Sanov-type large deviations and conditional limit theorems for high-dimensional Orlicz balls

Lorenz Frühwirth and Joscha Prochno

Abstract

In this paper, we prove a Sanov-type large deviation principle for the sequence of empirical measures of vectors chosen uniformly at random from an Orlicz ball. From this level-2 large deviation result, in a combination with Gibbs conditioning, entropy maximization and an Orlicz version of the Poincaré-Maxwell-Borel lemma, we deduce a conditional limit theorem for high-dimensional Orlicz balls. Roughly speaking, the latter shows that if $V_1$ and $V_2$ are Orlicz functions, then random points in the $V_1$-Orlicz ball, conditioned on having a small $V_2$-Orlicz radius, look like an appropriately scaled $V_2$-Orlicz ball. In fact, we show that the limiting distribution in our Poincaré-Maxwell-Borel lemma, and thus the geometric interpretation, undergoes a phase transition depending on the magnitude of the $V_2$-Orlicz radius.

Keywords. Entropy maximization, Gibbs conditioning principle, Large deviation principle, Orlicz space, Poincaré-Maxwell-Borel lemma, Sanov’s theorem

MSC. Primary 60F05, 60F10, 52A20; Secondary 52A23, 60D05

1 Introduction & Main results

It is a classical and famous result, independently observed by Borel [3] and Maxwell (e.g., [7]), and commonly referred to as the Poincaré–Maxwell–Borel lemma, that any fixed number of coordinates of random vectors uniformly distributed on a high-dimensional Euclidean sphere is approximately Gaussian distributed. More precisely, if $X^{(n)}(\mathbf{1}), \ldots, X^{(n)}(n)$ is uniformly distributed on $\mathbb{S}^{n-1} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : \sum_{i=1}^n |x_i|^2 = n\}$, then, for any fixed $k \leq n$, the distribution of the first $k$ coordinates $(X^{(n)}(1), \ldots, X^{(n)}(k))$ converges weakly to the distribution of a $k$-dimensional standard Gaussian random vector. The normalization of the sphere is taken to ensure that a typical coordinate of an element of $\mathbb{S}^{n-1}$ has unit order when the dimension $n$ of the ambient space is large.

In 1991, Mogul’skii [22] and Rachev and Rüschendorf [25] obtained an $\ell_p$ analogue of the Poincaré–Maxwell–Borel lemma for the surface and cone measure respectively on a properly scaled $\ell_p$-sphere (note that surface and cone measure coincide whenever $p \in (1, 2, \infty)$). A simplification of the arguments can be found in the work of Naor and Romik [23, Theorems 3 and 4] and a significant generalization (including the case of Orlicz balls) has recently been obtained by Johnston and Prochno in [11]. In the $\ell_p$-versions of the Poincaré–Maxwell–Borel lemma, instead of the standard Gaussian distribution, the so-called $p$-generalized Gaussian distribution appears in the weak limit, the Lebesgue density being given by

$$
\mathbb{R} \ni x \mapsto \frac{1}{2p^{1/p} \Gamma(1 + 1/p)} e^{-|x|^{p}/p}, \quad 1 \leq p < \infty.
$$

This density is intimately related to $\ell_p^n$-balls, we write $\mathbb{B}_p^n$, and of fundamental importance in essentially any probabilistic approach to understand their geometry; see, e.g., [13, 28, 29] and references cited therein.

Recently, Kim and Ramanan [16, Theorem 2.3] showed that a conditional analog to the $\ell_p$-version of the Poincaré–Maxwell–Borel lemma holds. Roughly speaking, they found that for $1 \leq q < p < \infty$, as the dimension $n$ of the space is high, a random point on the $\ell_p^n$-sphere, conditioned on having...
a sufficiently small $\ell_q$-norm, is weakly close to a random point drawn from an appropriately scaled $\ell_n^p$-sphere. As the authors point out, this means that conditioning on a sufficiently small $\ell_q$-norm induces a probabilistic change that admits a nice geometric interpretation. The strategy of proof is based on ideas from statistical mechanics and large deviations theory (see also [12] for more in this direction). The authors first prove a level-2 large deviation principle for the empirical measure of the coordinates of a random point on a properly scaled $\ell_n^p$-sphere, where the random choice is made with respect to the cone probability measure on the boundary. Fundamental to this proof is a probabilistic representation of the cone measure in terms of the $p$-generalized Gaussian distribution [16, Lemma 3.9], which goes back to Schechtman and Zinn [29] and Rachev and Rüschendorf [25], and allows one to go over to random vectors with independent coordinates. Let us already point out that in our generalized setting considered here, no such probabilistic representation is available. Having proved a Sanov-type large deviation principle for the empirical measure, the authors in [16] move forward deducing the conditional limit theorems from this large deviation result and the famous Gibbs conditioning principle; obviously numerous (technical) hurdles need to be taken here. The use of the Gibbs conditioning principle allows one to transform a Sanov-type large deviation principle for an empirical measure into a statement about the most likely behavior of the underlying sequence of random vectors conditioned on a rare event.

In this work, we prove that an analog result holds for the general and fundamental class of Orlicz balls; Orlicz spaces are natural generalizations of $\ell_p$-spaces and intensively studied in functional analysis and optimization theory (see, e.g., [10, 14, 17, 18, 21, 24, 27, 30]). One of the challenges that arises is that no probabilistic representation, like in the case of $\ell_p$-balls, is at our disposal. This makes an analysis more challenging. In addition, a general Orlicz function is not homogeneous, making any approach to understanding analytic, geometric or probabilistic aspects of these spaces and their unit balls even more complicated. Key to our generalization, which follows the general theme of first proving a level-2 large deviation principle for the sequence of empirical measures of the coordinates of random points chosen uniformly from a suitably scaled Orlicz ball and then using Gibbs conditioning, is the statistical mechanics point of view. Indeed, here conditional distributions appear in connection to the study of non-interacting particle systems (micro-canonical and canonical ensembles), where one is seeking to describe the most likely state of the system under an energy constraint (see, e.g., [9], [26], and [12, Subsection 1.2]). This point of view directly links such questions to certain Gibbs distributions. Considering such Gibbs measures at suitable critical temperatures with potentials given by an Orlicz function allows us to work around the problem arising from the lack of a probabilistic representation. This idea has recently been put forward by Kabluchko and Prochno [12] (see particularly Subsection 1.2 there) in their probabilistic approach to the geometry of Orlicz balls and has also successfully been used in [1, 2], and [15].

1.1 Main results

In order to present the main results of this paper, we need to introduce some notation; for the rest we refer the reader to Section 2 below.

Recall that an Orlicz function $V : \mathbb{R} \to [0,\infty)$ is a convex, symmetric function with $V(0) = 0$ and $V(x) > 0$ for $x \neq 0$. For $R \in (0,\infty)$, we denote by

$$
\mathbb{B}_R^n := \{x \in \mathbb{R}^n : \sum_{k=1}^n V(x_k) \leq R n\}.
$$

the associated Orlicz ball in $\mathbb{R}^n$. If $V : \mathbb{R} \to [0,\infty)$ is an Orlicz function and $\beta \in (-\infty,0)$, we define the corresponding Gibbs measure $\mu_{V,\beta}$ (with potential $V$ and at critical temperature $\beta$) via the Lebesgue density

$$
\frac{d\mu_{V,\beta}}{dx}(x) := \frac{e^{\beta V(x)}}{\int_{\mathbb{R}} e^{\beta V(t)} dt}, \quad x \in \mathbb{R}.
$$
We shall denote by $\mathcal{M}_1(\mathbb{R})$ the set of probability measures on the Borel $\sigma$-algebra on $\mathbb{R}$ and, for $\mu \in \mathcal{M}_1(\mathbb{R})$, we define
\[ m_V(\mu) := \int_{\mathbb{R}} V(x) \mu(dx) \quad \text{and} \quad \mathcal{M}_1^V(\mathbb{R}) := \{ \mu \in \mathcal{M}_1(\mathbb{R}) : m_V(\mu) < \infty \} . \]

In case of $V(x) = |x|^p$ for some $p \in [1, \infty)$, we write $\mathcal{M}_1^p(\mathbb{R}) := \mathcal{M}_1^V(\mathbb{R})$. For a metric $d_w$ inducing weak convergence on $\mathcal{M}_1(\mathbb{R})$, we can define a metric $d_V$ on $\mathcal{M}_1^V(\mathbb{R})$ with
\[ d_V(\mu, v) := d_w(\mu, v) + |m_V(\mu) - m_V(v)| . \]

Our first main result establishes a level-2 or Sanov-type large deviation principle for the empirical measure of the coordinates of a vector chosen uniformly at random from an Orlicz ball.

**Theorem A.** Let $V$ be an Orlicz function and $R \in (0, \infty)$. Assume that $X_1^{n,V} := (X_R^{n,V}(1), \ldots, X_R^{n,V}(n)) \sim \text{Unif}(\mathbb{R}_R^{n,V})$. Then, there exists a unique $\overline{\pi} \in (-\infty, 0)$ such that the sequence of empirical measures
\[ \mathcal{L}^{n,V} := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_R^{n,V}(i)}, \quad n \in \mathbb{N}, \]

satisfies an LDP in $\mathcal{M}_1(\mathbb{R})$ with the strictly convex good rate function $\ell_V : \mathcal{M}_1(\mathbb{R}) \to [0, \infty)$, where
\[ \ell_V(\mu) = \begin{cases} H(\mu|\overline{\mu}, \overline{\pi}) + \overline{\pi}|m_V(\mu) - R| & : m_V(\mu) \leq R, \\ \infty & : \text{else}. \end{cases} \]

In addition, if there exists another Orlicz function $W$ such that
\[ \frac{V(x)}{W(x)} \to \infty, \quad \text{as } |x| \to \infty, \]
then, $(\mathcal{L}^{n,V})_{n \in \mathbb{N}}$ satisfies an LDP in $(\mathcal{M}_1^W(\mathbb{R}), d_W)$.

**Remark 1.1.** The proof of Theorem A rests on a general version of the Gärtner–Ellis theorem. Beyond the Sanov-type large deviation principle on $\mathcal{M}_1(\mathbb{R})$ endowed with the weak topology, also the last part of the statement, the large deviation principle on the generalized Wasserstein space $(\mathcal{M}_1^V(\mathbb{R}), d_W)$, is essential for the proof of our second main result, Theorem B below.

Let $(X_R^{n,V})_{n \in \mathbb{N}}$ be a sequence of random variables, where each $X_R^{n,V}$ is uniformly distributed on the Orlicz ball $\mathbb{R}_R^{n,V}$. For fixed $k \in \mathbb{N}$, we denote by $\lambda^{(n,k)}$ the distribution of the first $k$ coordinates of $X_R^{n,V}$, i.e.,
\[ \lambda^{(n,k)}(\cdot) := \mathbb{P}\left[ (X_R^{n,V}(1), \ldots, X_R^{n,V}(k)) \in \cdot \right]. \]

Let $V_1, V_2$ be two Orlicz functions and $X_1^{n,V_1} \sim \text{Unif}(\mathbb{R}_1^{n,V_1})$. Then, for $R \in (0, \infty)$, we define
\[ \hat{\lambda}^{(R,k)}_{n,V_1|V_2} := \mathbb{P}\left[ (X_1^{n,V_1}(1), \ldots, X_1^{n,V_1}(k)) \in \cdot | X_1^{n,V_1} \in \mathbb{R}_1^{n,V_1} \right]. \]

We can now state the second main result of this paper, which establishes a conditional limit theorem for random points in Orlicz balls.

**Theorem B.** Let $V_1, V_2$ be Orlicz functions that satisfy
\[ \frac{V_1(x)}{V_2(x)} \to \infty, \quad \text{as } |x| \to \infty \quad \text{and} \quad \int_{\mathbb{R}} V_1(x) e^{\alpha V_2(x)} dx < \infty, \]
for all $\alpha \in (-\infty, 0)$. Then, there exists an $\overline{R} \in (0, \infty)$ such that for all $0 < R \leq \overline{R}$ and for fixed $k \in \mathbb{N}$, we have
\[ \lim_{\epsilon \to 0} \lim_{n \to \infty} d_w(\hat{\lambda}^{(R+\epsilon,k)}_{n,V_1|V_2}, \lambda^{(n,k)}) = 0, \]
where $d_w$ denotes some metric inducing weak convergence of probability measures.
We shall deduce Theorem B from our Sanov-type large deviation result in Theorem A, which we combine with the so-called Gibbs conditioning principle, entropy maximization, and an Orlicz version of the Poincaré-Maxwell-Borel lemma. Let us continue with a few remarks related to Theorem B.

Remark 1.2. Under the assumptions of Theorem B, roughly speaking, in high dimensions we have that random points in the \( V_1 \)-Orlicz ball, conditioned on having a small \( V_2 \)-Orlicz radius, look like an appropriately scaled \( V_2 \)-Orlicz ball.

Remark 1.3. Theorem B generalizes [16, Theorem 2.3]. There the authors proved a conditional limit theorem in an \( \ell_p \)-setting. Let us point out that there spheres rather than balls were considered.

Remark 1.4. In Corollary 4.9 we will see that, if the radius \( R \) in the condition of \( \hat{\lambda}^{(R+\epsilon,k)}_{n,V_1|V_2} \) is chosen sufficiently large, then the limiting distribution of \( \hat{\lambda}^{(R+\epsilon,k)}_{n,V_1|V_2} \) is the same as the one of \( \lambda^{(n,k)}_{1,V_1} \). Geometrically this means that, in high dimensions, points in a \( V_1 \)-Orlicz ball conditioned on being in a big \( V_2 \)-Orlicz ball are uniformly distributed in the \( V_1 \)-Orlicz ball, which is to be expected. We refer to Remark 4.11 on the phase transition in the “intermediate case”.

Remark 1.5. The second condition in Equation (4) ensures the existence of \( m_{V_2}(\mu_{V_1,\alpha}) \) for all \( \alpha \in (-\infty,0) \). This condition does not appear in case of \( \ell_p \)-balls, since
\[
\int_{\mathbb{R}} |x|^p e^{-|x|^q} \, dx < \infty
\]
for all \( 1 \leq p, q < \infty \). In particular, the second Condition in (4) is satisfied for all polynomially bounded \( V_1 \). A similar condition appears in [12, Theorem B].

Remark 1.6. Fix \( p \in [1,\infty) \) and take some Orlicz function \( V \) with
\[
\frac{V(x)}{|x|^p} \to \infty \quad \text{as} \quad |x| \to \infty.
\]
Then, the sequence of empirical measures \( (L_n^{V})_{n \in \mathbb{N}} \) from Equation (1) satisfies an LDP in the \( p \)-Wasserstein space \( M_{1}^p(\mathbb{R}) \) with good rate function \( I_V \) defined in Equation (2) (see also Corollary 3.5).

2 Notation and Preliminaries

2.1 Notation

We shall use the following standard notation. Given a Borel measurable set \( A \subseteq \mathbb{R}^n \), we shall denote by vol\( _n \) the \( n \)-dimensional Lebesgue measure. If \( X \) is a topological space, then for Borel sets \( A \subseteq X \), we write \( A^o \) and \( \overline{A} \) for the interior and the closure of \( A \) respectively. For a sequence \( (\nu_n)_{n \in \mathbb{N}} \subseteq \mathcal{M}_1(\mathbb{R}) \), which converges weakly to a distribution \( \nu \in \mathcal{M}_1(\mathbb{R}) \) as \( n \to \infty \), we write \( \nu_n \xrightarrow{\text{w}} \nu \) as \( n \to \infty \). For a measure \( \nu \in \mathcal{M}_1(\mathbb{R}) \), we define the relative entropy \( H(\cdot|\nu) : \mathcal{M}_1(\mathbb{R}) \to \mathbb{R} \) with
\[
H(\mu|\nu) := \left\{ \begin{array}{ll}
\int_{\mathbb{R}} \log \frac{d\mu}{d\nu}(x) \, d\nu(x) & : \frac{d\mu}{d\nu} \text{ exists} \\
\infty & : \text{else}
\end{array} \right.
\] (6)
A related quantity is the entropy \( h : \mathcal{M}_1(\mathbb{R}) \to \mathbb{R} \) of a measure, which is defined as
\[
h(\nu) := \left\{ \begin{array}{ll}
-\int_{\mathbb{R}} \log \frac{d\nu}{dx}(x) \, dx & : \frac{d\nu}{dx} \text{ exists} \\
\infty & : \text{else}
\end{array} \right.
\]
For $\mu \in \mathcal{M}_1(R)$ and a measurable function $f : R \to R$, we set
$$\mathbb{E}_\mu [f] := \int_R f(x) \mu(dx),$$
provided the latter exists. Also, for $\mu \in \mathcal{M}_1(R)$, we define $\delta_\mu : \mathcal{M}_1(R) \to R$ with
$$\delta_\mu (\nu) = \begin{cases} 1 : \nu = \mu \\
0 \quad \text{else.} \end{cases}$$
provided $\mu \notin \mathcal{M}_1(R)$. We will frequently work with the space of bounded and continuous functions mapping from $R$ to $R$, which we shall denote by $C_b(R)$.

### 2.2 Orlicz functions & Gibbs measures

Let us recall here what an Orlicz function is and introduce Gibbs measures whose potentials are given by Orlicz functions.

**Definition 2.1.** An Orlicz function $V : R \to [0, \infty]$ is a convex, symmetric function with $V(0) = 0$ and $V(x) > 0$ for $x \neq 0$. We define the associated Orlicz ball
$$B^n_R, V := \{ x \in R^n : \sum_{k=1}^n V(x_k) \leq R \}$$
for some $R > 0$ and in case of $R = 1$, we write $B^n, V := B^n, V_1$.

For some Orlicz function $V : R \to [0, \infty)$ and $\beta \in (-\infty, 0)$, we define the Gibbs measure $\mu_{V, \beta}$ via the Lebesgue density
$$\frac{d\mu_{V, \beta}}{dx}(x) := \frac{e^{\beta V(x)}}{\int_R e^{\beta V(t)} dt}, \quad x \in R.$$

### 2.3 Basics from large deviation theory and probability

Consider a sequence of random variables $(\xi_n)_{n \in \mathbb{N}}$ taking values in a topological space $X$, equipped with a Borel sigma algebra and a probability measure $P$ on it. We say that $(\xi_n)_{n \in \mathbb{N}}$ satisfies an LDP in $X$, if there exists a good rate function (GRF) $I : X \to [0, \infty]$, i.e. $I$ has compact level sets, such that
$$- \inf_{x \in A} I(x) \leq \liminf_{n \to \infty} \frac{1}{n} \log P[\xi_n \in A] \leq \limsup_{n \to \infty} \frac{1}{n} \log P[\xi_n \in \overline{A}] \leq - \inf_{x \in \overline{A}} I(x) \tag{7}$$
for all Borel sets $A \subseteq X$.

There are different ways to show that a sequence of random variables satisfies an LDP, the most commonly used is the so called contraction principle (see, e.g., Theorem 4.2.1 in [6]).

**Lemma 2.2.** Let $X$ and $Y$ be Hausdorff topological spaces and $f : X \to Y$ be a continuous function. Let $(\xi_n)_{n \in \mathbb{N}}$ be a sequence of random variables that satisfies an LDP in $X$ with GRF $I : X \to [0, \infty]$. Then, the sequence $(f(\xi_n))_{n \in \mathbb{N}}$ satisfies an LDP in $Y$ with the GRF $I' : Y \to [0, \infty)$, where
$$I'(y) := \inf_{x \in f^{-1}(y)} I(x).$$

When working with LDPs it may occur that one encounters different topologies on the same ambient space $X$. Corollary [2,4] is a useful tool when dealing with that situation.
Definition 2.3. Let \((\xi_n)_{n \in \mathbb{N}}\) be a sequence of random variables taking values in a topological space \(X\). The sequence \((\xi_n)_{n \in \mathbb{N}}\) is called exponentially tight if for all \(\alpha \in (-\infty, 0)\) there exists a compact set \(K_\alpha \subseteq X\) such that
\[
\limsup_{n \to \infty} \frac{1}{n} \log P[\xi_n \in K_\alpha^c] < \alpha.
\]

Corollary 2.4. (Corollary 4.2.6 in [6]) Let \((\xi_n)_{n \in \mathbb{N}}\) be an exponentially tight sequence of random variables on a topological space \(X\) equipped with a topology \(\tau_1\). If \((\xi_n)_{n \in \mathbb{N}}\) satisfies an LDP with respect to a Hausdorff topology \(\tau_2\) that is coarser than \(\tau_1\), then the same LDP holds with respect to the finer topology \(\tau_1\).

Lastly, we present a version of the Gärtner–Ellis Theorem, which will play an important role in the proof of Theorem A from Section 2.

Lemma 2.5. (Corollary 4.16.14 in [6]) Let \((\xi_n)_{n \in \mathbb{N}}\) be a sequence of exponentially tight random variables on the locally convex Hausdorff topological vector space \(E\). Suppose the Gärtner–Ellis limit \(\Lambda(\lambda) := \lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}[e^{n\lambda(\xi_n)}], \lambda \in E^*\), exists in \(\mathbb{R}\) and is Gateaux-differentiable, where \(E^*\) is the topological dual of \(E\). Then, \((\xi_n)_{n \in \mathbb{N}}\) satisfies the LDP in \(E\) with the convex GRF \(\Lambda^* : E \to [0, \infty]\), given by
\[
\Lambda^*(x) := \sup_{\lambda \in E^*} \{\lambda(x) - \Lambda(\lambda)\}.
\]

For a measure \(\nu \in \mathcal{M}_1(\mathbb{R})\) we recall the relative entropy \(H(\cdot|\nu) : \mathcal{M}_1(\mathbb{R}) \to \mathbb{R}\) from Equation (6) in Section 2.1, for which the following variational formula holds.

Proposition 2.6. (Theorem 5.6 in [26]) Let \(\mu, \nu \in \mathcal{M}_1(\mathbb{R})\). Then
\[
H(\mu|\nu) = \sup_{\lambda \in C_b(\mathbb{R})} \{\mathbb{E}_\mu[\lambda] - \log \mathbb{E}_\nu[e^{\lambda}]\}.
\]

When working with the space of probability measures \(\mathcal{M}_1(\mathbb{R})\), the notion of tightness plays a major role.

Definition 2.7. A subset \(L\) of \(\mathcal{M}_1(\mathbb{R})\) is called tight if, for all \(\varepsilon > 0\), there exists a compact set \(K_\varepsilon \subseteq \mathbb{R}\) such that
\[
\sup_{\mu \in L} \mu(K_\varepsilon) \geq 1 - \varepsilon.
\]

The following result is known as Gibbs conditioning principle, one of the fundamental tools in the theory of large deviations. It will be used in the proof of Proposition 4.7, which is itself used in the proof of Theorem B.

Proposition 2.8. (Gibbs conditioning principle, Theorem 7.1 in [20]) Let \(X\) be a topological space, and let \((\zeta_n)_{n \in \mathbb{N}}\) be a sequence of \(X\)-valued random variables that satisfies an LDP in \(X\) with GRF \(I\). In addition, let \(F \subseteq X\) be a subset that satisfies
\begin{enumerate}
  \item \(I(F) := \inf_{x \in F} I(x) < \infty\);
  \item \(F\) is closed;
  \item \(F = \bigcap_{\varepsilon > 0} F_\varepsilon\) for a family of sets \((F_\varepsilon)_{\varepsilon > 0}\) such that \(F_\varepsilon \subseteq X\) for \(\varepsilon > 0\) and \(P[\zeta_n \in F_\varepsilon] > 0\) for all \(\varepsilon > 0\) and \(n \in \mathbb{N}\);
  \item \(F \subseteq (F_\varepsilon)^0\) for all \(\varepsilon > 0\).
\end{enumerate}
Let $\mathcal{M}_F$ be the set of $x \in F$ which minimize $0$. That is,

$$
\mathcal{M}_F := \{ x \in F : \Pi(x) = 0(F) \}.
$$

Then, for all open $G \subseteq X$ such that $\mathcal{M}_F \subseteq G$, we have

$$
\limsup_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log P[\xi_n \notin G \mid \xi_n \in F_\epsilon] < 0.
$$

As a consequence, if $\mathcal{M}_F = \{ \overline{x} \}$ is a singleton, then

$$
\lim_{\epsilon \to 0} \lim_{n \to \infty} P[\xi_n \notin I_{\overline{x}}] = \delta_{\overline{x}}.
$$

In the proof of Proposition 2.7, we shall make use of Proposition 2.11, which relies on [31, Proposition 2.1]. We start with a definition, where we adapt the notation from [31] in order to stay consistent with the notation used in this article.

**Definition 2.9.** Let $(\mu^n_{\epsilon})_{n \in \mathbb{N}, \epsilon > 0}$ be a family of probability measures, where for each $\epsilon > 0$ and each $n \in \mathbb{N}$, $\mu^n_{\epsilon}$ is a symmetric distribution on $\mathbb{R}^n$. We say that $(\mu^n_{\epsilon})_{n \in \mathbb{N}, \epsilon > 0}$ is $\mu$-chaotic if there exists $\mu \in \mathcal{M}_1(\mathbb{R})$ such that, for fixed $k \in \mathbb{N}$ and $\phi_1, ..., \phi_k \in C_b(\mathbb{R})$, we have that

$$
\lim_{\epsilon \to 0} \lim_{n \to \infty} \int_{\mathbb{R}^k} k \prod_{i=1}^k \phi_i(x_i) \mu^n_{\epsilon,k}(d x) = \prod_{i=1}^k \int_{\mathbb{R}} \phi_i(x) \mu(d x),
$$

where $\mu^n_{\epsilon,k}$ denotes the marginal distribution of the $k$ first coordinates of $\mu^n_{\epsilon}$.

**Remark 2.10.** In particular, the limit in Equation (6) implies weak convergence of the family $(\mu^n_{\epsilon})_{n \in \mathbb{N}, \epsilon > 0}$ towards $\mu^\oplus k$, where the latter denotes the $k$-fold product measure of $\mu$.

We have the following characterization of the notion defined in Definition 2.9.

**Proposition 2.11.** Let $(\mu^n_{\epsilon,k})_{n \in \mathbb{N}, \epsilon > 0}$ be a family of symmetric probability measures, where for each $\epsilon > 0$ and each $n \in \mathbb{N}$, $\mu^n_{\epsilon}$ is a symmetric distribution on $\mathbb{R}^n$. For all $\epsilon > 0$ and $n \in \mathbb{N}$, let $X^n_\epsilon = (X^n_\epsilon(1), ..., X^n_\epsilon(n)) \in \mathbb{R}^n$ be distributed according to $\mu^n_{\epsilon}$, then, the family $(\mu^n_{\epsilon,k})_{n \in \mathbb{N}, \epsilon > 0}$ is $\mu$-chaotic if and only if there exists $\mu \in \mathcal{M}_1(\mathbb{R})$ such that

$$
\frac{1}{n} \sum_{i=1}^n \delta_{X^n_\epsilon(i)} \Rightarrow \mu,
$$

as $n \to \infty$ followed by $\epsilon \to 0$.

**Proof.** Similar to the proof of [31, Proposition 2.2]. \qed

### 2.4 The generalized Wasserstein spaces

Let $\mathcal{M}_1(\mathbb{R})$ be the set of probability measures on the Borel $\sigma$-algebra on $\mathbb{R}$ and $V$ be an Orlicz function and recall the following quantities from Section 2.

**Definition 2.12.** For $\mu \in \mathcal{M}_1(\mathbb{R})$, we define

$$
m_V(\mu) := \int_{\mathbb{R}} V(x) \mu(dx) \quad \text{and} \quad \mathcal{M}_1^V(\mathbb{R}) := \left\{ \mu \in \mathcal{M}_1(\mathbb{R}) : m_V(\mu) < \infty \right\}.
$$

In case of $V(x) = |x|^p$ for some $p \in [1, \infty)$, we write $m_p := m_V$ and $\mathcal{M}_1^p(\mathbb{R}) := \mathcal{M}_1^V(\mathbb{R})$. For a metric $d_\mu$ inducing weak convergence on $\mathcal{M}_1(\mathbb{R})$, we can define a metric $d_V$ on $\mathcal{M}_1^V(\mathbb{R})$ with

$$
d_V(\mu, \nu) := d_\mu(\mu, \nu) + |m_V(\mu) - m_V(\nu)|.
$$
Proposition 2.17. Let $V$.

Remark 2.13. By the very definition of the metric $d_V$ on $\mathcal{M}_1^V$, the mapping $m_V : \mathcal{M}_1^V(\mathbb{R}) \rightarrow \mathbb{R}$ is a continuous one.

Remark 2.14. In case of $V(x) = |x|^p$, for some $p \in [1, \infty)$, the space $(\mathcal{M}_1^V(\mathbb{R}), d_V)$ becomes the $p$-Wasserstein space, which is complete and separable (see, e.g., Chapter 6 in [32]).

Note that for general Orlicz functions $V$, it is not claimed that $(\mathcal{M}_1^V(\mathbb{R}), d_V)$ is Polish, but we have the following result which suffices for our purposes.

Proposition 2.15. (Lemma C.1 in [8]) Let $V$ be an Orlicz function and $\mathcal{M}_1^V(\mathbb{R})$ and $d_V$ be defined as in (11) and (12), respectively. Then, for a sequence $(\mu_n)_{n \in \mathbb{N}} \subseteq \mathcal{M}_1^V(\mathbb{R})$ and some $\mu \in \mathcal{M}_1^V(\mathbb{R})$, we have $d_V(\mu_n, \mu) \rightarrow 0$ as $n \rightarrow \infty$ if and only if

$$d_{\mu_n}(\mu_n, \mu) \rightarrow 0 \quad \text{and} \quad \limsup_{r \rightarrow \infty} \int_{|x| > r} V(x)\mu_n(dx) \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.$$ (13)

Furthermore, the metric space $(\mathcal{M}_1^V(\mathbb{R}), d_V)$ is separable.

Proposition 2.16. Let $M \in (0, \infty)$, then the set $K^M_{V_1} := \{v \in \mathcal{M}_1(\mathbb{R}) : \int_{\mathbb{R}} |x|dv(x) \leq M\}$ is compact in the weak topology on $\mathcal{M}_1(\mathbb{R})$.

Proof. The proof is analogue to the proof of Lemma 3.14 in [16].

Proposition 2.17. Let $V_1, V_2$ be two Orlicz functions with

$$\frac{V_1(x)}{V_2(x)} \rightarrow \infty \quad \text{as} \quad |x| \rightarrow \infty.$$ Then, for $M \in (0, \infty)$, the set $K^M_{V_1} := \{v \in \mathcal{M}_1(\mathbb{R}) : m_{V_1}(v) \leq M\}$ is compact in the $d_{V_2}$-topology on $\mathcal{M}_1(\mathbb{R})$.

Proof. Since the function $V_1 : \mathbb{R} \rightarrow (0, \infty)$ is convex, it has compact level sets. Thus, by Lemma 3.2 in [8], $K^M_{V_1}$ is tight. By Prohorov’s Theorem (see, e.g., Theorem D.9 in [6]), $K^M_{V_1}$ is precompact in the weak topology. To see that $K^M_{V_1}$ is also closed in the weak topology, we take a sequence $(\mu_n)_{n \in \mathbb{N}} \subseteq K^M_{V_1}$ with $\mu_n \rightarrow \mu \in \mathcal{M}_1(\mathbb{R})$ weakly. Then, for fixed $C \in (0, \infty)$, we get

$$M \geq \int_{\mathbb{R}} V_1(x)\mu_n(dx)$$
$$\geq \int_{\mathbb{R}} \max(C, V_1(x))\mu_n(dx) \rightarrow \int_{\mathbb{R}} \max(C, V_1(x))\mu(dx) \quad \text{as} \quad n \rightarrow \infty,$$

where the convergence in the second line holds since $\max(C, V_1)$ is continuous and bounded. By monotone convergence, we get

$$M \geq \lim_{C \rightarrow \infty} \int_{\mathbb{R}} \max(C, V_1(x))\mu(dx) = \int_{\mathbb{R}} V_1(x)\mu(dx).$$

Therefore, $\mu \in K^M_{V_1}$, which shows weak compactness of $K^M_{V_1}$. Now take a sequence $(\mu_n)_{n \in \mathbb{N}} \subseteq K^M_{V_1}$. Then, by weak compactness, there exists a subsequence $(\mu_{n_k})_{k \in \mathbb{N}}$ and $\mu \in \mathcal{M}_1(\mathbb{R})$ such that $\mu_{n_k} \rightarrow \mu$ weakly. We claim that $d_{V_1}(\mu_{n_k}, \mu) \rightarrow 0$ as $k \rightarrow \infty$ (see Equation (12) for the definition of this metric). In order to show that, we work with the characterization of convergence in $\mathcal{M}_1^V(\mathbb{R})$ given in Proposition 2.15. We get

$$\limsup_{r \rightarrow \infty} \left\{ \int_{|x|>r} V_2(x)\mu_{n_k}(dx) \right\} = \limsup_{r \rightarrow \infty} \left\{ \int_{|x|>r} \frac{V_2(x)}{V_1(x)} V_1(x)\mu_{n_k}(dx) \right\}$$
Thus, we get
\[
\phi \leq \limsup_{r \to \infty} \sup_{|x| \geq r} \left\{ \int_{|x| > r} V_1(x) \mu_{n_k}(d x) \right\} \leq M \limsup_{x \to -\infty} \frac{V_2(x)}{V_1(x)} = 0,
\]
where we used that \((\mu_{n_k})_{k \in \mathbb{N}} \subseteq K^M_{V_1}\) and \(\lim_{|x| \to -\infty} \frac{V_2(x)}{V_1(x)} = 0\). Thus, we have shown that every sequence in \(K^M_{V_1}\) has a \(d_{V_2}\) convergent subsequence, which implies that \(K^M_{V_1}\) is compact in \(\mathcal{M}^1_{V_2}(\mathbb{R})\).

\[\square\]

### 3 Proof of Theorem A

We shall now continue and present the proof of the Sanov-type large deviation principle presented in Theorem A.

**Lemma 3.1.** Let \(V \) be an Orlicz function and let \(R \in (0, \infty)\). Define \(\varphi_V : (-\infty, 0) \times C_b(\mathbb{R}) \to \mathbb{R}\) as
\[
\varphi_V(\alpha, \lambda) := \log \int_{\mathbb{R}} e^{\alpha V(x) + \lambda(x)} dx.
\]
Then, for each \(\lambda \in C_b(\mathbb{R})\), there exists an unique \(\alpha = \alpha(\lambda) \in (-\infty, 0)\) such that
\[
\frac{d}{d\alpha} \varphi_V(\alpha, \lambda) = R \quad \text{and} \quad \frac{d^2}{d\alpha^2} \varphi_V(\alpha, \lambda) \in (0, \infty).
\]
Moreover, the mapping \(\lambda \mapsto \varphi_V(\alpha(\lambda), \lambda)\) is Fréchet-differentiable and we have that
\[
\varphi_V(\alpha(\lambda), \lambda) - \alpha(\lambda)R = \inf_{\alpha \leq 0} \left\{ \varphi_V(\alpha, \lambda) - \alpha R \right\}.
\]

**Remark 3.2.** To be precise, the function \(\alpha(\lambda)\) also depends on \(R\) and on \(V\), but to keep the notational burden low, we omit these variables in this section.

**Proof.** (of Lemma 3.1) Fixing \(\lambda \in C_b(\mathbb{R})\), one can check easily that
\[
e^{-2||\alpha||_{\infty}} \frac{d}{d\alpha} \varphi_V(\alpha, 0) \leq \frac{d}{d\alpha} \varphi_V(\alpha, \lambda) \leq e^{2||\alpha||_{\infty}} \frac{d}{d\alpha} \varphi_V(\alpha, 0).
\]
We have that \(\frac{d}{d\alpha} \varphi_V(\alpha, 0) \to 0\) as \(\alpha \to -\infty\) as well as \(\frac{d}{d\alpha} \varphi_V(\alpha, 0) \to \infty\) as \(\alpha \to 0\). We show this. First, we observe that
\[
\alpha \to \frac{d}{d\alpha} \varphi_V(\alpha, 0) = \int_{\mathbb{R}} V(x) e^{\alpha V(x) - \varphi_V(\alpha, 0)} dx
\]
is, as a parameter integral, continuous on the negative real numbers. Now, for \(M \in (0, \infty)\),
\[
\int_{\mathbb{R}} V(x) e^{\alpha V(x) - \varphi_V(\alpha, 0)} dx \geq e^{-\varphi_V(\alpha, 0)} \int_{-M}^{\infty} V(x) e^{\alpha V(x)} dx \geq 2V(M) e^{-\varphi_V(\alpha, 0)} \int_{-M}^{\infty} e^{\alpha V(x)} dx \geq V(M) \left( 1 - 2e^{-\varphi_V(\alpha, 0)} \int_{-M}^{0} e^{\alpha V(x)} dx \right) \xrightarrow{\alpha \to 0} V(M).
\]
The limit in the last line holds, since \(\varphi_V(\alpha, 0) \to \infty\) as \(\alpha \to 0\) and \(\int_{-M}^{0} e^{\alpha V(x)} dx \leq M\) for all \(\alpha \in (-\infty, 0)\). Thus, we get
\[
\liminf_{\alpha \to 0} \int_{\mathbb{R}} V(x) e^{\alpha V(x) - \varphi_V(\alpha, 0)} dx \geq V(M) \quad \text{as} \quad M \to \infty.
\]
Next, we prove the other limit. In contrast, for \( \alpha \to -\infty \), we fix some \( x \in (0, \infty) \), where we have

\[
e^\alpha V(x) - \varphi_V(\alpha, 0) \leq \frac{e^\alpha V(x)}{\int_{-x/2}^{x/2} e^\alpha V(t) \, dt} \leq \frac{e^\alpha V(x)}{\left( \int_{-x/2}^{x/2} e^\alpha V(t) \, dt \right)^{1/2}} \leq \frac{1}{x} e^\alpha (V(x) - V(x/2)) \xrightarrow{\alpha \to -\infty} 0.
\]

In the last line, we used that \( V(x) - V(x/2) > 0 \), which holds since \( V \) is a convex function. Hence,

\[
\int_R V(x) e^{\alpha V(x) - \varphi_V(\alpha, 0)} \, dx \to 0 \quad \text{as} \quad \alpha \to -\infty.
\]

Thus, by Equation (16), we have that both

\[
\frac{d}{d\alpha} \varphi_V(\alpha, \lambda) \xrightarrow{\alpha \to -\infty} 0 \quad \text{and} \quad \frac{d}{d\alpha} \varphi_V(\alpha, \lambda) \xrightarrow{\alpha \to 0} \infty.
\]

This ensures that we find \( \overline{\alpha} \in (-\infty, 0) \) with \( \frac{d}{d\alpha} \varphi_V(\overline{\alpha}, \lambda) = 0 \). Now consider a random variable \( Z \) distributed according to the Gibbs density \( e^{\alpha V(x) + \lambda(x) - \varphi_V(\alpha, \lambda)} \). Then, we have that

\[
E[V(Z)] = \int_R V(x) e^{\overline{\alpha} V(x) + \lambda(x) - \varphi_V(\overline{\alpha}, \lambda)} \, dx = \frac{d}{d\alpha} \varphi_V(\overline{\alpha}, \lambda) = R
\]

and

\[
\mathcal{V}[V(Z)] = \int_R (V(x))^2 e^{\overline{\alpha} V(x) + \lambda(x) - \varphi_V(\overline{\alpha}, \lambda)} \, dx - \left( \int_R V(x) e^{\overline{\alpha} V(x) + \lambda(x) - \varphi_V(\overline{\alpha}, \lambda)} \, dx \right)^2
\]

\[
= \frac{d^2}{d\alpha^2} \varphi_V(\overline{\alpha}, \lambda).
\]

Hence, \( \frac{d^2}{d\alpha^2} \varphi_V(\overline{\alpha}, \lambda) \in (0, \infty) \). We now show that \( \overline{\alpha} \) is unique. The implicit function theorem (see, e.g., Theorem 5.9 in [19]) applied to the mapping \( \alpha \to \frac{d}{d\alpha} \varphi_V(\alpha, \lambda) - R \) guarantees that \( \overline{\alpha} = \alpha(\lambda) \) is unique and Fréchet-differentiable. Thus, as a composition of Fréchet-differentiable mappings, \( \varphi_V(\alpha(\lambda), \lambda) \) is Fréchet-differentiable. To see that \( \alpha(\lambda) \) is the minimizer in (15), we note that by the Cauchy-Schwarz inequality, \( \alpha \to \varphi_V(\alpha(\lambda), \lambda) \) is strictly convex and so is \( \alpha \to \varphi_V(\alpha(\lambda), \lambda) - \alpha R \). Since \( \varphi_V(\alpha(\lambda), \lambda) \to \infty \) as \( \alpha \to 0 \), we find an unique minimizer \( \overline{\alpha}(\lambda) \in (-\infty, 0) \) with

\[
\frac{d}{d\alpha} \varphi_V(\overline{\alpha}(\lambda), \lambda) - R = 0
\]

and hence \( \overline{\alpha}(\lambda) = \alpha(\lambda) \).

In the proof Theorem A, the Legendre transform of a certain functional containing \( \varphi_V(\alpha, \lambda) \) from Lemma 3.1 will appear, which we shall compute in the following proposition.

**Proposition 3.3.** Let \( V : \mathbb{R} \to [0, \infty) \) be an Orlicz function and let \( R \in (0, \infty) \). Define \( \Lambda_V : C_b(\mathbb{R}) \to \mathbb{R} \) as

\[
\Lambda_V(\lambda) := \varphi_V(\alpha(\lambda), \lambda) - \alpha(\lambda) R - \varphi_V(\alpha(0), 0) + \alpha(0) R,
\]

where \( \varphi_V(\alpha, \lambda) \) and \( \alpha(\lambda) \) are defined as in Lemma 3.1. Then, the Legendre transform \( \mathcal{L}_V : \mathcal{M}_1(\mathbb{R}) \to [0, \infty) \) of \( \Lambda_V \) is given as

\[
\mathcal{L}_V(\mu) := \begin{cases} H(\mu|\mu_V, a(0)) + \alpha(0) \left[ m_V(\mu) - R \right] : m_V(\mu) \leq R \\ \infty : \text{else}, \end{cases}
\]

where for \( \beta \in (-\infty, 0), \mu_{V, \beta} \) is the distribution with Lebesgue density

\[
\frac{d\mu_{V, \beta}(x)}{dx} = e^{\beta V(x) - \varphi_V(\beta, 0)}, \quad x \in \mathbb{R}.
\]

In addition, \( \mathcal{L}_V \) is strictly convex where finite.
Remark 3.4. If $\mu \in \mathcal{M}_1(\mathbb{R})$ with $m_V(\mu) \leq R$ and $\mu$ is absolutely continuous with respect to Lebesgue measure, then we have

$$l_V(\mu) = -h(\mu) + \varphi_V(\alpha(0), 0) - \alpha(0)R,$$

where $h(\mu)$ is the entropy of $\mu$. Indeed, for $\mu \in \mathcal{M}^V_1(\mathbb{R})$, we have

$$H(\mu|\mu_{V,a(0)}) = \int_{\mathbb{R}} \log \left( \frac{d\mu}{d\mu_{V,a(0)}}(x) \right) \mu(dx)$$

$$= \int_{\mathbb{R}} \log \left( \frac{d\mu}{d\mu}(x) \right) \mu(dx) + \int_{\mathbb{R}} \log \left( \frac{d\mu_{V,a(0)}}{d\mu}(x) \right) \mu(dx)$$

$$= -h(\mu) - \alpha(0)m_V(\mu) - \varphi_V(\alpha(0), 0),$$

where we used that $\frac{dx}{d\mu_{V,a}(x)}(x) = e^{\alpha(0)V(x)+\varphi_V(\alpha(0), 0)}$. For the GRF $\Lambda_V(\mu)$, we thus get

$$l_V(\mu) = H(\mu|\mu_{V,a(0)}) + \alpha(0)m_V(\mu) - R$$

$$= -h(\mu) - \alpha(0)m_V(\mu) - \varphi_V(\alpha(0), 0) + \alpha(0)m_V(\mu) - R$$

$$= -h(\mu) + \varphi_V(\alpha(0), 0) - \alpha(0)R.$$

Proof. (of Proposition 3.3) By definition, the Legendre transform $l_V$ of $\Lambda_V$ is given as

$$l_V(\mu) := \sup_{\lambda \in \mathcal{C}_a(\mathbb{R})} \left[ \mathbb{E}_\mu[\lambda] - \varphi_V(\alpha(0), 0) \right] + \varphi_V(\alpha(0), 0) - \alpha(0)R$$

$$= \sup_{\lambda \in \mathcal{C}_a(\mathbb{R})} \left[ \epsilon_{\mu}[\lambda] + \sup_{\beta < 0} \left\{ \beta R - \varphi_V(\beta, \alpha(0), 0) \right\} \right] + \varphi_V(\alpha(0), 0) - \alpha(0)R,$$

where the second equality uses the fact that $\alpha(0)$ minimizes $\beta \mapsto \varphi_V(\beta, \alpha(0)) - \beta R$ (see Lemma 3.1). By rearranging the terms and interchanging the suprema, we get

$$l_V(\mu) = \sup_{\beta < 0} \left[ \beta R + \sup_{\lambda \in \mathcal{C}_a(\mathbb{R})} \left[ \epsilon_{\mu}[\lambda] - \varphi_V(\beta, \alpha(0), 0) \right] \right] + \varphi_V(\alpha(0), 0) - \alpha(0)R$$

$$= \sup_{\beta < 0} \left[ \beta R + H(\mu|\mu_{V,\beta}) - \varphi_V(\alpha(0), 0) \right] + \varphi_V(\alpha(0), 0) - \alpha(0)R. (21)$$

Here we used the variational formula for the relative entropy $H$ given in Proposition 2.6. Next, we take a closer look at the relative entropy $H(\mu|\mu_{V,\beta})$ in order to obtain the desired representation of $l_V$. First, we note that

$$\frac{d\mu_{V,a(0)}}{d\mu}(x) = \exp \left( \varphi_V(\beta, 0) - \varphi_V(\alpha(0), 0) - \beta V(x) + \alpha(0)V(x) \right), \quad x \in \mathbb{R}.$$ 

This leads to

$$H(\mu|\mu_{\beta}) = \int_{\mathbb{R}} \log \left( \frac{d\mu}{d\mu_{\beta}}(x) \right) \mu(dx)$$

$$= \int_{\mathbb{R}} \log \left( \frac{d\mu}{d\mu}(x) \right) \mu(dx) + \int_{\mathbb{R}} \log \left( \frac{d\mu_{V,a(0)}}{d\mu}(x) \right) \mu(dx)$$

$$= H(\mu|\mu_{V,a(0)}) + \varphi_V(\beta, 0) - \varphi_V(\alpha(0), 0) + (\alpha(0) - \beta)m_V(\mu).$$

From Equation (21), it follows that

$$l_V(\mu) = \sup_{\beta < 0} \left[ \beta R + H(\mu|\mu_{V,\beta}) - \varphi_V(\beta, 0) \right] + \varphi_V(\alpha(0), 0) - \alpha(0)R$$

$$= H(\mu|\mu_{V,a(0)}) + \sup_{\beta < 0} \left[ \beta R - \beta m_V(\mu) \right] + \alpha(0)(m_V(\mu) - R).$$

11
\[
\begin{align*}
= & \begin{cases} 
  H(\mu|\nu_{(0)}) + \alpha(0)(m(\mu) - R) & : m(\mu) \leq R \\
  \infty & : \text{else}
\end{cases}
\end{align*}
\]

For the strict convexity, we just note that \( \mu \to H(\mu|\nu_{(0)}) \) is strictly convex (see, e.g., Ex. 5.5 in [26]) and for \( \kappa \in (0,1) \) and \( \mu_1, \mu_2 \in \mathcal{M}_1(\mathbb{R}) \) with \( m(\mu_1) < \infty \), we have that \( m((1-\kappa)\mu_1 + \kappa \mu_2) = (1-\kappa)m(\mu_1) + \kappa m(\mu_2) \).

Proof. (of Theorem 3.5) We are going to prove Theorem 3.5 by applying the version of the Gärtner–Ellis theorem stated in Lemma 2.5. We view \((\mathcal{L}^n, V)_{n \in \mathbb{N}}\) as \( \mathcal{M}(\mathbb{R}) \)-valued random variables, where \( \mathcal{M}(\mathbb{R}) \) is the space of finite signed measures on the Borel \( \sigma \)-algebra on \( \mathbb{R} \). Let \( \mathcal{M}(\mathbb{R})' \) denote the algebraic dual of \( \mathcal{M}(\mathbb{R}) \), i.e., the space of linear functions mapping from \( \mathcal{M}(\mathbb{R}) \) to \( \mathbb{R} \). Recall that \( C_b(\mathbb{R}) \) denotes the space of continuous and bounded functions mapping from \( \mathbb{R} \) to \( \mathbb{R} \). We define

\[\mathcal{V} := \left\{ \hat{\lambda} \in \mathcal{M}(\mathbb{R})' : \exists \lambda \in C_b(\mathbb{R}) : \forall \mu \in \mathcal{M}(\mathbb{R}) : \hat{\lambda}(\mu) = \int \lambda(x) \mu(dx) \right\} \subseteq \mathcal{M}(\mathbb{R})'.\]

The set \( \mathcal{V} \) is separating in \( \mathcal{M}(\mathbb{R}) \), i.e., for \( \nu, \mu \in \mathcal{M}(\mathbb{R}) \) with \( \mu \neq \nu \), we find a \( \hat{\lambda} \in \mathcal{V} \) such that \( \hat{\lambda}(\mu - \nu) \neq 0 \) (see [6] p. 261). The \( \mathcal{V} \)-topology on \( \mathcal{M}(\mathbb{R}) \) is the topology generated by the system of sets

\[U_{\delta, x, \hat{\lambda}} := \{ \mu \in \mathcal{M}(\mathbb{R}) : |\hat{\lambda}(\mu) - x| < \delta \},\]

where \( \hat{\lambda} \in \mathcal{V}, x \in \mathbb{R} \) and \( \delta \in (0, \infty) \). By definition, the relative topology on \( \mathcal{M}_1(\mathbb{R}) \) is the weak topology. By [6] Theorem B.8 we have that \( \mathcal{M}(\mathbb{R}) \) together with the \( \mathcal{V} \)-topology is a locally convex, Hausdorff topological vector space and \( \mathcal{V} \) is the topological dual \( \mathcal{M}(\mathbb{R})^* \) of \( \mathcal{M}(\mathbb{R}) \). Thus, we can identify \( \mathcal{M}(\mathbb{R})^* \) with \( C_b(\mathbb{R}) \). Now the framework for Lemma 2.5 is set up and we can compute the Gärtner–Ellis limit

\[\Lambda_V(\lambda) := \lim_{n \to \infty} \frac{1}{n} \log \mathbb{E} \left[ \exp \left( n \int \lambda(x) \mathcal{L}^n(x) dx \right) \right], \quad \lambda \in C_b(\mathbb{R}), \tag{22}\]

and then show that it is Gateaux-differentiable. This can be done by using an appropriate exponential tilting argument. In the following, we denote by \( X_{R}^{n,V} \) a random variable uniformly distributed on \( \mathbb{B}_{R}^{n,V} \). Then, for the quantity in the Gärtner–Ellis limit (22), where \( \lambda \in C_b(\mathbb{R}) \), we have

\[\frac{1}{n} \log \mathbb{E} \left[ \exp \left( n \int \lambda(x) \mathcal{L}^n(x) dx \right) \right] = \frac{1}{n} \log \mathbb{E} \left[ \exp \left( \sum_{i=1}^{n} \lambda \left( X_{R}^{n,V}(i) \right) \right) \right] = \frac{1}{n} \log \mathbb{E} \left[ \exp \left( \sum_{i=1}^{n} \lambda(x_i) \right) d(x_1, \ldots, x_n) \right] - \frac{1}{n} \log \text{vol}_n \left( \mathbb{B}_{R}^{n,V} \right). \tag{23}\]

We now consider a sequence of iid random variables \( (Z_i)_{i \in \mathbb{N}} \), all distributed according to the Gibbs density \( p : \mathbb{R} \to \mathbb{R}, \)

\[p(x) := e^{q(x) + \lambda(x) - \mathbb{Q}_V(\alpha(\lambda), \lambda)} \]

where \( \alpha(\lambda) \) and \( \mathbb{Q}_V(\alpha(\lambda), \lambda) \) are defined as in Lemma 3.1. Now, we define a new sequence of independent and identically distributed random variables, \( Y_i := V(Z_i) - R, \ i \in \mathbb{N} \). We obtain the identity

\[\int_{\mathbb{B}_{R}^{n,V}} \exp \left( \sum_{i=1}^{n} \lambda(x_i) \right) d(x_1, \ldots, x_n) = \int_{\mathbb{B}_{R}^{n,V}} \exp \left( \sum_{i=1}^{n} \lambda(x_i) \right) \prod_{i=1}^{n} p(x_i)^{-1} \prod_{i=1}^{n} p(x_i) d(x_1, \ldots, x_n) = \mathbb{E} \left[ \prod_{i=1}^{n} \mathbb{P}_{0, x_i} \exp \left( \sum_{i=1}^{n} \lambda(Z_i) - \alpha(\lambda) \sum_{i=1}^{n} V(Z_i) - \sum_{i=1}^{n} \lambda(Z_i) + n \mathbb{Q}_V(\alpha(\lambda), \lambda) \right) \right] \]
Therefore, we obtain the upper bound

\[ \lim_{n \to \infty} \frac{1}{n} \log \left( \int_{\mathbb{R}^n} \exp \left( \sum_{i=1}^n \lambda(x_i) \right) dx_1, \ldots, dx_n \right) \leq \varphi_V(\alpha(\lambda), \lambda) - \alpha(\lambda) R. \]

(24)

For a lower bound, we fix some \( c \in (0, \infty) \) and obtain

\[
\mathbb{E} \left[ \prod_{i=1}^n Y_i \right] < e^{-a(\lambda) \sum_{i=1}^n Y_i} \leq \mathbb{E} \left[ \prod_{i=1}^n Y_i \right] 
\]

\[
\geq \mathbb{E} \left[ \prod_{i=1}^n Y_i \right] e^{-a(\lambda) \sum_{i=1}^n Y_i} 
\]

\[
= e^{a(\lambda) \sum_{i=1}^n Y_i} \cdot \mathbb{E} \left[ - c \sqrt{n} \leq \sum_{i=1}^n Y_i \leq 0 \right]. 
\]

This yields the lower bound

\[
\frac{1}{n} \log \int_{\mathbb{R}^n} \exp \left( \sum_{i=1}^n \lambda(x_i) \right) dx_1, \ldots, dx_n 
\]

\[
\geq \frac{c a(\lambda)}{\sqrt{n}} + \frac{1}{n} \log \mathbb{P} \left[ - c \sqrt{n} \leq \sum_{i=1}^n Y_i \leq 0 \right] + \varphi_V(\alpha(\lambda), \lambda) - \alpha(\lambda) R. 
\]

(25)

By Equation (17) and Equation (18) in Lemma 3.1, we have that \( \mathbb{E}[Y_i] = \mathbb{E}[V(Z_1)] = R = 0 \) and \( \mathbb{V}[Y_i] = \mathbb{V}[V(Z_1)] \) \( \in (0, \infty) \). Hence, we can apply the central limit theorem to the sequence \( (Y_i)_{i \in \mathbb{N}} \). Thus, the probability in Equation (25) tends to \( \mathcal{N}(0, \mathbb{V}[Y_1])([-c, 0]) \in (0, 1) \) as \( n \to \infty \). Combining the upper bound (24) and the lower bound (25), we get

\[
\lim_{n \to \infty} \frac{1}{n} \log \int_{\mathbb{R}^n} \exp \left( \sum_{i=1}^n \lambda(x_i) \right) dx_1, \ldots, dx_n = \varphi_V(\alpha(\lambda), \lambda) - \alpha(\lambda) R. 
\]

(26)

In case of \( \lambda \equiv 0 \), we receive the asymptotic volume of an Orlicz ball with radius \( R \), i.e.,

\[
\lim_{n \to \infty} \frac{1}{n} \log \mathbb{V}_{n,R} = \lim_{n \to \infty} \frac{1}{n} \log \int_{\mathbb{B}^n_R} d(x_1, \ldots, x_n) = \varphi_V(\alpha(0), \lambda) - \alpha(0) R. 
\]

(27)

Putting (26) and (27) together, we obtain from (23) that the Gärtner–Ellis limit \( \Lambda_V : C_b(\mathbb{R}) \to \mathbb{R} \) (see (22)) is given by

\[
\Lambda_V(\lambda) = \varphi_V(\alpha(\lambda), \lambda) - \alpha(\lambda) R - \varphi_V(\alpha(0), \lambda) + \alpha(0) R. 
\]

(28)

In particular, by Lemma 3.1, \( \Lambda_V \) is Fréchet-differentiable. In order to be able to apply Lemma 2.5 it remains to show that the sequence of empirical measures \( (\mathcal{L}^n_{n,R})_{n \in \mathbb{N}} \) is exponentially tight. We define the quantity \( \bar{M} := \sup_{|x| \leq 1} \frac{|x|}{\mathbb{V}[X^n_{R}(x)]} \) and note that \( \bar{M} \in (0, \infty) \), since \( V \) is an Orlicz function. We set \( M := 1 + \bar{M} R \) and let \( X^n_{R} \sim \text{Unif}(\mathbb{B}^n_R) \), for \( n \in \mathbb{N} \). Then, recalling that \( m_1 \) denotes the first moment map, we have

\[
m_1(\mathcal{L}^n_{n,R}) = \frac{1}{n} \sum_{i=1}^n |X^n_{R}(i)| 
\]
by Proposition 3.3, the strictly convex function $I$.

Now consider the situation when there exists another Orlicz function $M$ satisfying the LDP in $(\mathcal{L}^n)_n$.

Thus, for every $n \in \mathbb{N}$, we have $\mathcal{L}^n \subset K^1$, where $K^1$ is the weakly compact set from Proposition 2.16. In other words, we have that

$$\frac{1}{n} \log \mathbb{P} [\mathcal{L}^n \notin (K^1)^c] = -\infty,$$

implying the exponential tightness of $(\mathcal{L}^n)_n$. It now follows from Lemma 3.5 that $(\mathcal{L}^n)_n$ satisfies an LDP in the weak topology on $\mathcal{M}_1(\mathbb{R})$. The GRF is the Legendre transform of $\Lambda_V$, which is, by Proposition 3.3, the strictly convex function $I_V : \mathcal{M}_1(\mathbb{R}) \to [0, \infty]$ with

$$I_V(\mu) = \begin{cases} H(\mu) \mu(0) + \alpha(0) [m_V(\mu) - R] & : m_V(\mu) \leq R \\ \infty & : \text{else} \end{cases}.$$

Now consider the situation when there exists another Orlicz function $W$ such that

$$\frac{V(x)}{W(x)} \to \infty \quad \text{as} \quad |x| \to \infty.$$

Recall the set $K^1$ from Proposition 2.17 and note that $(\mathcal{L}^n)_n \subseteq K^1$. Since $K^1$ is compact in $\mathcal{M}_1^W(\mathbb{R})$ by Proposition 2.17, $(\mathcal{L}^n)_n$ is exponentially tight in $\mathcal{M}_1^W(\mathbb{R})$. By Corollary 2.4, $(\mathcal{L}^n)_n$ satisfies the LDP in $(\mathcal{M}_1^W(\mathbb{R}), d_W)$ with strictly convex GRF $I_V$. This completes the proof.

**Corollary 3.5.** Let $V$ be an Orlicz function and $p \in (1, \infty)$ such that

$$\frac{V(x)}{|x|^p} \to \infty \quad \text{as} \quad |x| \to \infty$$

and let $X^n \sim \text{Unif}([a,b]^n)$. Then, for $1 \leq q < p$, we have that $(n^{-1/p}||X^n||_p)_n$ satisfies an LDP in $\mathcal{M}_1^W(\mathbb{R})$ with GRF $I_V : \mathcal{M}_1(\mathbb{R}) \to [0, \infty]$, where

$$J_p^V(x) := \begin{cases} -\max_{a < 0} \{a - \log \int_R e^{aV(t)} dt\} - \max \{h(v) : v \in \mathcal{M}_1(\mathbb{R}), m_2(v)^{1/p} = x, m_V(v) \leq R\} & : x \geq 0 \\ \infty & : x < 0. \end{cases}$$

**Remark 3.6.** In case of $p = 2$ and $R = 1$, the GRF $I^V_2$ from Corollary 3.5 is the same as the GRF from [15, Section 3]. There, the authors used a different approach, but encountered the same growth condition on $V$, i.e., they assumed

$$\frac{V(x)}{|x|^2} \to \infty \quad \text{as} \quad |x| \to \infty.$$

This is necessary to introduce the exponential change of measure used in the proof of Proposition 3.11 in [15]. In [15] this condition may seem artificial at first, but looking at the proof of Theorem 3.1 one sees that this condition appears in a quite natural way.
Proof. (of Corollary 3.5) Under our assumption, by Theorem A, $\mathcal{L}^{n, V}$ satisfies an LDP in $\mathcal{M}_1^P(\mathbb{R})$ with strictly convex GRF $J_V : \mathcal{M}_1(\mathbb{R}) \to [0, \infty)$, where

$$J_V(\mu) = \left\{ \begin{array}{ll} H(\mu | \mu_{V, a(0)}) + \alpha(0) \left[ m_V(\mu) - R \right] & : m_V(\mu) \leq R \\ \infty & \text{else}. \end{array} \right.$$ 

We can apply the contraction principle to the continuous mapping $m_P : \mathcal{M}_1^P(\mathbb{R}) \to [0, \infty)$ with $m_P(v) = \int_\mathbb{R} |x|^p dV(x)$ (see Remark 2.13) in order to establish an LDP for $(n^{-1} \|X_{n, V}^p\|_p)_{n \in \mathbb{N}}$ with GRF $J^P_V : \mathbb{R} \to [0, \infty)$, given as

$$J^P_V(x) = \left\{ \begin{array}{ll} \inf \left\{ H(\mu | \mu_{V, a(0)}) + \alpha(0) |R - m_V(\mu)| : \mu \in \mathcal{M}_1(\mathbb{R}), m_V(\mu) \leq R, m_P(\mu) = x \right\} & : x \geq 0 \\ \infty & \text{else}. \end{array} \right.$$

Further, because the relative entropy appears, in the infimum we only need to consider measures $\mu$ which are absolutely continuous with respect to $\mu_{V, a(0)}$ (and hence absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$). By Remark 3.1 we get

$$J^V_P(x) = \inf \left\{ -h(\mu) : \mu \in \mathcal{M}_1(\mathbb{R}), m_V(\mu) \leq R, m_P(\mu) = x \right\} + \varphi_V(\alpha(0), 0) = \sup \left\{ h(\mu) : \mu \in \mathcal{M}_1(\mathbb{R}), m_V(\mu) \leq R, m_P(\mu) = x \right\} + \varphi_V(\alpha(0), 0),$$

where we note that the supremum is a maximum, since it is of the form given in Equation (39) in the maximum entropy principle (see Lemma 4.8). Another application of the contraction principle to $(n^{-1} \|X_{n, V}^p\|_p)_{n \in \mathbb{N}}$ via the continuous mapping $x \mapsto x^{1/p}$ gives us the LDP for $(n^{-1/p} \|X_{n, V}^p\|_p)_{n \in \mathbb{N}}$ with the GRF $J^V_P : \mathbb{R} \to [0, \infty)$ given by

$$J^V_P(x) = \left\{ \begin{array}{ll} \max_{\alpha > 0} \left\{ \alpha - \log \int_{\mathbb{R}} e^{aV(t)} d t \right\} - \max_{\alpha \leq 0} \left\{ h(v) : v \in \mathcal{M}_1(\mathbb{R}), m_P(v)^{1/p} = x, m_V(v) \leq R \right\} & : x \geq 0 \\ \infty & : x < 0. \end{array} \right.$$ 

Here we used Lemma 3.1, where we showed that

$$\varphi_V(\alpha(0), 0) = \max_{\alpha > 0} \left\{ \alpha R - \log \int_{\mathbb{R}} e^{aV(t)} d t \right\}.$$

This completes the proof. \hfill \Box

4 Proof of Theorem B

We prove Theorem B by employing the Gibbs conditioning principle (see Proposition 2.8). This leads to the maximization of the entropy function $h$ under certain moment constraints.

For an Orlicz function $V$ and $\alpha \in (-\infty, 0)$, we recall the Gibbs distribution $\mu_{V, \alpha}$ with potential $V$ and critical temperature $\alpha$, which is given by the Lebesgue density

$$\frac{d\mu_{V, \alpha}}{dx}(x) = e^{aV(x) - \varphi_V(\alpha, 0)}, \quad x \in \mathbb{R}.$$ 

Here, $e^{-\varphi_V(\alpha, 0)}$ is the normalizing constant, i.e., $\varphi_V(\alpha, 0) = \log \int_{\mathbb{R}} e^{aV(t)} d t$. Note that in Section 4, $\alpha = \alpha(\lambda)$ depends implicitly on a parameter $\lambda$. In the following, we will only face the situation where $\lambda \equiv 0$ and $\alpha = \alpha(R)$ depends on a parameter $R \in (0, \infty)$.

Lemma 4.1. Let $V_1, V_2$ be two Orlicz functions that satisfy both

$$\frac{V_1(x)}{V_2(x)} \to \infty \quad \text{as} \quad |x| \to \infty \quad \text{and} \quad \int_{\mathbb{R}} V_1(x) e^{\alpha V_2(x)} dx < \infty,$$

15
for all $\alpha \in (-\infty, 0)$. Then, there exists an $\overline{R} \in (0, \infty)$ such that for all $0 < R \leq \overline{R}$, we have
\[
m_{V_1}(\mu_{V_2,R}) \leq m_{V_1}(\mu_{V_2,R}^\alpha) = 1,
\]
where $\alpha(R)$ is the solution of
\[
R = \int_{\mathbb{R}} V_2(x) \mu_{V_2,R}(dx).
\]

Remark 4.2. The quantity $\alpha(R) \in (-\infty, 0)$ from Equation (30) exists and is unique by Lemma 3.1.

Proof. (of Lemma 4.1) Analogously to the first part of the proof of Lemma 3.1, we can show that there exists an $\overline{R} \in (-\infty, 0)$ such that
\[
\int_{\mathbb{R}} V_1(x) e^{\overline{R}_V(x) - \phi_{V_2}(\alpha,0)} dx = 1.
\]
We define
\[
\overline{R} := \int_{\mathbb{R}} V_2(x) e^{\overline{R}_V(x) - \phi_{V_2}(\alpha,0)} dx \in (0, \infty).
\]

Since the solution $\overline{R}$ of Equation (31) is unique by Lemma 3.1, we can write $\overline{R} = \alpha(\overline{R})$. Now take some $R > 0$ with $R \leq \overline{R}$. For the solution $\alpha(R)$ of Equation (30), we have that $\alpha(R) \leq \alpha(\overline{R})$. This holds since the map $\alpha \rightarrow \int_{\mathbb{R}} V_2(x) e^{\alpha V_2(x) - \phi_{V_2}(\alpha,0)} dx$ is monotone increasing (see, e.g., p. 5 in [11]). Thus, if we are able to show that the function
\[
\alpha \rightarrow \int_{\mathbb{R}} V_1(x) e^{\alpha V_2(x) - \phi_{V_2}(\alpha,0)} dx
\]
is monotone increasing on $(-\infty, 0)$, we obtain the claim in Equation (29) since $\alpha(R) \leq \alpha(\overline{R}) = \overline{R}$. Take $\alpha_1, \alpha_2$ with $-\infty < \alpha_1 \leq \alpha_2 < 0$. Then, there exists an $\overline{x} \geq 0$ such that
\[
e^{\alpha_1 V_2(x) - \phi_{V_2}(\alpha_1,0)} \geq e^{\alpha_2 V_2(x) - \phi_{V_2}(\alpha_2,0)} \quad \text{for} \quad x \in [0, \overline{x})
\]
and
\[
e^{\alpha_1 V_2(x) - \phi_{V_2}(\alpha_1,0)} \leq e^{\alpha_2 V_2(x) - \phi_{V_2}(\alpha_2,0)} \quad \text{for} \quad x > \overline{x}.
\]

In the following, we denote by $V_1^{-1}$ the inverse of $V_1$ restricted on the positive real numbers, which exists since $V_1$ is an Orlicz function. Then we get
\[
\int_{\mathbb{R}} V_1(x) e^{\alpha_1 V_2(x) - \phi_{V_2}(\alpha_1,0)} dx
\]
\[
= 2 \int_{0}^{V_1^{-1}(\overline{R})} \int_{V_1(x)}^{\infty} e^{\alpha_1 V_2(t) - \phi_{V_2}(\alpha_1,0)} dt dx
\]
\[
= 2 \int_{0}^{V_1^{-1}(\overline{R})} \int_{V_1(x)}^{\infty} e^{\alpha_1 V_2(t) - \phi_{V_2}(\alpha_1,0)} dt dx + 2 \int_{V_1^{-1}(\overline{R})}^{V_1(0)} \int_{V_1(x)}^{\infty} e^{\alpha_1 V_2(t) - \phi_{V_2}(\alpha_1,0)} dt dx
\]
\[
= 2 \int_{0}^{V_1^{-1}(\overline{R})} \left[ \frac{1}{2} - \int_{0}^{V_1(x)} e^{\alpha_1 V_2(t) - \phi_{V_2}(\alpha_1,0)} dt \right] dx + 2 \int_{V_1^{-1}(\overline{R})}^{V_1(0)} \int_{V_1(x)}^{\infty} e^{\alpha_1 V_2(t) - \phi_{V_2}(\alpha_1,0)} dt dx
\]
\[
\leq \int_{0}^{V_1^{-1}(\overline{R})} \left[ \frac{1}{2} - \int_{0}^{V_1(x)} e^{\alpha_2 V_2(t) - \phi_{V_2}(\alpha_2,0)} dt \right] dx + 2 \int_{V_1^{-1}(\overline{R})}^{V_1(0)} \int_{V_1(x)}^{\infty} e^{\alpha_2 V_2(t) - \phi_{V_2}(\alpha_2,0)} dt dx
\]
\[
= \int_{0}^{V_1^{-1}(\overline{R})} \left[ \frac{1}{2} - \int_{0}^{V_1(x)} e^{\alpha_1 V_2(t) - \phi_{V_2}(\alpha_1,0)} dt \right] dx + 2 \int_{V_1^{-1}(\overline{R})}^{V_1(0)} \int_{V_1(x)}^{\infty} e^{\alpha_2 V_2(t) - \phi_{V_2}(\alpha_2,0)} dt dx.
\]

Remark 4.3. In the proof of Lemma 4.1 we have shown that $\alpha \rightarrow m_{V_1}(\mu_{V_2,a})$ is monotone increasing. The same argument can be used to prove that $\alpha \rightarrow m_{V_1}(\mu_{V_2,a})$ and $\alpha \rightarrow m_{V_1}(\mu_{V_2,a})$ are monotone increasing. Here, note that the latter two functions are well-defined even if we do not assume that
\[
\int_{\mathbb{R}} V_1(x) e^{\alpha V_2(x) - \phi_{V_2}(\alpha,0)} dx < \infty
\]
for all $\alpha \in (-\infty, 0)$.  

16
In order to get a nice geometric interpretation of Theorem B we need the following limit theorem from [11]. We recall that \( \lambda^{(n,k)}_{R,V} \) is the distribution of the first \( k \) components of a random vector uniformly distributed on the Orlicz ball \( B^{n}_{R} \), where \( R \in (0,\infty) \). The distribution \( \lambda^{(n,k)}_{R,V} \) has a Lebesgue density on \( \mathbb{R}^{k} \) (see p.6 in [11]) which we shall denote by \( \frac{d\lambda^{(n,k)}_{R,V}}{dx} \). Further, for an Orlicz function \( V \) and \( \alpha \in (-\infty,0) \), we denote by \( \mu^{\kappa}_{V,\alpha} \) the \( k \)-fold product measure of \( \mu_{V,\alpha} \).

**Proposition 4.4.** (Theorem A in [11]) Let \( V \) be an Orlicz function, \( \alpha \in (-\infty,0), R \in (0,\infty), k \in \mathbb{N} \). Then, there exists a constant \( C \in (0,\infty) \) such that

\[
\int_{\mathbb{R}} \left| \frac{d\lambda^{(n,k)}_{R,V}}{dx}(x) - \frac{d\mu^{\kappa}_{V,\alpha}}{dx}(x) \right| dx \leq \frac{Ck}{n},
\]

(34)

where \( \alpha(R) \in (-\infty,0) \) is the unique solution of

\[
R = \int_{\mathbb{R}} V(x) e^{\alpha(R)V(x)} - \rho_{V}(\alpha(R),0) \, dx.
\]

(35)

**Remark 4.5.** In particular, Equation (34) in Proposition 4.4 implies convergence in distribution of \( \lambda^{(n,k)}_{R,V} \) towards \( \mu^{\kappa}_{V,\alpha} \) as \( n \to \infty \).

**Remark 4.6.** In [11], the authors considered a more general class of Orlicz functions, for our purpose the version stated Proposition 4.4 suffices.

**Proposition 4.7.** Let \( V_{1} \) and \( V_{2} \) be two Orlicz functions which satisfy

\[
\frac{V_{1}(x)}{V_{2}(x)} \to \infty \quad \text{as} \quad |x| \to \infty.
\]

Suppose that \( X^{n}_{V_{1}} \sim \text{Unif}([B^{n}_{1},V_{1}]) \) and denote by \( (\mathcal{L}^{n}_{V_{1}})_{n \in \mathbb{N}} \) the associated sequence of empirical measures. For \( R \in (0,\infty) \), we define \( C := [0,R] \subseteq \mathbb{R} \) and for \( \varepsilon > 0 \), let \( C_{\varepsilon} := [-\varepsilon,R+\varepsilon] \). Then, the measure

\[
\nu_{*} := \arg\max \{ h(v) : m_{V_{1}}(v) \leq 1, m_{V_{2}}(v) \in C \},
\]

(36)

which is the measure maximizing the entropy under the constraints \( m_{V_{1}}(v) \leq 1, m_{V_{2}}(v) \in C \), exists, is unique and

\[
\lim_{\varepsilon \to 0} \lim_{n \to \infty} P[\mathcal{L}^{n}_{V_{1}} \in \cdot | m_{V_{2}}(\mathcal{L}^{n}_{V_{1}}) \in C_{\varepsilon}] = \delta_{\nu_{*}},
\]

(37)

where \( \delta_{\nu_{*}} : \mathcal{M}_{1}(\mathbb{R}) \to \mathbb{R} \) with

\[
\delta_{\nu_{*}}(\mu) = \begin{cases} 1 & : \mu = \nu_{*} \\ 0 & \text{else.} \end{cases}
\]

Moreover, for \( k \in \mathbb{N} \),

\[
P \left[ (X^{n}_{V_{1}}(1),...,X^{n}_{V_{1}}(k)) \in \cdot \left| \frac{1}{n} \sum_{i=1}^{n} V_{2}(X^{n}_{V_{1}}(i)) \in C_{\varepsilon} \right| \right] \overset{d}{\to} \nu_{*}^{\kappa},
\]

(38)

as \( n \to \infty \) followed by \( \varepsilon \to 0 \) and \( \nu_{*}^{\kappa} \) is the \( k \)-fold product measure of \( \nu_{*} \).

**Proof.** For the proof we mainly use the Gibbs conditioning principle from Proposition 2.8. Due to our assumptions, \( (\mathcal{L}^{n}_{V_{1}})_{n \in \mathbb{N}} \) satisfies an LDP in \((\mathcal{M}_{1}(\mathbb{R}),d_{V_{1}})\) by Theorem A. Further, we recall that the map \( m_{V_{2}} : \mathcal{M}_{1}(\mathbb{R}) \to \mathbb{R} \) is continuous. Then, for \( \varepsilon > 0 \), the set \( F_{\varepsilon} := \{ \nu \in \mathcal{M}_{1}(\mathbb{R}) \mid m_{V_{2}}(v) \in C_{\varepsilon} \} \) is closed in the \( d_{V_{2}} \)-topology on \( \mathcal{M}_{1}(\mathbb{R}) \). Moreover, the set \( F := \bigcap_{\varepsilon > 0} F_{\varepsilon} \) is closed as an intersection of closed sets and, for \( \varepsilon > 0 \), we have

\[
F = m_{V_{2}}^{-1}(C) \subseteq m_{V_{2}}^{-1}(C_{\varepsilon}) \subseteq [m_{V_{2}}^{-1}(C_{\varepsilon})]^{0} = F_{\varepsilon},
\]

17
where the second inclusion holds since $m_{V_2}$ is continuous. This shows properties (2) and (4) of Proposition 3.3. For Property (3) we need to check that $P[\mathcal{L}^{n,V_1} \in F_\epsilon] > 0$ for all sufficiently large $n \in \mathbb{N}$ and all $\epsilon > 0$. Since $(\mathcal{L}^{n,V_1})_{n \in \mathbb{N}}$ satisfies an LDP in $(\mathcal{M}_1^V(\mathbb{R}),d_{V_2})$ with strictly convex GRF $\lambda_{V_1}$ given in Equation (20), we have that

$$\liminf_{n \to \infty} \frac{-1}{n} \log P[\mathcal{L}^{n,V_1} \in F_\epsilon] \geq - \inf_{v \in F_\epsilon} \lambda_{V_1}(v) \geq - \inf_{v \in F} \lambda_{V_1}(v).$$

If we are able to show that there exists a measure $v \in F$ with $\lambda_{V_1}(v) < \infty$, then it follows that $P[\mathcal{L}^{n,V_1} \in F_\epsilon] > 0$ for sufficiently large $n \in \mathbb{N}$ and all $\epsilon > 0$. For $\alpha \in (-\infty,0)$, by Remark 4.3 both mappings $\alpha \mapsto m_{V_1}(\mu_{V_1,\alpha})$ and $\alpha \mapsto m_{V_2}(\mu_{V_1,\alpha})$ are monotone increasing and we have that $\lim_{\alpha \to -\infty} m_{V_1}(\mu_{V_1,\alpha}) = \lim_{\alpha \to -\infty} m_{V_2}(\mu_{V_1,\alpha}) = 0$. Thus, there exists an $\overline{\alpha} \in (-\infty,0)$ such that $m_{V_1}(\mu_{V_1,\overline{\alpha}}) \leq 1$ and $m_{V_2}(\mu_{V_1,\overline{\alpha}}) \leq R$ and hence $\mu_{V_1,\overline{\alpha}} \in F$. The distribution $\mu_{V_1,\overline{\alpha}}$ is absolutely continuous with respect to Lebesgue measure and thus by Remark 3.3 we have

$$\lambda_{V_1}(\mu_{V_1,\overline{\alpha}}) = -h(\mu_{V_1,\overline{\alpha}}) + c,$$

where $c \in \mathbb{R}$ denotes some finite constant and $h$ is the entropy function. For $-h(\mu_{V_1,\overline{\alpha}})$, we have that

$$-h(\mu_{V_1,\overline{\alpha}}) = \int_{\mathbb{R}} e^{\overline{\alpha} V_1(x) - \psi_{V_1}(\overline{\alpha},0)} \log \left(e^{\overline{\alpha} V_1(x) - \psi_{V_1}(\overline{\alpha},0)}\right) dx = -\psi_{V_1}(\overline{\alpha},0) + \overline{\alpha} m_{V_1}(\mu_{V_1,\overline{\alpha}})$$

$$< \infty,$$

showing that $\lambda_{V_1}(\mu_{V_1,\overline{\alpha}}) < \infty$.

For property (1), we observe that $F$ is closed and convex. Indeed, for any $v_1, v_2 \in F$ and $\lambda \in [0,1]$, we have

$$m_{V_2}(\lambda v_1 + (1-\lambda) v_2) = \lambda m_{V_2}(v_1) + (1-\lambda) m_{V_2}(v_2) \leq 0,$$

where we used that $m_{V_2}(v_i) \in [0,R]$ for $i = 1,2$. Moreover, we have shown that $\inf_{v \in F} \lambda_{V_1}(v) < \infty$ and since $\lambda_{V_1}$ is a strictly convex GRF, $\lambda_{V_1}$ uniquely attains its minimum over such a set $F$. In total we get property (1) from Proposition 2.8 and that $\mathcal{M}_F$ is a singleton with $\mathcal{M}_F = \{v_\ast\}$, where

$$v_\ast = \arg \min \{\lambda_{V_1}(v) : v \in F\} = \arg \max \{h(v) : v \in \mathcal{M}_1^V(\mathbb{R}), m_{V_1}(v) \leq 1, m_{V_2}(v) \in C\}.$$

The equality uses Remark 3.4 and the observation that in the minimization only those $v \in \mathcal{M}_1^V(\mathbb{R})$ are relevant which are absolutely continuous with respect to Lebesgue measure and $m_{V_1}(v) \leq 1$ and $m_{V_2}(v) \in C$. The Gibbs conditioning principle (more precisely 3) now gives us the limit in Equation 37. We now want to employ Proposition 2.11 to derive the limit in Equation 39 from the limit

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} P[\mathcal{L}^{n,V_1} \in \cdot | m_{V_2}(\mathcal{L}^{n,V_1}) \in C_\epsilon] = \delta_{v_\ast}(\cdot).$$

We introduce the family of measures $(\mu_\epsilon^n)_{n \in \mathbb{N}, \epsilon > 0}$, where, for $n \in \mathbb{N}$ and $\epsilon > 0$, we define

$$\mu_\epsilon^n(\cdot) := P\left[X_{1,n,V_1}^{n,\epsilon} \in \cdot | X_1^{n,V_1} \in \mathcal{B}_{1+\epsilon}^{n,V_2}\right].$$

Here, for $n \in \mathbb{N}$, $X_{1,n,V_1}^{n,\epsilon} \sim \text{Unif}(\mathcal{B}_{1+\epsilon}^{n,V_2})$ (and hence, $\mathcal{L}^{n,V_1}$ is the empirical measure associated to $X_{1,n,V_1}^{n,\epsilon}$). We note that the family $(\mu_\epsilon^n)_{n \in \mathbb{N}, \epsilon > 0}$ is a family of symmetric probability measures, since, by the symmetry of the Orlicz function $V_1$, $X_{1,n,V_1}^{n,\epsilon}$ has a symmetric distribution.

For $n \in \mathbb{N}$ and $\epsilon > 0$, we consider random variables $X_\epsilon^n$ with distribution $\mu_\epsilon^n$ and let $\mathcal{L}_\epsilon^n$ be the associated empirical distribution, i.e.,

$$\mathcal{L}_\epsilon^n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_\epsilon^n(i)}.$$
Then, we have
\[
\delta_{\nu_*}(\cdot) = \lim_{\epsilon \to 0} \lim_{n \to \infty} \mathbb{P} \left[ \mathcal{L}^n_{V_1} \in \cdot \mid m_{V_1}(\mathcal{L}^n_{V_1}) \in C_\epsilon \right]
\]
\[
= \lim_{\epsilon \to 0} \lim_{n \to \infty} \mathbb{P} \left[ \mathcal{L}^n_{V_1} \in \cdot \mid X^n_{1,V_1} \in \mathbb{P}^n_{R+\epsilon} \right] = \lim_{\epsilon \to 0} \lim_{n \to \infty} \mathbb{P} \left[ \mathcal{L}^n_{V_1} \in \cdot \right].
\]

Here, we used that the event \(m_{V_1}(\mathcal{L}^n_{V_1}) \in C_\epsilon\) can be expressed as \(X^n_{1,V_1} \in \mathbb{P}^n_{R+\epsilon}\). Thus, by Proposition 2.11 and Remark 2.10 we get that the marginal distribution of the first \(k\) coordinates of \(\nu_*^n\) converges weakly to \(\nu_*^k\). In other words, we have
\[
\nu_*^k(\cdot) = \lim_{\epsilon \to 0} \lim_{n \to \infty} \mathbb{P} \left[ (X^n_{1,V_1}(1), \ldots, X^n_{1,V_1}(k)) \in \cdot \right]
\]
\[
= \lim_{\epsilon \to 0} \lim_{n \to \infty} \mathbb{P} \left[ (X^n_{1,V_1}(1), \ldots, X^n_{1,V_1}(k)) \in \cdot \mid X^n_{V_1} \in \mathbb{P}^n_{R+\epsilon} \right]
\]
\[
= \lim_{\epsilon \to 0} \lim_{n \to \infty} \mathbb{P} \left[ (X^n_{1,V_1}(1), \ldots, X^n_{1,V_1}(k)) \in \cdot \right]
\]
\[
\left( \frac{1}{n} \sum_{i=1}^{n} V_2(\chi^n_{V_1}(i)) \in C_\epsilon \right).
\]

This completes the proof. \(\square\)

In the preceding Proposition 4.7 we have seen that the limiting distribution \(\nu_*\) follows from a maximum entropy principle. In order to compute \(\nu_*\), we need the following classical result, which relies on Ex. 5.3 in [4] and chapter 12 in [5]. We state it in the version given in [16].

**Lemma 4.8.** Fix \(r_i : \mathbb{R} \to \mathbb{R}, \alpha_i \in \mathbb{R}, \) for \(i = 1, \ldots, m\) and \(s_j : \mathbb{R} \to \mathbb{R}, \beta_j \in \mathbb{R}, \) for \(j = 1, \ldots, n\). Consider the following maximization problem

\[
\max_{\nu \in \mathcal{M}_1(\mathbb{R})} h(\nu)
\]

subject to

\[
\int_{\mathbb{R}} r_i(x) d\nu(x) = \alpha_i \text{ for } i = 1, \ldots, m
\]

\[
\int_{\mathbb{R}} s_j(x) d\nu(x) \leq \beta_j \text{ for } j = 1, \ldots, n.
\]

Then, a probability measure \(\nu_* \in \mathcal{M}_1(\mathbb{R})\) attains the maximum in (39) if and only if it is of the following form

\[
\nu_*(dx) = \exp \left\{ -1 - \kappa_0^* - \sum_{i=1}^{m} \lambda_i^* r_i(x) - \sum_{i=1}^{n} \mu_j^* s_j(x) \right\} dx,
\]

with non-negative constants \(\kappa_0^*, (\lambda_i^*_j)_{i=1}^{m}, (\mu_j^*_i)_{j=1}^{n}\) chosen such that \(\nu_*\) lies in \(\mathcal{M}_1(\mathbb{R})\) and satisfies the constraints in (39). Moreover, if \(\nu_*\) attains the maximum in (39), then

\[
\left( \int_{\mathbb{R}} s_j(x) d\nu_*(x) - \beta_j \right) \mu_j^* = 0,
\]

for all \(j = 1, \ldots, n\).

**Proof.** (of Theorem 4.6) In Proposition 4.7 we have seen that the limit in distribution of the family of measures \(\lambda_{n,V_1(V_2)}^{(R+\epsilon,k)}\) is given by \(\nu_*^k\), where \(\nu_*\) is given as

\[
\nu_* = \arg\max \{ h(\nu) : m_{V_1}(\nu) \leq 1, m_{V_2}(\nu) \in [0,R] \}.
\]

By Lemma 4.1 there exists a \(R \in (0,\infty)\) such that \(m_{V_1}(\mu_{V_2,a(R)}) = 1\) and \(m_{V_2}(\mu_{V_2,a(R)}) = \mathbb{R}\). We claim that for \(0 < R \leq \mathbb{R}\), the maximizing measure in (42) is the Gibbs distribution \(\mu_{V_2,a(R)}\) with Lebesgue density

\[
\frac{d\mu_{V_2,a(R)}}{dx}(x) = e^{a(R)\psi_{V_2}(a(R),0)}, \quad x \in \mathbb{R}.
\]
First, we observe that the density of \( \mu_{V_2,a(R)} \) is of the form in Equation (40) from Lemma 4.3 (with the choice of \( n = m = 1, \kappa^*_n = 1 + q_{V_2}(a(R),0) \in \mathbb{R}, \lambda^*_n = -a(R) \) and \( \mu^*_1 = 0 \)). By assumption, we have that \( R \leq \tilde{R} \) and hence \( m_{V_1}(\mu_{V_2,a(R)}) \leq 1 \) (see Equation (29) in Lemma 4.1) and by the choice of \( a(R) \), we get that \( m_{V_1}(\mu_{V_2,a(R)}) = R \). Hence, by Lemma 4.8, we have that \( \nu_* = \mu_{V_2,a(R)} \). Let \( d_w \) be some metric inducing weak convergence of probability measures (e.g., the bounded Lipschitz metric). Then, by Proposition 4.4 and Remark 4.10,

\[
d_w(\lambda^{(R+\epsilon,k)}_{n,V_1,V_2}, \lambda^{(n,k)}_{n,V_1,V_2}) \leq d_w(\lambda^{(R+\epsilon,k)}_{n,V_1,V_2}, \mu_{\lambda^{(n,k)}_{n,V_1,V_2}}) + d_w(\mu_{\lambda^{(n,k)}_{n,V_1,V_2}}, \mu_{\lambda^{(n,k)}_{n,V_1,V_2}}) \to 0
\]
as \( n \to \infty \) followed by \( \epsilon \to 0 \). \( \Box \)

At this point one might ask what happens in the other cases, i.e., when the quantity \( R \in (0, \infty) \) in the definition of the Orlicz ball \( \mathbb{B}_R^{V_2} \) is „big”. This leads to the following corollary of Proposition 4.7.

**Corollary 4.9.** Let \( V_1, V_2 \) be two Orlicz functions that satisfy

\[
\frac{V_1(x)}{V_2(x)} \to \infty \quad \text{as} \quad |x| \to \infty.
\]

Take some \( R \in (\tilde{R}, \infty) \) with

\[
\tilde{R} := \sup_{v \in K^{V_1}_{\lambda^*_1}} m_{V_1}(v), \tag{43}
\]

where \( K^{V_1}_{\lambda^*_1} = m_{V_1}^{-1}([0, 1]) \) is the set from Proposition 2.17. Then, for fixed \( k \in \mathbb{N} \), we have the weak limit

\[
\lim_{\epsilon \to 0} \lim_{n \to \infty} d_w(\lambda^{(R+\epsilon,k)}_{n,V_1,V_2}, \lambda^{(n,k)}_{n,V_1,V_2}) = 0,
\]

where \( d_w \) is some metric inducing weak convergence of probability measures and \( \lambda^{(n,k)}_{n,V_1,V_2} \) is the distribution of the first \( k \) components of a uniform distribution on \( \mathbb{B}^{V_2}_{1/\lambda^*_1} \).

**Remark 4.10.** By Proposition 2.17 \( K^{V_1}_{\lambda^*_1} \) is compact in \( (\mathcal{M}^{V_2}_1(\mathbb{R}), d_{V_2}) \) and \( m_{V_2}: \mathcal{M}_1^{V_2}(\mathbb{R}) \to \mathbb{R} \) is continuous, which implies that \( R \) from Equation (43) is finite.

**Proof.** (of Corollary 4.9) By Proposition 4.7 we have that the limit in distribution, as \( n \to \infty \) followed by \( \epsilon \to 0 \), of \( \lambda^{(R+\epsilon,k)}_{n,V_1,V_2} \) is given by \( \nu^{*k} \), where

\[
\nu_* := \arg\max \{ h(v) : v \in \mathcal{M}^{V_2}_1(\mathbb{R}), m_{V_1}(v) \leq 1, m_{V_2}(v) \leq R \}.
\]

Since we assume \( R \geq \tilde{R} \), we see that for each \( v \in K^{V_1}_{\lambda^*_1} \), the inequality \( m_{V_1}(v) \leq 1 \) implies \( m_{V_2}(v) \leq R \).

Thus \( \nu_* \) is of the form \( \nu_* = \arg\max \{ h(v) : v \in \mathcal{M}^{V_2}_1(\mathbb{R}), m_{V_1}(v) \leq 1 \} \), where the maximum is attained by the Gibbs measure \( \mu_{V_1,a(1)} \) (see proof of Theorem 3). By Proposition 4.4 we have that \( \lambda^{(n,k)}_{1,V_1} \) converges in distribution to \( \mu_{\lambda^{(n,k)}_{1,V_1}} \) as \( n \to \infty \) and thus

\[
\lim_{\epsilon \to 0} \lim_{n \to \infty} d_w(\lambda^{(R+\epsilon,k)}_{n,V_1,V_2}, \lambda^{(1,k)}_{1,V_1}) \leq \lim_{\epsilon \to 0} \lim_{n \to \infty} d_w(\lambda^{(R+\epsilon,k)}_{n,V_1,V_2}, \mu_{\lambda^{(1,k)}_{1,V_1}}) + \lim_{n \to \infty} d_w(\mu_{\lambda^{(n,k)}_{1,V_1}}) = 0,
\]

which completes the proof. \( \Box \)

**Remark 4.11.** In the „intermediate“ case, where \( R \in (\tilde{R}, \tilde{R}) \) with \( \tilde{R} \) being the quantity from Theorem 2 and \( \tilde{R} \) from Equation (43), we get that the limit in distribution of \( \lambda^{(R+\epsilon,k)}_{n,V_1,V_2} \) is given by \( \nu^{*k} \). \( \nu_* \) is some distribution proportional to

\[
e^{-\mu^*_1(V_1(x)) - \mu^*_V(V_2(x))}, \quad x \in \mathbb{R}.
\]

The constants \( \mu^*_1 \geq 0 \) follow from the maximum entropy principle Lemma 4.8. In total, one can say that the limiting distribution of \( \lambda^{(R+\epsilon,k)}_{n,V_1,V_2} \) undergoes a phase transition depending on the magnitude of the parameter \( R \).
Acknowledgement

LF and JP are supported by the Austrian Science Fund (FWF) Project P32405 Asymptotic geometric analysis and applications and FWF Project F5513-N26, which is a part of the Special Research Program Quasi-Monte Carlo Methods: Theory and Applications.

References

[1] D. Alonso-Gutiérrez and J. Prochno. Thin-shell concentration for random vectors in Orlicz balls via moderate deviations and Gibbs measures. *J. Funct. Anal.*, 282(1):Paper No. 109291, 2022.

[2] F. Barthe and P. Wolff. Volume properties of high-dimensional Orlicz balls. *arxiv preprint arXiv:2106.01675*, 2021.

[3] E. Borel. *Introduction géométrique à quelques théories physiques*. Cornell University Library historical math monographs. Gauthier-Villars, 1914.

[4] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004.

[5] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2006.

[6] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.

[7] P. Diaconis and D. Freedman. A dozen de Finetti-style results in search of a theory. *Ann. Inst. H. Poincaré Probab. Statist.*, 23(2, suppl.):397–423, 1987.

[8] P. Dupuis, V. Laschos, and K. Ramanan. Large deviations for configurations generated by Gibbs distributions with energy functionals consisting of singular interaction and weakly confining potentials. *Electron. J. Probab.*, 25, 2020.

[9] R. S. Ellis. *Entropy, large deviations, and statistical mechanics*, volume 271 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1985.

[10] C. Hao, A. Kamińska, and N. Tomczak-Jaegermann. Orlicz spaces with convexity or concavity constant one. *Journal of Mathematical Analysis and Applications*, 320(1):303–321, 2006.

[11] S. G. G. Johnston and J. Prochno. A Maxwell principle for generalized Orlicz balls. *arXiv preprint arXiv:2012.11568*, 2020.

[12] Z. Kabluchko and J. Prochno. The maximum entropy principle and volumetric properties of Orlicz balls. *Journal of Mathematical Analysis and Applications*, 495(1):124687, 2021.

[13] Z. Kabluchko, J. Prochno, and C. Thäle. High-dimensional limit theorems for random vectors in $\ell^p_n$-balls. *Commun. Contemp. Math.*, 21(1):1750092, 30, 2019.

[14] A. Kamińska. The criteria for local uniform rotundity of Orlicz spaces. *Studia Math.*, 79(3):201–215, 1984.

[15] S. S. Kim, Y.-T. Liao, and K. Ramanan. An asymptotic thin shell condition and large deviations for random multidimensional projections. *arXiv preprint arXiv:1912.13447*, 2021.
S. S. Kim and K. Ramanan. A conditional limit theorem for high-dimensional $\ell^p$-spheres. *J. Appl. Probab.*, 55(4):1060–1077, 2018.

P. Kosmol and D. Müller-Wichards. *Optimization in function spaces: with stability considerations in Orlicz spaces*, volume 13. Walter de Gruyter, 2011.

S. Kwapien and C. Schütt. Some combinatorial and probabilistic inequalities and their application to Banach space theory. *Studia Math.*, 82(1):91–106, 1985.

S. Lang. *Fundamentals of differential geometry*, volume 191. Springer Science & Business Media, 2012.

C. Léonard. Entropic projections and dominating points. *ESAIM Probab. Stat.*, 14:343–381, 2010.

J. Lindenstrauss and L. Tzafriri. *Classical Banach spaces. I*. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 92. Springer-Verlag, Berlin-New York, 1977. Sequence spaces.

A. A. Mogul’skii. de Finetti-type results for $l^p_n$. *Sibirsk. Mat. Zh.*, 32(4):88–95, 228, 1991.

A. Naor and D. Romik. Projecting the surface measure of the sphere of $\ell^n_p$. *Ann. Inst. H. Poincaré Probab. Statist.*, 39(2):241–261, 2003.

J. Prochno and C. Schütt. Combinatorial inequalities and subspaces of $L_1$. *Studia Math.*, 211(1):21–39, 2012.

S. T. Rachev and L. Rüschendorf. Approximate independence of distributions on spheres and their stability properties. *Ann. Probab.*, 19(3):1311–1337, 1991.

F. Rassoul-Agha and T. Seppäläinen. *A course on large deviations with an introduction to Gibbs measures*, volume 162 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2015.

Y. Raynaud and C. Schütt. Some results on symmetric subspaces of $L_1$. *Studia Math.*, 89(1):27–35, 1988.

G. Schechtman and M. Schmuckenschläger. Another remark on the volume of the intersection of two $L^n_p$ balls. In *Geometric aspects of functional analysis (1989–90)*, volume 1469 of *Lecture Notes in Math.*, pages 174–178. Springer, Berlin, 1991.

G. Schechtman and J. Zinn. On the volume of the intersection of two $L^n_p$ balls. *Proc. Amer. Math. Soc.*, 110(1):217–224, 1990.

C. Schütt. On the embedding of 2-concave Orlicz spaces into $L^1$. *Studia Math.*, 113(1):73–80, 1995.

A. Sznitman. Topics in propagation of chaos. In *École d’Été de Probabilités de Saint-Flour XIX—1989*, volume 1464 of *Lecture Notes in Math.*, pages 165–251. Springer, Berlin, 1991.

C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.

LORENZ FRÜHWIRTH: Institute of Mathematics and Scientific Computing, University of Graz, Heinrichstraße 36, 8010 Graz, Austria

22
E-mail: lorenz.fruehwirth@uni-graz.at

JOSCHA PROCHNO: Faculty of Computer Science and Mathematics, University of Passau, Innstraße 33, 94032 Passau, Germany.
E-mail: joscha.prochno@uni-passau.de