n}\sum_{i=1}^n \Xi_i^2 < z + \delta\right]
\end{aligned}
$$

converging to 1 as $n \to \infty$ by the weak law of large numbers. Using this result leads to

$$
\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log |\mathbb{B}^n_{2,V}[U]| &\geq \log Z_{s,t} - s(y-\delta) - t(z \pm \delta) \\
&= \log \int_{\mathbb{R}} e^{sV(x)+tx^2} dx - sy - tz + \delta s \pm t\delta \\
&= -\mathcal{J}(y,z) + \delta s \pm t\delta \\
&\geq \inf_{x\in U} \mathcal{J}(1,x) - \epsilon + \delta s \pm t\delta.
\end{aligned}
$$

Now we let $\epsilon, \delta \to 0$, showing the lower bound.
3. Step: We study the asymptotics of $|\mathbb{B}^n_V|$ using the bounds we found in the previous steps. For this note that

$$
|\mathbb{B}^n_V| = |\mathbb{B}^n_{2,V}[[0,\infty)]| = |\mathbb{B}^n_{2,V}[(0,\infty)]|
$$

an thus

$$-\inf_{x\in(0,\infty)} \mathcal{J}(1,x) \le \lim_{n\to\infty} \frac{1}{n} \log |\mathbb{B}_V^n| \le -\inf_{x\in[0,\infty)} \mathcal{J}(1,x).$$

A short calculation shows that the limit in the middle exists, since the bounds coincide. This can be seen by taking a closer look at $\mathcal{J}(1,0)$, where we get

$$\begin{aligned}\mathcal{J}(1,0) &= \sup_{s,t\in\mathbb{R}} \left\{ s - \log \int_{\mathbb{R}} e^{sV(x)+tx^2} dx \right\} \\ &\ge \lim_{t\to-\infty} \left\{ -\log \int_{\mathbb{R}} e^{tx^2} dx \right\} \\ &= \infty.\end{aligned}$$

Hence the point $x = 0$ can be omitted in the infimization. In total we get the limit

$$\lim_{n\to\infty} \frac{1}{n} \log |\mathbb{B}_V^n| = -\inf_{x\in[0,\infty)} \mathcal{J}(1,x) = -\sup_{s<0} \left\{ s - \log \int_{\mathbb{R}} e^{sV(x)} dx \right\},$$

where we once more used property 2 of Lemma 5.2.1.
Finally, this establishes an LDP for $\{\frac{1}{n}\sum_{i=1}^n (X_i^n)^2\}_{n\in\mathbb{N}}$ at speed n and GRF

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

The claimed LDP for $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ at speed n and rate function $J_{X,V}$ follows by applying the contraction principle for the mapping $x \mapsto \sqrt{x}$. □

This finishes the situation of symmetric Orlicz balls, we continue with the general case, which is, as we will see in a moment, not so general as it may appear.

5.2.2 Generalized Orlicz balls

Consider the generalized Orlicz ball $\mathbb{B}^n_{V_1,\dots,V_n}$, where we choose suitable functions $V_1, V_2, ..$ such that another Orlicz function $\bar{V}$ exists with

$$\frac{1}{n} \log \left(\frac{|\mathbb{B}^n_{V_1,\dots,V_n} \triangle \mathbb{B}^n_{\bar{V}}|}{|\mathbb{B}^n_{V_1,\dots,V_n} \cup \mathbb{B}^n_{\bar{V}}|} \right) \longrightarrow -\infty, \tag{5.2.2}$$

where $A \triangle B := A\backslash B \cup B\backslash A$ denotes the symmetric difference of sets A and B. This condition basically means that $\mathbb{B}^n_{V_1,\dots,V_n}$ is very similar to a symmetric Orlicz ball on an exponential scale. Considering the important role of exponential equivalence in large deviation theory one is not surprised that we are able to establish an LDP for random variables uniformly distributed on such $\mathbb{B}^n_{V_1,\ldots,V_n}$. We start with the construction of the corresponding uniform distribution. For a measurable set $A \subseteq \mathbb{R}^n$ define

$$\mu(A) := \frac{|A \cap \mathbb{B}^n_{V_1,\ldots,V_n}|}{|\mathbb{B}^n_{V_1,\ldots,V_n}|}.$$

Analogously to the symmetric case the denominator is finite and not zero, hence μ is a probability distribution on $\mathbb{R}^n$ and we can construct a random variable $X^{(n)} \sim \mu = Unif(\mathbb{B}^n_{V_1,\ldots,V_n})$. We can now state the main result in this subsection.

Theorem 5.2.3. *(Lemma 3.13 in [18])*
Given Orlicz functions $\{V_i\}_{i\in\mathbb{N}}$ and $\bar{V}$ that satisfy Condition 5.2.2, suppose $X^{(n)} \sim Unif(\mathbb{B}^n_{V_1,\ldots,V_n})$ and there exists a constant $c_V \in (0,\infty)$ such that for $x > c_V$, $V_i(x) > x^2$ for $i \in \mathbb{N}$ and $\bar{V}(x) > x^2$. Then, $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption A with $\mathcal{J}_{X,\bar{V}}$, defined as in Theorem 5.2.2, but with V replaced with $\bar{V}$.*

Proof. Consider sequences $\{\bar{X}^{(n)}\}_{n\in\mathbb{N}}$ and $\{U^{(n)}\}_{n\in\mathbb{N}}$ with $\bar{X}^{(n)} \sim Unif(\mathbb{B}^n_{\bar{V}})$ and $U^{(n)} \sim Unif(\mathbb{B}^n_{\bar{V}} \cup \mathbb{B}^n_{V_1,\ldots,V_n})$ (such a distribution can be constructed in the same way as $X^{(n)}$ and $\bar{X}^{(n)}$ respectively). Assume that all involved sequences are independent and constructed on some common probability space with measure $\mathbb{P}$. Then we define the two quantities

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

Fist note that the conditional distribution $\mathbb{P}\big[U^{(n)} \in \cdot | U^{(n)} \in \mathbb{B}^n_{V_1,\ldots V_n}\big]$ is the uniform distribution on $\mathbb{B}^n_{V_1,\ldots V_n}$. To see this, we take a measurable $A \subseteq \mathbb{B}^n_{V_1,\ldots V_n}$, then

$$\begin{aligned} \mathbb{P}\big[U^{(n)} \in A | U^{(n)} \in \mathbb{B}^n_{V_1,\ldots V_n}\big] &= \frac{\mathbb{P}\big[U^{(n)} \in A \cap \mathbb{B}^n_{V_1,\ldots V_n}\big]}{\mathbb{P}\big[U^{(n)} \in \mathbb{B}^n_{V_1,\ldots V_n}\big]} \\ &= \frac{\big|A \cap \mathbb{B}^n_{V_1,\ldots V_n} \cap \big(\mathbb{B}^n_{V_1,\ldots V_n} \cup \mathbb{B}^n_{\bar{V}}\big)\big|}{\big|\mathbb{B}^n_{V_1,\ldots V_n} \cap \big(\mathbb{B}^n_{V_1,\ldots V_n} \cup \mathbb{B}^n_{\bar{V}}\big)\big|} \frac{\big|\mathbb{B}^n_{V_1,\ldots V_n} \cup \mathbb{B}^n_{\bar{V}}\big|}{\big|\mathbb{B}^n_{V_1,\ldots V_n} \cap \big(\mathbb{B}^n_{V_1,\ldots V_n} \cup \mathbb{B}^n_{\bar{V}}\big)\big|} \\ &= \frac{\big|A \cap \mathbb{B}^n_{V_1,\ldots V_n}\big|}{\big|\mathbb{B}^n_{V_1,\ldots V_n}\big|}. \end{aligned}$$

Next we look at the distribution of $W^{(n)}$. Again, let $A \subseteq \mathbb{R}$ be some measurable set.

Then

$$
\begin{aligned}
\mathbb{P}\big[W^{(n)} \in A\big] &= \mathbb{P}\big[W^{(n)} \in A | U^{(n)} \in \mathbb{B}^n_{V_1,\dots V_n}\big]\mathbb{P}\big[U^{(n)} \in \mathbb{B}^n_{V_1,\dots V_n}\big] \\
&\quad + \mathbb{P}\big[W^{(n)} \in A | U^{(n)} \notin \mathbb{B}^n_{V_1,\dots V_n}\big]\mathbb{P}\big[U^{(n)} \notin \mathbb{B}^n_{V_1,\dots V_n}\big] \\
&= \mathbb{P}\big[U^{(n)} \in A | U^{(n)} \in \mathbb{B}^n_{V_1,\dots V_n}\big]\mathbb{P}\big[U^{(n)} \in \mathbb{B}^n_{V_1,\dots V_n}\big] \\
&\quad + \mathbb{P}\big[X^{(n)} \in A | U^{(n)} \notin \mathbb{B}^n_{V_1,\dots V_n}\big]\mathbb{P}\big[U^{(n)} \notin \mathbb{B}^n_{V_1,\dots V_n}\big] \\
&= \mathbb{P}\big[X^{(n)} \in A\big]\mathbb{P}\big[U^{(n)} \in \mathbb{B}^n_{V_1,\dots V_n}\big] + \mathbb{P}\big[X^{(n)} \in A\big]\mathbb{P}\big[U^{(n)} \notin \mathbb{B}^n_{V_1,\dots V_n}\big] \\
&= \mathbb{P}\big[X^{(n)} \in A\big].
\end{aligned}
$$

We see that $W^{(n)} \overset{(d)}{=} X^{(n)}$ and hence $W^{(n)} \sim Unif(\mathbb{B}^n_{V_1,\dots,V_n})$. Basically the same calculation can be applied to $\bar{W}^{(n)}$, establishing $\bar{W}^{(n)} \overset{(d)}{=} \bar{X}^{(n)}$. The desired result now follows by showing exponential equivalence of $\{\|W^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ and $\{\|\bar{W}^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$. This follows by direct computation. Before we do this, we define $\kappa := \sqrt{1+c_V^2}$. Then $\mathbb{B}^n_{V_1,\dots,V_n} \subseteq \kappa\sqrt{n}\mathbb{B}^n_2$, since if $x \in \mathbb{B}^n_{V_1,\dots,V_n}$, we have

$$
\begin{aligned}
\sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_i^2 \mathbb{1}_{x_i \le c_V} + \sum_{i=1}^n x_i^2 \mathbb{1}_{x_i > c_V} \\
&\le \sum_{i=1}^n c_V^2 + \sum_{i=1}^n V(x_i) \\
&\le n(c_V^2 + 1) \\
&= n\kappa^2.
\end{aligned}
$$

In the same way we are also able to show $\mathbb{B}^n_{\bar{V}} \subseteq \kappa\sqrt{n}\mathbb{B}^n_2$. This leads to the estimate

$$
\begin{aligned}
\left|\frac{\|\bar{W}^{(n)}\|_2}{\sqrt{n}} - \frac{\|W^{(n)}\|_2}{\sqrt{n}}\right| &= \frac{1}{\sqrt{n}}\left|\left(\|\bar{W}^{(n)}\|_2 - \|W^{(n)}\|_2\right)\mathbb{1}_{\{U^{(n)} \in \mathbb{B}^n_{V_1,\dots,V_n} \cap \mathbb{B}^n_{\bar{V}}\}}\right| \\
&\quad + \frac{1}{\sqrt{n}}\left|\left(\|\bar{W}^{(n)}\|_2 - \|W^{(n)}\|_2\right)\mathbb{1}_{\{U^{(n)} \in \mathbb{B}^n_{V_1,\dots,V_n} \triangle \mathbb{B}^n_{\bar{V}}\}}\right| \\
&\le 0 + \frac{1}{\sqrt{n}}\|\bar{W}^{(n)} - W^{(n)}\|_2 \mathbb{1}_{\{U^{(n)} \in \mathbb{B}^n_{V_1,\dots,V_n} \triangle \mathbb{B}^n_{\bar{V}}\}} \\
&\le 2\kappa \mathbb{1}_{\{U^{(n)} \in \mathbb{B}^n_{V_1,\dots,V_n} \triangle \mathbb{B}^n_{\bar{V}}\}}.
\end{aligned}
$$

Here we used, that $\|\bar{W}^{(n)} - W^{(n)}\|_2 \le 2\kappa\sqrt{n}$ and on the set $\big[U^{(n)} \in \mathbb{B}^n_{V_1,\dots,V_n} \cap \mathbb{B}^n_{\bar{V}}\big]$ we have $W^{(n)} = \bar{W}^{(n)}$. Now we fix $\delta > 0$ and consider

$$\begin{aligned}
\frac{1}{n}\log \mathbb{P}\big[\big|\|W^{(n)}\|_2/\sqrt{n} - \|\bar{W}^{(n)}\|_2/\sqrt{n}\big| > \delta\big] &\leq \frac{1}{n}\log \mathbb{P}\big[2\kappa \mathbb{1}_{\{U^{(n)} \in \mathbb{B}^n_{V_1,\dots,V_n} \triangle \mathbb{B}^n_{\bar{V}}\}} > \delta\big] \\
&= \frac{1}{n}\log \mathbb{P}\big[U^{(n)} \in \mathbb{B}^n_{V_1,\dots,V_n} \triangle \mathbb{B}^n_{\bar{V}}\big] \\
&= \frac{1}{n}\log\Big(\frac{|\mathbb{B}^n_{V_1,\dots,V_n} \triangle \mathbb{B}^n_{\bar{V}}|}{|\mathbb{B}^n_{V_1,\dots,V_n} \cup \mathbb{B}^n_{\bar{V}}|}\Big).
\end{aligned}$$

Thus, if we apply the limsup on both sides, we obtain the desired exponential equivalence. □

5.3 Gibbs measures

In this section we introduce the notion of Gibbs measures and deduce Assumption A* for a certain sequence of random variables. We start with some notation and definitions. The following is based on [10]. Consider a set of "particles" $\mathbf{x}^n := (\mathbf{x_1}, \dots, \mathbf{x_n})$ in $\mathbb{R}^d$, which are exposed to some external force $V : \mathbb{R}^d \to (-\infty, \infty]$ and some kind of pairwise interaction (e.g. electrons repulsing each other) $W : \mathbb{R}^d \times \mathbb{R}^d \to (-\infty, \infty]$. The behaviour of this system can be modelled via the Hamilton operator

$$\begin{aligned}
H_n : \mathbb{R}^{d\times n} &\to (-\infty, \infty] \\
\mathbf{x}^n &\mapsto \frac{1}{n}\sum_{i=1}^n V(\mathbf{x}_i) + \frac{1}{2n^2}\sum_{i,j=1, i\neq j}^n W(\mathbf{x}_i, \mathbf{x}_j).
\end{aligned}$$

Further, for such an n-tuple of particles we consider the respective sequence of empirical measures

$$\begin{aligned}
L_n : \mathbb{R}^{d\times n} &\to P(\mathbb{R}^d) \\
\mathbf{x}^n &\mapsto \frac{1}{n}\sum_{i=1}^n \delta_{\mathbf{x}_i}.
\end{aligned}$$

Later we make use of the notation $L_n(\mathbf{x}^n, A) = \frac{1}{n}\sum_{i=1}^n \delta_{\mathbf{x}_i}(A)$, for $\mathbf{x}^n \in \mathbb{R}^{d\times n}$ and $A \in \mathcal{B}(\mathbb{R}^d)$.

Lemma 5.3.1. *The function $L_n : \mathbb{R}^{d\times n} \to P(\mathbb{R}^d)$ is a continuous mapping.*

Proof. Since we consider a mapping between metric spaces, continuity can be decided via sequences. Further we use the bounded Lipschitz-metric on $P(\mathbb{R})$ and we assume

$d = 1$, the general case works analogously. Let $\{x^m\}_{m\in\mathbb{N}}$ be a sequence of reals with $x^m \to x$. We claim

$$L_n(x^m,.) \xrightarrow{d_{bL}} L_n(x,.).$$

Denote by $BL(\mathbb{R})$ the set of bounded Lipschitz continuous functions with Lipschitz constant equal to 1. Then we need to consider

$$\begin{aligned}\sup_{f\in BL(\mathbb{R})}\left|\int_{\mathbb{R}} f(t)dL_n(x^m,t) - \int_{\mathbb{R}} f(t)dL_n(x,t)\right| &= \sup_{f\in BL(\mathbb{R})}\left|\frac{1}{n}\sum_{i=1}^n f(x_k^m) - \frac{1}{n}\sum_{i=1}^n f(x_k)\right| \\ &\le \sup_{f\in BL(\mathbb{R})}\frac{1}{n}\sum_{i=1}^n |f(x_k^m) - f(x_k)| \\ &\le \frac{1}{n}\sum_{i=1}^n |x_k^m - x_k| \xrightarrow{m\to\infty} 0.\end{aligned}$$

In the last step we used the uniform Lipschitz-continuity of the involved function space. □

Let $\{\beta_n\}_{n\in\mathbb{N}}$ be a sequence of positive real numbers with $\beta_n \to \infty$ and ℓ, a σ-finite measure on $\mathbb{R}$ (e.g. the Lebesgue measure). Then we use our Hamiltonian to define "Gibbs-measures" $P_n \in P(\mathbb{R}^{d\times n})$ as

$$P_n(d\mathbf{x}_1,...,d\mathbf{x}_n) := \frac{\exp(-\beta_n H_n(\mathbf{x}_1,...,\mathbf{x}_n))}{Z_n}\ell(d\mathbf{x}_1)...\ell(d\mathbf{x}_n),$$

provided V and W are chosen in a way such that

$$Z_n := \int_{\mathbb{R}^d}...\int_{\mathbb{R}^d}\exp(-\beta_n H_n(\mathbf{x}_1,...,\mathbf{x}_n))\ell(d\mathbf{x}_1)...\ell(d\mathbf{x}_n)$$

is finite.

Definition 5.3.1. *Given measurable spaces $(S,\mathcal{F})$ and $(\tilde{S},\tilde{\mathcal{F}})$, a measurable mapping $f : S \to \tilde{S}$ and a measure $\mu : \mathcal{F} \to [0,\infty]$, the pushforward of μ is the measure induced on $(\tilde{S},\tilde{\mathcal{F}})$ by μ under f, that is, the measure $f_\#\mu : \tilde{\mathcal{F}} \to [0,\infty]$ given by*

$$(f_\#\mu)(B) = \mu(f^{-1}(B)) \quad \textit{for } B \in \tilde{\mathcal{F}}.$$

In other words $f_\#\mu$ is the image measure of μ under f.

Finally, we introduce probability measures Q_n by pushing P_n forward under the mapping L_n, i.e.

$$Q_n(A) := P_n(L_n \in A)$$

for a Borel set $A \subseteq P(\mathbb{R}^d)$, where we equip $P(\mathbb{R}^d)$ with the weak topology and note that $L_n : \mathbb{R}^{d\times n} \to P(\mathbb{R}^d)$ is continuous (see Lemma 5.3.1) and hence measurable.

Theorem 5.3.2. *(Theorem 2.7 and Lemma 2.6 from [18])*
Suppose the following assumptions hold for the potentials V and W:

1. *V and W are lower semicontinuous on the respective sets on which they are finite.*
2. *There exists $1 > a \geq 0$ and $c \in \mathbb{R}$ such that V satisfies*

$$\int_{\mathbb{R}^d} e^{-(1-a)V(x)}\ell(dx) < \infty$$
$$\inf_{x\in\mathbb{R}^d} V(x) > c, \quad \inf_{(x,y)\in\mathbb{R}^d\times\mathbb{R}^d} [W(x,y) + a(V(x) + V(y))] > c.$$

3. *There exists a set $A \in \mathcal{B}(\mathbb{R}^d)$ with $\ell(A) > 0$ such that*

$$\int_{A\times A} (V(x) + V(y) + W(x,y))\ell(dx)\ell(dy) < \infty.$$

4. *For all $\lambda \in \mathbb{R}^d$, we have*

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \exp\left[\lambda(x^2 + y^2) - V(x) - V(y) - W(x,y)\right]\ell(dx)\ell(dy) < \infty.$$

Then $\{Q_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $P_2(\mathbb{R}^d)$ equipped with the 2-Wasserstein topology, at speed n, with GRF $\mathcal{J}_* : P_2(\mathbb{R}^d) \to [0,\infty]$ defined by

$$\mathcal{J}_*(\mu) := \mathcal{J}(\mu) - \inf_{\nu\in P_2(\mathbb{R}^d)} \mathcal{J}(\nu),$$
$$\mathcal{J}(\mu) := H(\mu|\ell) + \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W(x,y)\mu(dx)\mu(dy) + \int_{\mathbb{R}^d} V(x)\mu(dx),$$

where H denotes the relative entropy.

Remark. In [10] the authors proved Theorem 5.3.2 using Assumptions 1,2,3 and a different one (they refer to it as "Assumption B", which is not so easy to handle). Lemma 2.6 in [10] proves that 1-3 and 4 imply Assumption B.

Remark. In [10] a more abstract setting is treated, where one considers a generalization of Wasserstein spaces, the result in Theorem 5.3.2 is a special case.

The previous theorem can be used to establish an LDP for norms of random variables distributed according to the Gibbs measure. To simplify formulas, we restrict to the case $d = 1$.

Corollary. (Proposition 3.15 in [18])
For $n \in \mathbb{N}$, suppose $X^{(n)}$ is drawn from P_n, and suppose that the assumptions of Theorem 5.3.2 hold. Then, $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption A* with GRF

$$J_X(x) = \inf\left\{\mathcal{J}_*(\mu) : \mu \in P_2(\mathbb{R}), x = \sqrt{m_2(\mu)}\right\}, \qquad x \geq 0.$$

Proof. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space such that $X^{(n)} : \Omega \to \mathbb{R}^n$ has distribution P_n. First, note the relation

$$\begin{aligned}\frac{\|X^{(n)}\|_2^2}{n} &= \int_{\mathbb{R}} x^2 dL_n(X^{(n)}, x) \\ &= m_2(L_n(X^{(n)}, .)),\end{aligned}$$

and recall that $m_2 : P_2(\mathbb{R}) \to [0, \infty)$ is continuous by the very definition of the 2-Wasserstein topology. Then, since

$$\mathbb{P}\left[L_n(X^{(n)}, .) \in A\right] = P_n(L_n \in A) = Q_n(A)$$

for $A \in \mathcal{B}(P_2(\mathbb{R}))$, we can apply the contraction principle to establish an LDP for $\{\|X^{(n)}\|_2^2/n\}_{n\in\mathbb{N}}$ at speed n. By taking the square root we finally get the desired LDP with the claimed GRF. □

6 Appendix

Theorem 6.0.1. *(Hall marriage theorem)*
Let (G, E) be a bipartite graph with $G = X \cup Y$ and $|X| = |Y|$ be the bipartition classes (i.e. E consists only of edges between X and Y). For $A \subseteq X$ we denote by $K(A)$ all neighbours of A, i.e.

$$K(A) = \{y \in Y \mid (x, y) \in E\}.$$

Then there exists a bijective function $f : X \to Y$ such that $x \sim f(x)$ iff $|K(A)| \geq |A|$.

Definition 6.0.1. *Let $\mathcal{X}$ be a vector space over the real numbers. Then $\mathcal{X}'$ denotes the algebraic dual, i.e. the space of linear functions mapping from $\mathcal{X}$ to $\mathbb{R}$. Moreover, if one chooses a topology on $\mathcal{X}$, we consider the topological dual $\mathcal{X}^* \subseteq \mathcal{X}'$, where we additionally assume continuity.*

Definition 6.0.2. *Let $\mathcal{X}$ be a topological vector space over the real numbers. We call $\mathcal{Y} \subseteq \mathcal{X}'$ a separating subset, if for any $x \in \mathcal{X}$ there exists $y' \in \mathcal{Y}$ with $\langle y', x \rangle \neq 0$. The $\mathcal{Y}$-topology is the coarsest topology on $\mathcal{X}$, such that all functionals in $\mathcal{Y}$ are continuous.*

Theorem 6.0.2. *Let $\mathcal{X}$ be a vector space over the real numbers, and $\mathcal{Y} \subseteq \mathcal{X}'$ be a separating subspace. Then the $\mathcal{Y}$-topology makes $\mathcal{X}$ into a locally convex topological vector space with $\mathcal{X}^* = \mathcal{Y}$.*

The case when $\mathcal{Y} = \mathcal{X}^*$ is of special interest, as we will see in the next theorem.

Definition 6.0.3. *(weak*-topology)*
Let $(\mathcal{X}, \|\cdot\|)$ be a normed vector space over $\mathbb{R}$. The weak topology is defined as the coarsest topology, such that all mappings $\hat{x} \in \mathcal{X}^{**}$ with $\hat{x}(f) = f(x)$ for some $x \in \mathcal{X}$, are continuous.*

Theorem 6.0.3. *(Version of Banach-Alaoglu theorem)*
Let $(\mathcal{X}, \|\cdot\|)$ be a normed vector space over $\mathbb{R}$. Then the set

$$M = \{\phi \in \mathcal{X}^* \mid \|\phi\| \leq 1\}$$

is compact with respect to the weak topology. If $\mathcal{X}$ is separable, then every sequence $\{x_n\}_{n\in\mathbb{N}} \subseteq \mathcal{X}^*$ has a weak* convergent subsequence.*

Definition 6.0.4. *Let (M, d) be a metric space. We say that M is compact if for every family $\{O_i\}_{i\in I}$ of open sets with $M = \bigcup_{i\in I} O_i$ there exists a finite subfamily $O_{i_1}, \ldots, O_{i_n}$ with $M = \bigcup_{j=1}^{n} O_{i_j}$.*

Definition 6.0.5. *Let (M, τ) be a Hausdorff space (e.g. if (M, d) is a compact metric space and τ is induced by d) and $\mathcal{B}$ the Borel σ algebra. A Radon measure μ on $\mathcal{B}$ is inner regular and locally finite, i.e.*

for all measurable sets A: $\mu(A) = \sup\{\mu(K) \mid K \subseteq A, K \text{ compact}\}$, and

for every $m \in M$ there exists a neighbourhood U of m with $\mu(U) < \infty$.

Theorem 6.0.4. *(Version of Riesz-Markov representation theorem)*
Let (M, d) be a compact metric space and let $C(M)$ denote the continuous functions mapping from M to $\mathbb{R}$. Suppose we are given a linear functional $I : C(M) \to \mathbb{R}$ with

$$f \geq 0 \implies I(f) \geq 0.$$

Then there exists a unique Radon measure on M representing I, i.e. for all $f \in C(M)$

$$I(f) = \int_M f(x) d\mu(x).$$

Theorem 6.0.5. *(Cantor's theorem)*
Let $\{K_n\}_{n\in\mathbb{N}}$ be a sequence of non-empty compact sets in a metric space $(\mathcal{X}, d)$ with $K_{n+1} \subseteq K_n$ for all $n \in \mathbb{N}$. Then $\bigcap_{n\in\mathbb{N}} K_n$ is non-empty and compact.

Definition 6.0.6. *Given a topological space $\mathcal{X}$ and a point $x \in \mathcal{X}$, a basis of open neighbourhoods $\mathcal{B}(x)$ satisfies the following properties*

(1) For any $U \in \mathcal{B}(x), x \in U$.

(2) For any $U_1, U_2 \in \mathcal{B}(x) : \exists U_3 \in \mathcal{B}(x)$ such that $U_3 \subseteq U_1 \cap U_2$.

(3) If $y \in U \in \mathcal{B}(x)$, then $\exists W \in \mathcal{B}(y)$ such that $W \subseteq U$.

Theorem 6.0.6. *(Hausdorff's criterion)*
Let τ and τ' two topologies on the same set $\mathcal{X}$. For each $x \in \mathcal{X}$, let $\mathcal{B}(x)$ a basis of neighbourhoods of x in $(\mathcal{X}, \tau)$ an $\mathcal{B}'(x)$ a Basis of neighbourhoods of x in (X, τ'). Then $\tau \subseteq \tau'$ iff $\forall x \in X, \forall U \in \mathcal{B}(x) : \exists V \in \mathcal{B}'(x)$ such that $x \in V \subseteq U$.

Theorem 6.0.7. *(Version of Radon-Nikodym Theorem)*
Let $(\Sigma, \mathcal{F})$ be a measurable space and let μ, ν be probability measures on $\mathcal{F}$. Assume that ν is absolutely continuous w.r.t. μ, denoted by $\nu \ll \mu$, i.e.

$$\forall F \in \mathcal{F} : \mu(F) = 0 \implies \nu(F) = 0.$$

Then there exists a μ almost surely unique measurable function $f : \Sigma \to \mathbb{R}_+$, such that for all $F \in \mathcal{F}$

$$\nu(F) = \int_F f(x) d\mu(x).$$

We often denote f by $d\nu/d\mu$ and call it the Radon-Nikodym derivative.

Remark. In the case of $\nu \ll \mu$ and $\mu \ll \nu$, also $d\mu/d\nu$ exists and we have the nice relation

$$\frac{d\mu}{d\nu} = \left(\frac{d\nu}{d\mu}\right)^{-1}.$$

Further, if η is another probability measure on $(\Sigma, \mathcal{F})$ with $\nu \ll \mu \ll \eta$, then $d\eta/d\nu$ exists and

$$\frac{d\eta}{d\nu} = \frac{d\eta}{d\mu}\frac{d\mu}{d\nu}.$$

Theorem 6.0.8. *(transformation rule for probability measures)*
Let μ be a Borel probability measure on $\mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ a measurable function. For a measurable function $f : \mathbb{R} \to \mathbb{R}$ and a measurable set $A \subseteq \mathbb{R}$ it holds that

$$\int_{g^{-1}(A)} f \circ g(x) d\mu(x) = \int_A f(x) d(\mu \circ g^{-1})(x),$$

provided at least one of those integrals exist.

Theorem 6.0.9. *(Theorem D.4 in [9])*
Let $\{\Sigma_i\}_{i=1}^N$ be metric spaces and N be either finite or $N = \infty$. Then

(a) $\bigotimes_{i=1}^N \mathcal{B}_{\Sigma_i} \subseteq \mathcal{B}_{\times_{i=1}^N \Sigma_i}$.

(b) *If Σ_i are separable, then* $\bigotimes_{i=1}^N \mathcal{B}_{\Sigma_i} = \mathcal{B}_{\times_{i=1}^N \Sigma_i}$.

Definition 6.0.7. *(Tightness)*
A probability measure μ on a metric space Σ is tight if for each $\epsilon > 0$, there exists a compact set $K_\epsilon \subseteq \Sigma$ such that $\mu(K_\epsilon^c) < \epsilon$. A family of probability measures $\{\mu_i\}_{i \in I}$ on the metric space Σ is called a tight family if the set K_ϵ may be chosen independently of i.

Theorem 6.0.10. *(Theorem D.7 in [9])*
Each probability measure on a Polish space Σ is tight.

Theorem 6.0.11. *(Prohorov's theorem, D.9 in [9])*
Let Σ be Polish, and let $\Gamma \subseteq P(\Sigma)$. Then $\bar{\Gamma}$ is compact iff Γ is tight.

Theorem 6.0.12. *(Portmanteau Theorem, Theorem D.10 in [9])*
Let Σ be Polish with Borel sigma algebra $\mathcal{B}_\Sigma$. The following statements are equivalent

(1) $\mu_n \to \mu$ *weakly.*

(2) $\forall g : \Sigma \to \mathbb{R}$ *bounded and uniformly continuous,* $\lim_{n\to\infty} \int_\Sigma g(x)d\mu_n(x) = \int_\Sigma g(x)d\mu(x)$.

(3) $\forall F \subseteq \Sigma$ *closed,* $\limsup_{n\to\infty} \mu_n(F) \le \mu(F)$.

(4) $\forall G \subseteq \Sigma$ *open,* $\liminf_{n\to\infty} \mu_n(G) \ge \mu(G)$.

(5) $\forall A \subseteq \mathcal{B}_\Sigma$ *which is a continuity set, i.e., such that* $\mu(\overline{A}\backslash\mathring{A}) = 0$, $\lim_{n\to\infty} \mu_n(A) = \mu(A)$.

Remark. There are several other characterisations of weak convergence. In particular, criteria using certain function spaces can be convenient. Beside the set of continuous and bounded functions one may use

$$\{f : \Sigma \to \mathbb{R} \text{ Lipschitz continuous } : ||f||_\infty + ||f||_{Lip} \le 1\},$$

where $||f||_{Lip}$ denotes the Lipschitz constant.

Theorem 6.0.13. *(Slutsky's theorem)*
Let $\{X_n\}_{n\in\mathbb{N}}, \{Y_n\}_{n\in\mathbb{N}}, \{Z_n\}_{n\in\mathbb{N}}$ be sequences of $\mathbb{R}$-valued random variables with

$$X_n \xrightarrow{(d)} X$$
$$(Y_n, Z_n) \xrightarrow{\mathbb{P}} (Y, Z)$$

for some random variables X, Y, Z. Then $Z_n + Y_n X_n \overset{(d)}{\to} Z + YX$.

Theorem 6.0.14. *Suppose a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and independent random variables $X, Y : \Omega \to \mathbb{R}$. Further let $h : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a measurable function such that $\mathbb{E}[|h(X,Y)|] < \infty$. Then*

$$\mathbb{E}[h(X,Y)|Y] = H(Y),$$

where $H(y) = \mathbb{E}[h(X,y)], y \in \mathbb{R}$.

Theorem 6.0.15. *(Theorem of Ionescu-Tulcea)*
For all $n \in \mathbb{N}$, let $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ be a probability space. Then there exists an unique probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ where

$$\Omega := \bigtimes_{n\in\mathbb{N}} \Omega_n \quad \textit{and} \quad \mathcal{F} := \bigotimes_{n\in\mathbb{N}} \mathcal{F}_n$$

such that

$$\mathbb{P}\Big[A_1 \times \ldots \times A_n \times \bigtimes_{k=n+1}^{\infty} \Omega_k\Big] = \prod_{k=1}^{n} \mathbb{P}_k[A_k]$$

for $A_k \in \mathcal{F}_k$, $k = 1, \ldots, n$.

Since we are often dealing with a certain distribution μ on the real numbers (or on $\mathbb{R}^d$, for some d), we usually want to work with random variables distributed according μ.

Theorem 6.0.16. *Given a distribution function F, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a real random variable $X : \Omega \to \mathbb{R}$ such that $\mathbb{P}[X \le x] = F(x)$ for all $x \in \mathbb{R}$.*

Proof. (Theorem 3.1.7)
We see that $\mathbb{I}' = \mathbb{I} \circ g$ is again a rate function, because $\mathbb{I}' \ge 0$ and for $\alpha < \infty$ we have

$$\begin{aligned}\{y \in \mathcal{Y} : \mathbb{I}'(y) \le \alpha\} &= \{y \in \mathcal{Y} : \mathbb{I}(g(y)) \le \alpha\} \\ &= g^{-1}\big(\{y \in \mathcal{Y} : \mathbb{I}(y) \le \alpha\}\big) \\ &= g^{-1}(\psi_{\mathbb{I}}(\alpha)).\end{aligned}$$

The latter set is the preimage of a closed set under a continuous bijection and hence closed. Due to exponential tightness it suffices to show a weak LDP for $\{\nu_\epsilon\}_{\epsilon>0}$ with rate function $\mathbb{I}'$. For the upper bound, we fix a compact set $K \subseteq \mathcal{Y}$ and apply the large deviation upper bound for $\{\nu_\epsilon \circ g^{-1}\}_{\epsilon>0}$ on the compact set $g(K)$

$$\begin{aligned}\limsup_{\epsilon\to 0} \epsilon \log \nu_\epsilon(K) &= \limsup_{\epsilon\to 0} \epsilon \log \nu_\epsilon(g^{-1} \circ g(K)) \\ &\le - \inf_{x\in g(K)} \mathbb{I}(x) \\ &= - \inf_{x\in K} \mathbb{I}(g(x)).\end{aligned}$$

For the lower bound we fix $y \in \mathcal{Y}$ with $\mathbb{I}'(y) = \mathbb{I}(g(y)) =: \alpha < \infty$, and a neighbourhood

G of y. Since $\{\nu_\epsilon\}_{\epsilon>0}$ is exponentially tight, there exists a a compact set $K_\alpha \subseteq \mathcal{Y}$ with

$$\limsup_{\epsilon\to 0} \epsilon \log \nu_\epsilon(K_\alpha^c) < -\alpha.$$

Since g is a bijection we have $g(K_\alpha^c) = g(K_\alpha)^c$, also $g(K_\alpha)$ is compact. The lower large deviation bound for $\{\nu_\epsilon \circ g^{-1}\}_{\epsilon>0}$ applied to the open set $g(K_\alpha^c)$ yields

$$-\inf_{x\in g(K_\alpha^c)} \mathbb{I}'(x) \leq \liminf_{\epsilon\to 0} \epsilon \log \nu_\epsilon \circ g^{-1}(g(K_\alpha^c)) < -\alpha.$$

In particular, we get $\inf_{x\in g(K_\alpha^c)} \mathbb{I}(x) = \inf_{x\in K_\alpha^c} \mathbb{I}(g(x)) > \alpha$ and hence $y \in K_\alpha$. Since g is a continuous bijection, the restriction

$$g_{|K_\alpha} : K_\alpha \to g(K_\alpha)$$

becomes a homeomorphism, i.e. the inverse $g^{-1}_{|K_\alpha}$ is also continuous. The set $G \cap K_\alpha$ is open in the induced topology on K_α and hence $g(G\cap K_\alpha)$ is open in the induced topology on $g(K_\alpha)$. We can use this observation to find a neighbourhood G' of $g(y)$ in $\mathcal{X}$ with

$$G' \subseteq g(G \cap K_\alpha) \cup g(K_\alpha)^c = g(K_\alpha^c \cup G).$$

Applying the measure $\nu_\epsilon \circ g^{-1}$ yields

$$\begin{aligned}\nu_\epsilon \circ g^{-1}(G') &\leq \nu_\epsilon(K_\alpha^c \cup G)\\ &\leq \nu_\epsilon(K_\alpha^c) + \nu_\epsilon(G).\end{aligned}$$

We now make use of the trivial inequality $\log(a+b) \leq \log 2 + \log a \vee \log b$, where we get

$$\begin{aligned}\liminf_{\epsilon\to 0} \epsilon \log \nu_\epsilon \circ g^{-1}(G') &\leq \liminf_{\epsilon\to 0} \epsilon \log \left(\nu_\epsilon(G) + \nu_\epsilon(K_\alpha^c)\right)\\ &\leq \liminf_{\epsilon\to 0} \epsilon \log \nu_\epsilon(G) \vee \log \nu_\epsilon(K_\alpha^c),\end{aligned}$$

where the $\log 2$ term vanishes in the $\liminf$. The latter expression can further bounded from above, namely

$$\liminf_{\epsilon\to 0} \epsilon \log \nu_\epsilon(G) \vee \log \nu_\epsilon(K_\alpha^c) \leq \liminf_{\epsilon\to 0} \epsilon \log \nu_\epsilon(G) \vee \limsup_{\epsilon\to 0} \log \nu_\epsilon(K_\alpha^c).$$

In total, using the large deviation lower bound for $\nu_\epsilon \circ g^{-1}$, we receive

$$
\begin{aligned}
-\mathbb{I}'(y) &\leq -\inf_{x \in G'} \mathbb{I}(x) \\
&\leq \liminf_{\epsilon \to 0} \epsilon \log \nu_\epsilon \circ g^{-1}(G') \\
&\leq \liminf_{\epsilon \to 0} \epsilon \log \nu_\epsilon(G) \vee \limsup_{\epsilon \to 0} \log \nu_\epsilon(K_\alpha^c).
\end{aligned}
$$

Since $-\mathbb{I}'(y) = -\alpha$ by the definition of α and using the exponential tightness of $\{\nu_\epsilon\}_{\epsilon>0}$, we can neglect the right-hand term in the maximum. For an open set $O \subseteq \mathcal{Y}$ and any $y \in O$ we hence get

$$-\mathbb{I}'(y) \leq \liminf_{\epsilon \to 0} \epsilon \log \nu_\epsilon(G).$$

Taking the infimum over all $y \in O$ establishes the weak large deviation lower bound. $\square$

Definition 6.0.8. *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a probability space and $(\mathfrak{X}, ||\cdot||)$ a Banach space over $\mathbb{R}$. A function $\phi : \mathcal{X} \to \mathfrak{X}$ of the form*

$$\phi(x) = \sum_{i=1}^{n} x_i \mathbb{1}_{A_i}(x),$$

for $n \in \mathbb{N}$, $x \in \mathcal{X}$, $\{x_i\}_{i \in \{1,\dots,n\}} \subseteq \mathfrak{X}$ and pairwise disjoint $\{A_i\}_{i \in \{1,\dots,n\}} \subseteq \mathcal{F}$ is called a simple function. We define the Bochner-integral of ϕ w.r.t. μ

$$\int_{\mathcal{X}} \phi(x) d\mu(x) := \sum_{i=1}^{n} x_i \mu(A_i) \in \mathfrak{X}.$$

Definition 6.0.9. *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a probability space and $(\mathfrak{X}, ||\cdot||)$ a Banach space over $\mathbb{R}$. A function $f : \mathcal{X} \to \mathfrak{X}$ is called Bochner-integrable, if there exists a sequence $\{\phi_n\}_{n \in \mathbb{N}}$ of simple functions with*

$$f(x) = \lim_{n \to \infty} \phi_n(x)$$

almost everywhere, and if

$$\forall \epsilon > 0 : \exists n_0 \in \mathbb{N} : \forall n, m \geq n_0 : \int_{\mathcal{X}} ||\phi_n(x) - \phi_m(x)|| d\mu(x) < \epsilon.$$

The integral in the expression above is the usual Lebesgue integral.

Remark. There are equivalent characterisations of Bochner-integrability. In particular a function $f : \mathcal{X} \to \mathfrak{X}$ is Bochner-integrabel iff $||f|| : \mathcal{X} \to \mathbb{R}$ is Lebesgue-integrabel.

Theorem 6.0.17. *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a probability space and $(\mathfrak{X}, \|\cdot\|)$ a Banach space over $\mathbb{R}$. For a Bochner-integrable function $f : \mathcal{X} \to \mathfrak{X}$ with an approximating sequence of simple functions $\{\phi_n\}_{n\in\mathbb{N}}$,*

$$\lim_{n\to\infty} \int_{\mathcal{X}} \phi_n(x) d\mu(x) \in \mathcal{X}$$

exists and is independent of the choice of the sequence $\{\phi_n\}_{n\in\mathbb{N}}$. Hence, we define

$$\int_{\mathcal{X}} f(x) d\mu(x) := \lim_{n\to\infty} \int_{\mathcal{X}} \phi_n(x) d\mu(x).$$

Remark. It is possible to construct a more general version of the Bochner integral on certain measure spaces, for our purpose Theorem 6.0.17 suffices.

Theorem 6.0.18. *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a probability space and $(\mathfrak{X}, \|\cdot\|)$,$(\mathfrak{Z}, \|\cdot\|_Z)$ Banach spaces over $\mathbb{R}$. Moreover, let $T : \mathfrak{X} \to \mathfrak{Z}$ be a continuous linear operator and $f : \mathcal{X} \to \mathfrak{X}$ a Bochner-integrabel function. Then $T(f) : \mathcal{X} \to \mathfrak{Z}$ is also Bochner-integrabel and we have*

$$T\Big(\int_{\mathcal{X}} f(x) d\mu(x) \Big) = \int_{\mathcal{X}} T(f(x)) d\mu(x).$$

Definition 6.0.10. *Let $f : X \to \mathbb{R}$ be a differentiable convex function. Then f is called co-finite if for all $\{x_i\}_{i\in\mathbb{N}}$ with $\lim_{i\to\infty} \|x_i\|_2 = \infty$ we have*

$$\lim_{i\to\infty} \|\nabla f(x_i)\|_2 = \infty.$$

Theorem 6.0.19. *(Theorem 26.6. in [25])*
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable convex function. In order that $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ is bijective, it is necessary and sufficient that f be strictly convex and co-finite. When these conditions hold, f^* is likewise a differentiable convex function on $\mathbb{R}^n$ which is strictly convex and co-finite, and f^* is the same as the Legendre conjugate of f^*, i.e. for $x^* \in \mathbb{R}^n$

$$f^*(x^*) = \langle (\nabla f)^{-1}(x^*), x^* \rangle - f((\nabla f)^{-1}(x^*)),$$

where $\langle \cdot, \cdot \rangle$ denotes the standard scalar product in $\mathbb{R}^n$.

Remark. In [18] the authors make frequently use of Theorem 6.0.19 to prove Lemma 5.2.1. In general it is not true that $f : \mathbb{R}_- \times \mathbb{R} \to \mathbb{R}$ with

$$f(s,t) = \log \int_{\mathbb{R}} e^{sV(x)+tx^2} dx$$

is co-finite. Consider the case $V(x) = x^4$ and the sequence $\{(-i, 0)\}_{i\in\mathbb{N}} \subseteq \mathbb{R}_- \times \mathbb{R}$ with

$\|(-i,0)\|_2 \to \infty$ as $i \to \infty$. Then we have

$$\begin{aligned} \frac{d}{ds} f(s,t)\Big|_{(s,t)=(-i,0)} &= Z_{-i,0}^{-1} \int_{\mathbb{R}} x^4 e^{-ix^4} dx \\ \frac{d}{dt} f(s,t)\Big|_{(s,t)=(-i,0)} &= Z_{-i,0}^{-1} \int_{\mathbb{R}} x^2 e^{-ix^4} dx, \end{aligned} \tag{6.0.1}$$

where $Z_{-i,0} = \int_{\mathbb{R}} e^{-ix^4} dx$. We can calculate the quantities in 6.0.1. For the normalisation constant we have

$$Z_{-i,0} = \int_{\mathbb{R}} e^{-ix^4} dx = i^{-1/4} \frac{1}{2} \Gamma\left(\frac{1}{4}\right),$$

where Γ denotes the Gamma function. Similar one can calculate the partial derivatives of f, where we get that

$$\begin{aligned} \int_{\mathbb{R}} x^4 e^{sx^4} dx &= i^{-5/4} \frac{1}{2} \Gamma\left(\frac{5}{4}\right) \\ \int_{\mathbb{R}} x^2 e^{sx^4} dx &= i^{-3/4} \frac{1}{2} \Gamma\left(\frac{3}{4}\right). \end{aligned}$$

Thus, the gradient of f evaluated in $(-i,0)$ can be expressed in the following way

$$\nabla f(-i,0) = \left(\frac{1}{i} \frac{\Gamma(5/4)}{\Gamma(1/4)}, \frac{1}{\sqrt{i}} \frac{\Gamma(3/4)}{\Gamma(1/4)} \right).$$

In total we get that $\|\nabla f(-i,0)\|_2 \to 0$ even though $\|(-i,0)\|_2 \to \infty$ as $i \to \infty$.

Theorem 6.0.20. *(Householder transformation)*
Let $a, e \in \mathbb{R}^n$ be normed vectors, i.e. $\|a\|_2 = \|e\|_2 = 1$ with $a \neq e$. Then there exists a normed vector $v \in \mathbb{R}^n$ such that $H_v a = e$, where

$$H_v := I_n - 2vv^T$$

is an orthogonal matrix.

Remark. One can verify directly that -1 is an eigenvalue of H_v with multiplicity 1 and that 1 is the only other eigenvalue of H_v with multiplicity $n-1$. The determinant is -1 and hence, H_v belongs to the set of „reflection“ matrices.

Proof. (Proof of Theorem 6.0.20)
We define

$$v := \frac{a-e}{\|a-e\|_2}$$

and calculate $H_v a$, where we get (recall that a and e are normed)

$$\begin{aligned} H_v a &= a - 2vv^T a \\ &= a - 2(a-e)\frac{\|a\|_2^2 - \langle e, a\rangle}{\|a\|_2 - 2\langle a, e\rangle + \|e\|_2^2} \\ &= a - 2(a-e)\frac{1 - \langle e, a\rangle}{2 - 2\langle a, e\rangle} \\ &= e. \end{aligned}$$

□

Remark. Householder transforms play an important role in numerical mathematics, where one uses the previous construction to calculate the so-called QR decomposition.

Bibliography

[1] David Alonso-Gutiérrez, Joscha Prochno, and Christoph Thäle. "Large deviations for high-dimensional random projections of ℓ_p^n -balls". In: *Adv. in Appl. Math.* 99 (2018), pp. 1–35.

[2] David Alonso-Gutiérrez, Joscha Prochno, and Christoph Thäle. "Large deviations, moderate deviations, and the KLS conjecture". In: *J. Funct. Anal.* 280.1 (2021), pp. 108779, 33.

[3] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2005.

[4] Milla Anttila, Keith Ball, and Irini Perissinaki. "The central limit problem for convex bodies". In: *Trans. Amer. Math. Soc.* 355.12 (2003), pp. 4723–4735.

[5] Gérard B. Arous, Amir Dembo, and Alice Guionnet. "Aging of spherical spin glasses". In: *Probab. Theory Related Fields* 120.1 (2001), pp. 1–67.

[6] Ludwig Boltzmann. *Über die Beziehung zwischen dem zweiten Hauptsatze des mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung, respective den Sätzen über das Wärmegleichgewicht*. Kk Hof-und Staatsdruckerei, 1877.

[7] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Second. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2006.

[8] Harald Cramér and Hugo Touchette. "On a new limit theorem in probability theory (Translation of 'Sur un nouveau théoréme-limite de la théorie des probabilités')". In: *arXiv preprint arXiv:1802.05988* (2018).

[9] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*. Vol. 38. Stochastic Modelling and Applied Probability. Corrected reprint of the second (1998) edition. Springer-Verlag, Berlin, 2009.

[10] Paul Dupuis, Vaios Laschos, and Kavita Ramanan. “Large deviations for configurations generated by Gibbs distributions with energy functionals consisting of singular interaction and weakly confining potentials”. In: *Electron. J. Probab.* 25 (2020).

[11] Richard S. Ellis. *Entropy, large deviations, and statistical mechanics*. Classics in Mathematics. Reprint of the 1985 original. Springer-Verlag, Berlin, 2006.

[12] Richard S. Ellis. “The theory of large deviations: from Boltzmann’s 1877 calculation to equilibrium macrostates in 2D turbulence”. In: vol. 133. 1-4. Predictability: quantifying uncertainty in models of complex phenomena (Los Alamos, NM, 1998). 1999, pp. 106–136.

[13] Mark I. Freidlin and Alexander D. Wentzell. *Random perturbations of dynamical systems*. Third. Vol. 260. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Translated from the 1979 Russian original by Joseph Szücs. Springer, Heidelberg, 2012.

[14] Nina Gantert, Steven S. Kim, and Kavita Ramanan. “Cramér’s theorem is atypical”. In: *Advances in the mathematical sciences*. Vol. 6. Assoc. Women Math. Ser. Springer, [Cham], 2016, pp. 253–270.

[15] Nina Gantert, Steven S. Kim, and Kavita Ramanan. “Large deviations for random projections of ℓ^p balls”. In: *Ann. Probab.* 45.6B (2017), pp. 4419–4476.

[16] Zakhar Kabluchko, Joscha Prochno, and Christoph Thäle. “High-dimensional limit theorems for random vectors in ℓ_p^n-balls”. In: *Commun. Contemp. Math.* 21.1 (2019), pp. 1750092, 30.

[17] Zakhar Kabluchko, Joscha Prochno, and Christoph Thäle. “Sanov-type large deviations in Schatten classes”. In: *Ann. Inst. Henri Poincaré Probab. Stat.* 56.2 (2020), pp. 928–953.

[18] Steven S. Kim, Yin-Ting Liao, and Kavita Ramanan. “An asymptotic thin shell condition and large deviations for random multidimensional projections”. In: (2020). arXiv: `1912.13447 [math.PR]`.

[19] Steven S. Kim and Kavita Ramanan. “A conditional limit theorem for high-dimensional ℓ^p -spheres”. In: *J. Appl. Probab.* 55.4 (2018), pp. 1060–1077.

[20] Bo’az Klartag. “A central limit theorem for convex sets”. In: *Invent. Math.* 168.1 (2007), pp. 91–131.

[21] Alexander V. Kolesnikov and Emanuel Milman. "The KLS isoperimetric conjecture for generalized Orlicz balls". In: *Ann. Probab.* 46.6 (2018), pp. 3578–3615.

[22] Yin-Ting Liao and Kavita Ramanan. "Geometric sharp large deviations for random projections of ℓ_p^n spheres". In: *arXiv preprint arXiv:2001.04053* (2020).

[23] Vitali D. Milman and Gideon Schechtman. *Asymptotic theory of finite-dimensional normed spaces*. Vol. 1200. Lecture Notes in Mathematics. Springer-Verlag, Berlin, 1986.

[24] Marcin Pilipczuk and Jakub O. Wojtaszczyk. "The negative association property for the absolute values of random variables equidistributed on a generalized Orlicz ball". In: *Positivity* 12.3 (2008), pp. 421–474.

[25] Ralph T. Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[26] Srinivasa Varadhan. "Asymptotic probabilities and differential equations". In: *Comm. Pure Appl. Math.* 19 (1966), pp. 261–286.

[27] Srinivasa Varadhan. *Large deviations and applications*. Vol. 46. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1984.

[28] Cédric Villani. *Optimal transport*. Vol. 338. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Old and new. Springer-Verlag, Berlin, 2009.

[29] Royce KP Zia, Edward F Redish, and Susan R McKay. "Making sense of the Legendre transform". In: *American Journal of Physics* 77.7 (2009), pp. 614–622.