Advances in Applied Mathematics 134 (2022) 102306

Contents lists available at ScienceDirect

Advances in Applied Mathematics

www.elsevier.com/locate/yaama

An asymptotic thin shell condition and large deviations for random multidimensional projections

Steven Soojin Kim, Yin-Ting Liao, Kavita Ramanan *,1

Brown University, United States of America

A R T I C L E   I N F O

Article history:
Received 19 October 2020
Received in revised form 1 November 2021
Accepted 8 November 2021
Available online xxxx

MSC:
60F10
52A23
46B06

Keywords:
Large deviations
Random projections
Stiefel manifold
Rate function
Thin shell condition
Asymptotic thin shell condition
Central limit theorem for convex sets
ℓ_p^n balls
Orlicz balls
Gibbs measures
KLS conjecture

A B S T R A C T

It is well known that fluctuations of marginals of high-dimensional random vectors that satisfy a certain concentration estimate called the thin shell condition are approximately Gaussian. In this article we identify a general condition on a sequence of high-dimensional random vectors under which one can identify the exponential decay rate of large deviation probabilities of the corresponding sequence of marginals. More precisely, consider the projection of an n-dimensional random vector onto a random k_n-dimensional basis, $k_n \leq n$, drawn uniformly from the Haar measure on the Stiefel manifold of orthonormal k_n-frames in $\mathbb{R}^n$, in three different asymptotic regimes as $n \to \infty$: "constant" ($k_n = k$), "sublinear" ($k_n \to \infty$ but $k_n/n \to 0$) and "linear" ($k_n/n \to \lambda$ with $0 < \lambda \leq 1$). When the sequence of random vectors satisfies a certain "asymptotic thin shell condition", we establish large deviation principles for the corresponding sequence of random projections in the constant regime, and for the sequence of empirical measures of the coordinates of the random projections in the sublinear and linear regimes. We also establish large deviation principles for scaled ℓ_q norms of the random projections in all three regimes. Moreover, we show that the asymptotic thin shell condition holds for various sequences of random vectors of interest, including the uniform measure on suitably scaled ℓ_p^n balls, for $p \in [1, \infty)$, and generalized Orlicz balls defined via a superquadratic function, as well as a class

* Corresponding author.
E-mail addresses: steven_kim@alumni.brown.edu (S.S. Kim), yin-ting_liao@brown.edu (Y.-T. Liao), kavita_ramanan@brown.edu (K. Ramanan).

[1] Ramanan and Y.-T. Liao were supported by the National Science Foundation under grants DMS-1713032 and DMS-1954351.

https://doi.org/10.1016/j.aam.2021.102306

of Gibbs measures with superquadratic interaction potential. Along the way, we obtain logarithmic asymptotics of volumes of high-dimensional Orlicz balls, which may be of independent interest. We also show that the decay rate of large deviation probabilities of Euclidean norms of multi-dimensional projections of ℓ_p^n balls, when $p \in [1, 2)$, exhibits an unexpected phase transition in the sublinear regime, thus disproving an earlier conjecture due to Alonso-Gutiérrez et al. Random projections of high-dimensional random vectors are of interest in a range of fields including asymptotic convex geometry and high-dimensional statistics.

1. Introduction

1.1. Motivation and context

The study of high-dimensional probability distributions through their lower-dimensional projections, especially random projections, is a common theme in a wide range of areas, including geometric functional analysis [25], statistics and data analysis [15,18], information retrieval [24,48], machine learning [42] and asymptotic geometric analysis [6,34]. In the latter case, the typical probability measure of interest is the uniform distribution on a high-dimensional convex body (i.e., a compact convex set with non-empty interior). Questions about the geometry of convex bodies in high dimensions often take on a certain probabilistic flavor. A significant result in this direction is the so-called central limit theorem (CLT) for convex sets, which roughly says that most k-dimensional projections (equivalently, marginals) of an n-dimensional isotropic convex body are close to Gaussian in the total variation distance, when n is sufficiently large and k is of a smaller order than n^α for some universal constant $\alpha \in (0, 1)$. Although foreshadowed by results of Sudakov [52], Diaconis and Freedman [15] and von Weizsäcker [54], the conjecture that most marginals are Gaussian was precisely formulated by Anttila, Ball and Perissinaki in [6] (see also [11]). Specifically, they showed that if the Euclidean norm of a symmetric high-dimensional random vector $X^{(n)}$ satisfies a certain concentration estimate referred to as the "thin shell" condition, then "most" of its marginals are approximately Gaussian. They also verified this condition for random vectors uniformly distributed on a certain class of convex sets whose modulus of convexity and diameter satisfy certain assumptions. Subsequently, the thin shell condition was verified for various classes of convex bodies by several authors, with a breakthrough verification due to Klartag [34,35] for any isotropic log-concave distribution, which, in particular, includes the uniform distribution on an isotropic convex body (see also [19] for a simplified proof). The result of [6] was further extended to general high-dimensional measures by Meckes [43], who showed that whenever an n-dimensional random vector satisfies a quantitative version of the thin shell condition, then most k-dimensional marginals are close to

Gaussian (in the bounded-Lipschitz distance) if $k < 2\log n/\log\log n$, and that the latter cutoff for k is in some sense the best possible.

One broad aim of studying lower-dimensional projections is to obtain information about less tractable high-dimensional measures. While the central limit theorem for convex sets and related theorems are beautiful universality results, they imply the somewhat negative result that (fluctuations of) most lower-dimensional projections do not provide much information about the high-dimensional measure. In contrast, large deviation principles (LDPs), which characterize the rate of decay of tail probabilities in an asymptotically exact way, are typically non-universal and distribution-dependent, and thus may allow one to distinguish high-dimensional probability measures or convex bodies via their lower-dimensional projections. Moreover, LDPs are also useful for computation of limits of scaled logarithmic volumes (e.g., of Orlicz balls; see also [27] and Remark 3.15) and for the development of computationally efficient (importance sampling) algorithms for numerically estimating such volumes or other tail probabilities for finite n (see, e.g., [40]). Furthermore, they can also be used to obtain information on the conditional distribution of the high-dimensional measure, given that its projections deviate significantly from their means, via the so-called Gibbs conditioning principle (as elucidated in [13] or [14, Section 3.3]); see [33] or the more recent extension in [26], for example, for demonstrations in a geometric context.

The focus of this article is to identify general conditions on sequences of high-dimensional random vectors under which one can characterize asymptotic tail probabilities of their multi-dimensional random projections, and obtain corollaries that may be of interest in convex geometry. More precisely, the goal is to establish LDPs for random projections of general sequences of random vectors, both for projections onto fixed lower-dimensional spaces, as well as onto spaces with dimension growing with n. In the latter case, when the dimension of the projected vectors grows with n, the space in which one should look for an LDP for the sequence of projected vectors is *a priori* unclear. We show that the empirical measure of the coordinates of the projected vector is a convenient object to look at, and allows us to establish LDPs in the space of probability measures on $\mathbb{R}$ in these cases. Unlike in the case of the CLT for convex sets, where a transition occurs at a projection dimension of $k_n = 2\log n/\log\log n$ (or $k_n = n^\alpha$ when restricted to isotropic logconcave measures), these LDPs hold for all growing k_n as long as $k_n/n \to \lambda \in [0,1]$. Additional motivation for studying projections onto growing subspaces arises from the fact that the speeds and rate functions of such LDPs for scaled Euclidean norms of sequences of log-concave isotropic random vectors have implications for the Kannan-Lovász-Simonovits (KLS) conjecture, which is one of the major open problems in convex geometry (see [4, Theorem A] for details of this connection, and also Remark 2.4).

1.2. Contributions and outline of the paper

In this article, we introduce a large deviation analogue of the results in [6]. Whereas the latter work shows that fluctuations of (most) random projections of high-dimensional vectors that satisfy a *thin shell condition* can be characterized (as almost Gaussian), our work characterizes tail behavior (at the level of annealed LDPs) for projections and their associated norms onto (possibly growing) random subspaces of high-dimensional random vectors that satisfy an *asymptotic thin shell condition.* In particular, our work goes beyond the more studied specific setting of measures on ℓ_p^n balls or spheres, and also univariate LDPs, although we also obtain new results in this setting (see Remark 3.6). Specifically, for any sequence of random vectors $\{X^{(n)}\}_{n\in\mathbb{N}}$ whose scaled Euclidean norms satisfy an LDP (see Assumption A and its specific case Assumption A*), we characterize the tail behavior of the corresponding sequence of orthogonal projections of $X^{(n)}$ onto a random k_n-dimensional basis, $k_n \leq n$, drawn from the Haar measure on the Stiefel manifold $\mathbb{V}_{n,k}$ of orthonormal k_n-frames in $\mathbb{R}^n$, as the dimension n goes to infinity with $k_n/n \to \lambda \in [0, 1]$. Assumption A (or rather, a slight strengthening of it) can be viewed as an "asymptotic thin shell" condition since it implies that for all sufficiently large n, the random vector $X^{(n)}$ satisfies the thin shell condition (see the discussion at the end of Section 2.1). Note, however, that for growing subspaces, in contrast to CLT results where approximate Gaussian marginals hold only for $k_n < \frac{2\log n}{\log\log n}$ [43] (or $k_n \sim n^\alpha$ if one assumes additional regularity of $X^{(n)}$ such as logconcavity [35]), the annealed LDP results indicate three crucial regimes for $\{k_n\}_{n\in\mathbb{N}}$, constant, sublinear and linear (see also [3] for ℓ_p^n balls).

A summary of our main results is as follows (the precise definition of LDPs, rate functions, and the Stiefel manifold are given in Section 1.4 and Section 2):

1. *LDPs in the constant regime (Theorem 2.7):* Given Assumption A*, or a modification of it stated as Assumption B, in the setting where $k_n = k$ for every n, we establish an (annealed) LDP for the sequence of k-dimensional random projections of $X^{(n)}$.
2. *LDPs in the sublinear and linear regimes (Theorems 2.9 and 2.15):* Given Assumption A, in the setting where $\{k_n\}_{n\in\mathbb{N}}$ satisfies $k_n \to \infty$, we establish (annealed) LDPs for the sequence of empirical measures of the coordinates of the k_n-dimensional projections of $X^{(n)}$. This LDP is established with respect to the q-Wasserstein topology for $q < 2$, which is stronger than the weak topology. The rate function is shown to have a different form in the sublinear ($k_n/n \to 0$) and linear ($k_n/n \to \lambda \in (0, 1]$) regimes.
3. *LDPs for norms of random projections (Corollary 2.8 and Theorems 2.12, 2.11 and 2.16):* We establish LDPs for sequences of ℓ_q norms of the multi-dimensional random projections in all regimes, with two different scalings considered in the sublinear regime.
4. *Illustrative examples (Section 3):* To show that our theory unites disparate examples under a common framework, recovering, and in some cases extending, existing results

for ℓ_p^n balls, while also covering new examples, we verify our assumptions for many sequences $\{X^{(n)}\}_{n\in\mathbb{N}}$ of interest, including product measures, the uniform measure on certain scaled ℓ_p^n balls and generalized Orlicz balls, and Gibbs measures (see Remark 2.3). Along the way, we obtain several results of potentially independent interest. Specifically, in Theorem 3.5 we show that when $p \in [1,2)$, the LDP for Euclidean norms of multi-dimensional projections of ℓ_p^n balls exhibits an interesting phase transition depending on whether the ratio $k_n/n^{2p/(2+p)}$ is asymptotically finite or infinite, thus disproving a conjecture in [3] (see Remark 3.6 for more details). In addition, we also obtain the asymptotic logarithmic volume of Orlicz balls, as elaborated in Remark 3.15. (See also more recent extensions due to [27] and [39]).

The formulation of the correct form of the asymptotic thin shell condition that would also allow one to consider ℓ_p^n balls, when $p \in [1,2)$, is somewhat subtle, and involves a suitable rescaling argument. Verification of the asymptotic thin shell condition for ℓ_p^n balls, for all $p \geq 1$, makes use of a probabilistic representation for the uniform measure on ℓ_p^n balls (see [50,51] and Section 3.2 for details). However, such probabilistic representations are not available for Orlicz balls, and so we develop a new approach in which the tail probability is expressed as a volume ratio (see (3.18) in the proof of Proposition 3.11), and the asymptotics of this volume ratio is determined using a tilted measure with respect to Lebesgue measure. Subsequent work that uses similar ideas to obtain the volumetric properties and the thin shell concentration for random vectors in Orlicz balls includes [2,27]. More recently, detailed estimates of intersections and differences of Orlicz balls have been obtained in [40]. We leave for future work the identification of more sequences $\{X^{(n)}\}_{n\in\mathbb{N}}$ of random vectors that satisfy the asymptotic thin shell (and related) conditions. Furthermore, although in this article we focus on Euclidean (and more general ℓ_q) norms of the random projections, since they are of special relevance in asymptotic convex geometry, our results could potentially be used to also investigate other symmetric functionals of lower-dimensional projections that are of interest, such as the volume or barycenter of the projected body, or other functionals of interest in high-dimensional statistics.

1.3. Summary of prior work

First steps towards studying LDPs of sequences of one-dimensional projections were taken in [20], [32] and [21], where it was shown that such LDPs capture geometric information about the convex body. However, LDPs for random projections have been first largely restricted to the setting of high-dimensional product measures [20] or the uniform measure on the (suitably renormalized) unit ball or sphere in the space ℓ_p^n for some $p \in [1,\infty)$ [3,4,21,28,29,40], and secondly, limited to univariate LDPs (i.e., in $\mathbb{R}$) involving either the projection of a high-dimensional random vector onto a random one-dimensional subspace [21] or (annealed) LDPs of the Euclidean norm of an orthogonal projection onto a k_n-dimensional subspace, with k_n possibly tending to infinity [3,4].

Indeed, the first paper to consider LDPs for norms of projections of scaled ℓ_p^n balls onto growing subspaces was [3], with further results obtained in subsequent papers (see, e.g., [4,28,29]). (Strictly speaking, in [3] the authors analyze norms of random vectors uniformly distributed on scaled ℓ_p^n balls projected onto k-dimensional subspaces indexed by the Grassmannian, whereas we analyze norms of random projections of more general random n-dimensional vectors (including those uniformly distributed on scaled ℓ_p^n balls) projected onto k-dimensional orthogonal bases indexed by the Stiefel manifold. In the annealed setting considered here, the Grassmanian and Stiefel manifold perspectives can be seen to be equivalent due to the exchangeability of the coordinates of the projections, but in quenched settings, as considered in [38], it is more appropriate to consider the Stiefel manifold.) The only prior example of a multivariate LDP that we know in this context is for the particular case of the sequence of projections of a random vector sampled from a scaled ℓ_p^n ball onto the first k canonical directions [8, Theorem 3.4]. Further, in all cases, the analysis for non-product measures has focused on ℓ_p^n balls, where the analysis is greatly facilitated by a convenient probabilistic representation of the uniform measure on the ℓ_p^n ball (see [51, Lemma 1] or [50]), or a slightly more general class of measures supported on the ℓ_p^n ball that admit a similar probabilistic representation (see [9] or Section 2.2 of [4]). Such a representation has also been exploited in recent work [40] that obtains refined or sharp large deviations estimates for (quenched) random projections of ℓ_p^n balls and spheres.

1.4. Basic definitions and background results

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

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

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

Then, a sequence of probability measure $\{\nu_n\}_{n\in\mathbb{N}} \subset \mathcal{P}_q(\mathbb{R}^d)$ converges to $\nu \in \mathcal{P}_q(\mathbb{R}^d)$ with respect to the *q-Wasserstein topology* if we have weak convergence $\nu_n \Rightarrow \nu$ and convergence of q-th moments $\int_{\mathbb{R}^d} |x|^q \nu_n(dx) \to \int_{\mathbb{R}^d} |x|^q \nu(dx)$. In fact, as elaborated in [53, Sec. 6], the q-Wasserstein topology can be metrized by a distance function called the q-Wasserstein metric, which we denote $\mathcal{W}_q$. Next, for $q > 0$, let

$$M_q(\nu) := \int_{\mathbb{R}^d} |x|^q \nu(dx), \quad \nu \in \mathcal{P}(\mathbb{R}^d), \tag{1.1}$$

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

$$K_{2,j} := \left\{ \nu \in \mathcal{P}(\mathbb{R}^d) : M_2(\nu) \le j \right\}. \tag{1.2}$$

Lemma 1.1. *Fix $j \in \mathbb{N}$. For any $q \in [1, 2)$, the set $K_{2,j} \subset \mathcal{P}_2(\mathbb{R}^d)$ is compact with respect to the q-Wasserstein topology. In addition, $K_{2,j}$ is convex and non-empty.*

Proof. The proof is a simple modification of the proof of the $j = 1$ case given in [33, Lemma 3]. □

We refer to [14] for general background on large deviations theory. In particular, we recall the definition:

Definition 1.2. Let $\mathcal{X}$ be a regular topological space with Borel σ-algebra $\mathcal{B}$. A sequence $\{P_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(\mathcal{X})$ is said to satisfy a *large deviation principle (LDP)* at *speed* s_n with rate function $I : \mathcal{X} \to [0, \infty]$ if I is lower-semicontinuous and for all $\Gamma \in \mathcal{B}$,

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

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

Remark 1.3. Let $\{P_n\}_{n\in\mathbb{N}}$ be a sequence of probability measures satisfying an LDP at speed s_n with GRF I that has a unique minimizer m. It is immediate from the definition of the LDP that $I(m) = 0$ and that then, for any speed t_n such that $t_n/s_n \to 0$, the sequence $\{P_n\}_{n\in\mathbb{N}}$ also satisfies an LDP at speed t_n, but with a degenerate rate function χ_m, which is defined to be 0 at m and $+\infty$ elsewhere.

Remark 1.4. Let $\{P_n\}_{n\in\mathbb{N}}$ be a sequence of probability measures satisfying an LDP at speed s_n with GRF I and suppose b_n is another sequence such that $s_n/b_n \to \lambda \in (0, \infty)$. Then it is immediate from the definition that $\{P_n\}_{n\in\mathbb{N}}$ also satisfies an LDP at speed b_n with GRF λI.

As a useful tool, we recall the following definition.

Definition 1.5. Let $\mathcal{X}$ be a metric space with distance d, equipped with its Borel σ-algebra. Two sequences of $\mathcal{X}$-valued random variables $\{\eta_n\}_{n\in\mathbb{N}}$ and $\{\tilde{\eta}_n\}_{n\in\mathbb{N}}$ are *exponentially equivalent* at speed s_n if for all $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{s_n} \log \mathbb{P}(d(\eta_n, \tilde{\eta}_n) > \delta) = -\infty. \tag{1.3}$$

Remark 1.6. The notion of exponential equivalence is valuable because if an LDP holds for $\{\eta_n\}_{n\in\mathbb{N}}$, then an LDP holds for an exponentially equivalent sequence $\{\tilde{\eta}_n\}_{n\in\mathbb{N}}$ with the same GRF (see, e.g., Theorem 4.2.13 of [14]).

Remark 1.7. Let $\{a_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ that converges to $a \in \mathbb{R}$ as $n \to \infty$. Suppose $\{U_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ at speed b_n with a GRF I. Then the sequences $\{a_nU_n\}_{n\in\mathbb{N}}$ and $\{aU_n\}_{n\in\mathbb{N}}$ are exponentially equivalent at speed b_n. Indeed, for any $M \in (0, \infty)$, since $\{U_n\}_{n\in\mathbb{N}}$ satisfies an LDP with a GRF, there exists $K \in (0, \infty)$ such that $\limsup_{n\to\infty} \frac{1}{b_n} \log \mathbb{P}(|U_n| \geq K) < -M$. Given $\delta, M \in (0, \infty)$, pick $\varepsilon > 0$ such that $\delta/\varepsilon > K$. Then

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

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

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

$$h(\nu) := -\int_{\mathbb{R}} \log\left(\tfrac{d\nu}{dx}\right) d\nu, \tag{1.4}$$

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

$$H(\nu|\mu) := \int_{\mathbb{R}} \log\left(\frac{d\nu}{d\mu}\right) d\nu \tag{1.5}$$

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

$$\Lambda^*(x) = \sup_{t\in\mathbb{R}}\{xt - \Lambda(t)\}, \quad x \in \mathbb{R}. \tag{1.6}$$

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

Lemma 1.8 *(Contraction principle). [14, Theorem 4.2.1] Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces and $f : \mathcal{X} \to \mathcal{Y}$ a continuous function. Suppose $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of $\mathcal{X}$-valued random variables and satisfies an LDP in $\mathcal{X}$ at speed s_n with GRF $\mathcal{I}_X$. Then $\{f(X_n)\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{Y}$ at speed s_n with GRF $\mathcal{I}_Y$ defined by*

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

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

Lemma 1.9. *Suppose $\{U_n\}_{n\in\mathbb{N}}$, $\{V_n\}_{n\in\mathbb{N}}$ and $\{W_n\}_{n\in\mathbb{N}}$ satisfy LDPs in $\mathbb{R}$ at speed α_n, β_n and γ_n with GRFs J_U, J_V and J_W, respectively. Let $\alpha_n = \beta_n \ll \gamma_n$, (i.e., $\beta_n/\gamma_n \to 0$ as $n \to \infty$). Assume $\{U_n\}_{n\in\mathbb{N}}$ is independent of $\{V_n\}_{n\in\mathbb{N}}$ and J_W has a unique minimizer m. Then $\{(U_n, V_n, W_n)\}_{n\in\mathbb{N}}$ is exponentially equivalent to $\{(U_n, V_n, m)\}_{n\in\mathbb{N}}$ and satisfies an LDP at speed α_n with GRF $J : \mathbb{R}^3 \to [0, \infty]$ defined by*

$$J(u, v, w) := \begin{cases} J_U(u) + J_V(v), & w = m, \\ +\infty, & \textit{otherwise}. \end{cases} \tag{1.7}$$

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

Proof. Define $Y_n := (U_n, V_n, W_n)$ and $\tilde{Y}_n := (U_n, V_n, m)$. By the independence of $\{U_n\}_{n\in\mathbb{N}}$ and $\{V_n\}_{n\in\mathbb{N}}$, $\{\tilde{Y}_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed α_n with GRF J defined in (1.7) (this is easily deduced from the definition of the LDP, but the fact that the rate function for (U_n, V_n) is $J_U(u) + J_V(v)$ follows from [14, Exercise 4.2.7] with $\mathcal{X} = \mathbb{R}$, $\mathcal{Y} = \mathbb{R}^2$ and F being the identity mapping; the stated rate function for (U_n, V_n, m) follows as an immediate consequence). Now, for $\epsilon > 0$, we have

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

where the last equality follows because J_W has a unique minimizer at m and $\{W_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $\gamma_n \gg \alpha_n$, as in the observation of Remark 1.3. Thus, $\{Y_n\}_{n\in\mathbb{N}}$ is exponentially equivalent to $\{\tilde{Y}_n\}_{n\in\mathbb{N}}$ at speed α_n. By Remark 1.6, this implies $\{Y_n\}_{n\in\mathbb{N}}$ satisfies an LDP with speed α_n with the same GRF as $\{\tilde{Y}_n\}_{n\in\mathbb{N}}$. This proves the first assertion of the lemma.

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

2. Main results

2.1. Projection regimes and assumptions

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

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

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

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

Definition 2.1. Given a sequence $\{k_n\}_{n\in\mathbb{N}} \subset \mathbb{N}$, we say:

1. $\{k_n\}_{n\in\mathbb{N}}$ is *constant* at k, for some $k \in \mathbb{N}$, denoted $k_n \equiv k$, if $k_n = k$ for all $n \in \mathbb{N}$;
2. $\{k_n\}_{n\in\mathbb{N}}$ *grows sublinearly*, denoted $1 \ll k_n \ll n$, if $k_n \to \infty$ but $k_n/n \to 0$;
3. $\{k_n\}_{n\in\mathbb{N}}$ *grows linearly* with rate λ, for some $\lambda \in (0,1]$, denoted $k_n \sim \lambda n$, if $k_n/n \to \lambda$.

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

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

Remark 2.2. Note that the law of $\mathbf{A}_{n,k_n}$ is invariant under permutation of its k_n columns, so the k_n coordinates of $\mathbf{A}_{n,k_n}^T X^{(n)}$ are exchangeable. Thus, the empirical measure L^n in (2.1) encodes the essential distributional properties of the coordinates of the projection, and hence, serves as a natural choice for a common infinite-dimensional embedding of the k_n coordinates of $\mathbf{A}_{n,k_n}^T X^{(n)}$, for all $n \in \mathbb{N}$.

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

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

When a special case of Assumption A holds with speed $s_n = n$, we say that Assumption A* holds.

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

Remark 2.3. The special case of Assumption A* is important because, as shown in Section 3, it is satisfied by several important sequences of measures, including those whose elements are taken from a large family of product measures (see Proposition 3.1), the uniform measure on an ℓ_p^n ball of radius $n^{1/p}$, with $p \geq 2$ (see Proposition 3.7) or, in fact, a more general class of measures that includes the uniform measure on an Orlicz ball defined via a superquadratic function (see Proposition 3.11), and a general class of Gibbs measures with superquadratic potential and interaction functions (see Proposition 3.18). However, Assumption A* no longer holds when $X^{(n)}$ is uniformly chosen from an ℓ_p^n ball of radius $n^{1/p}$ with $p \in [1, 2)$. Indeed, this can be deduced from the sharp large deviation upper bounds obtained in [47] and [22]. In addition, Theorem 1.3 of [28] shows that $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP with speed $s_n = n^{p/2}$. These results motivate the more general condition stated in Assumption A.

Remark 2.4. The general form of Assumption A is also of interest because of its connection to the Kannan-Lovász-Simonovits (KLS) conjecture formulated in [31] (also see [1] for a nice exposition), which is one of the major open problems in convex geometry. Indeed, Theorem A in [4] states that the KLS conjecture is false if there exists a sequence of isotropic and log-concave random vectors $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfying Assumption A either with $s_n \ll \sqrt{n}$ and nontrivial J_X or with $s_n = \sqrt{n}$ and $\inf_{t>t_0}\{\inf_{x>t} J_X(x)/t\} = 0$ for some $t_0 \in (1, \infty)$. It is shown in [4] that when $X^{(n)}$ is uniformly distributed on the ℓ_1^n ball of radius n, then $\{X^{(n)}\}$ satisfies Assumption A with $s_n = \sqrt{n}$, but the condition on the rate function J_X is not satisfied (which is consistent with the fact that the KLS conjecture is widely believed to be true). In view of this observation, it would be interesting to extend our verification of Assumption A to more general measures such as Orlicz balls (respectively, Gibbs measures) with superlinear functions (respectively, potentials), extending the corresponding results obtained in this article for superquadratic functions (respectively, potentials).

When Assumption A is satisfied with a GRF J_X that has a unique minimum at $m \in \mathbb{R}_+$, go then for all sufficiently large n, the random variable $X^{(n)}$ satisfies the thin shell condition of [6, Equation (1)]. To be more precise, note that the observation that the minimum of J_X, which is equal to zero, is achieved uniquely at m, together with the fact that $s_\ell \to \infty$, implies that for every $c > 0$, there exists $\{\delta_\ell\}_{\ell\in\mathbb{N}}$ with $\delta_\ell \downarrow 0$ such that

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

Setting $\varepsilon_\ell := \max(\delta_\ell, 2e^{-c\sqrt{s_\ell}})$, it then follows from Assumption A and the definition of the LDP (see Definition 1.2) that $\varepsilon_\ell \downarrow 0$ as $\ell \to \infty$, and for every $\ell \in \mathbb{N}$, there exists $N_\ell \in \mathbb{N}$ such that

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

In particular, this implies the following weak limit:

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

Thus, we refer to the strengthening of Assumption A with J_X having a unique minimum as the *asymptotic thin shell condition*. Note that the thin shell condition is usually stated for isotropic random vectors $X^{(n)}$ and with $m = 1$, but since we do not restrict our consideration to only isostropic random vectors $X^{(n)}$, we phrased the condition above for arbitrary $m > 0$. Just as the thin shell condition yields a central limit theorem in the sense that the projections of $X^{(n)}$ can be shown to be close to Gaussian (see, e.g., [6,35]), our results, summarized in the next three sections, show that the asymptotic thin shell condition implies that the empirical measures $\{L^n\}_{n\in\mathbb{N}}$ of the coordinates of the projections of $X^{(n)}$ satisfy an LDP in the sublinear and linear regimes (with the weaker Assumption A sufficing in the latter regime).

Remark 2.5. There exist sequences $\{X^{(n)}\}_{n\in\mathbb{N}}$ for which the corresponding random projections satisfy an LDP, but the thin shell condition fails to hold. For example, this can happen when Assumption A holds with a rate function J_X that has multiple minima, as can happen for certain Gibbs measures (see Section 3.4) with non-convex potentials F and G. As another example, let $X^{(n)}$ be distributed according to a mixture of two Gaussian distributions in $\mathbb{R}^n$ both with mean 0 and covariance matrices I_n and $2I_n$, respectively. In other words, the density of $X^{(n)}$ takes the form $f_{X^{(n)}} = \frac{1}{2}\phi_{I^n} + \frac{1}{2}\phi_{2I^n}$ for a positive definite matrix C, where ϕ_C denotes the Gaussian density with mean 0 and covariance matrix C. Then $\{X^{(n)}\}_{n\in\mathbb{N}}$ does not satisfy the thin shell condition since half its mass is concentrated around the thin shell of radius $\sqrt{n}$ and the other half around the thin shell of radius $\sqrt{2n}$. However, it is easy to show that the sequence $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption A* (e.g., since the sequence of norms of each of the Gaussian distributions in the mixture satisfies an LDP trivially, the LDP for the sequence of norms of the mixtures can be deduced from results in [16]).

2.2. Results in the constant regime

To establish LDPs in the constant regime, when $\{k_n\}_{n\in\mathbb{N}}$ is constant at k for some $k \in \mathbb{N}$, we will require either Assumption A* or, to cover more general sequences $\{X^{(n)}\}_{n\in\mathbb{N}}$ like the uniform measure on an ℓ_p^n ball with $p \in [1, 2)$, the following modification of Assumption A*.

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 $J_X : \mathbb{R} \to [0, \infty]$.

Remark 2.6. It is easy to see, by a simple rescaling argument, that Assumption A* is equivalent to a modified version of Assumption B, in which one requires that $\{\sqrt{s_n}\|X^{(n)}\|_2/n)\}_{n\in\mathbb{N}}$ satisfies an LDP at a speed s_n that satisfies $s_n/n \to r$, with $r \in (0, \infty)$ rather than $r = 0$. Indeed, this modified version of Assumption B would hold with GRF $J_X^{(r)}$ if and only if Assumption A* is satisfied with GRF $J_X(x) := rJ_X^{(r)}(\sqrt{r}x)$.

Theorem 2.7 *(constant, $k_n \equiv k$). 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 $\{n^{-1/2}\mathbf{A}_{n,k}^T X^{(n)}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^k$ at speed s_n, with GRF $I_{\mathbf{A}X,k} : \mathbb{R}^k \to [0, \infty]$ defined by*

$$I_{\mathbf{A}X,k}(x) := \begin{cases} \inf_{0<c<1} \left\{ J_X\left(\frac{\|x\|_2}{c}\right) - \frac{1}{2}\log\left(1 - c^2\right) \right\}, & \text{if Assumption A* holds,} \\ \inf_{c>0} \left\{ J_X\left(\frac{\|x\|_2}{c}\right) + \frac{c^2}{2} \right\}, & \text{if Assumption B holds.} \end{cases} \tag{2.3}$$

The proof of Theorem 2.7 is given in Section 4.2, where the role of Assumption B in the proof is discussed in detail. The rate function takes the form in (2.3) because rare events for $n^{-1/2}\mathbf{A}_{n,k}^T X^{(n)}$ occur through a combination of the "radial" component represented by J_X, and the "angular" component represented by $c \mapsto -\frac{1}{2}\log(1 - c^2)$ (when Assumption A* holds) or $c \mapsto c^2/2$ (when Assumption B holds).

As an immediate corollary of Theorem 2.7, we have the following LDP for the corresponding scaled ℓ_q^k norms (or in the case $q \in (0,1)$, quasi-norms) of random projections. For any $q \in (0,\infty)$ and $n \in \mathbb{N}$, define

$$Y_{q,k}^n := \|\mathbf{A}_{n,k}^T X^{(n)}\|_q. \tag{2.4}$$

Corollary 2.8 *(constant, $k_n \equiv k$). Suppose $\{k_n\}_{n\in\mathbb{N}}$, $\{X^{(n)}\}_{n\in\mathbb{N}}$, $\{s_n\}_{n\in\mathbb{N}}$ and $I_{\mathbf{A}X,k}$ are as in Theorem 2.7. Then $\{n^{-1/2}Y_{q,k}^n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF*

$$\mathbb{J}_q^{\mathsf{con}}(x) := \inf_{z\in\mathbb{R}^k} \{I_{\mathbf{A}X,k}(z) : \|z\|_q = x\}, \quad x \in \mathbb{R}_+. \tag{2.5}$$

Proof. This is an immediate consequence of the LDP for $\{n^{-1/2}\mathbf{A}_{n,k}^T X^{(n)}\}_{n\in\mathbb{N}}$ in Theorem 2.7 and the contraction principle (Lemma 1.8) applied to the continuous mapping $\mathbb{R}^k \ni x \mapsto \|x\|_q \in \mathbb{R}$. □

2.3. Results in the sublinear regime

Recall that if, instead of being constant, the sequence k_n tends to infinity as $n \to \infty$, then our goal is to establish an LDP for the sequence of empirical measures $\{L^n\}_{n\in\mathbb{N}}$ of (2.1). We start in this section by analyzing the sublinear regime. Section 2.3.1 summarizes our LDP results for the sequences of empirical measures $\{L^n\}_{n\in\mathbb{N}}$ and Euclidean norms of the randomly projected vectors. Section 2.3.2 contains additional results on LDPs for a different scaling of Euclidean norms as well as more general q-norms, with $q \in [1,2)$, of projected vectors that is suitable for the sublinear regime.

2.3.1. LDPs for the empirical measures and norms of projected vectors

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

$$\mathcal{P}(\mathbb{R}) \ni \gamma_\sigma \sim \mathcal{N}(0,\sigma^2). \tag{2.6}$$

Theorem 2.9 *(sublinear, $1 \ll k_n \ll n$). Suppose $\{k_n\}_{n\in\mathbb{N}}$ grows sublinearly and Assumption 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 defined in (1.5). Then, for every $q \in [1,2)$,*

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

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

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

$$\mathbb{I}_{L,k_n}(\mu) := \inf_{c>0} \{H(\mu|\gamma_c) + J_X(c)\}.$$

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

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

As in Theorem 2.7, the rate functions above can again be decomposed into a radial component, represented by J_X (as a consequence of Assumption A), and "an angular" component $\mathbf{A}_{n,k_n}$, which is captured by the relative entropy term. Depending on the relative rate of growth of k_n and s_n, different parts dominate the rate function, with both terms being present only when $s_n = k_n$.

Next, as in the constant regime (see Corollary 2.8), we also establish LDPs for the sequence of scaled Euclidean norms of the random projections. To deal with cases when Assumption A* is not satisfied, it is not sufficient to consider Assumption B as in the constant regime. Instead, we will need to introduce the following refinement of Assumption A.

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 $J_X^{(\infty)}$.

The following simple observation is analogous to the one made in Remark 2.6.

Remark 2.10. It is easy to see that Assumption C holds with $r \in (0,\infty)$, $\{s_n\}_{n\in\mathbb{N}}$ and GRF $J_X^{(r)}$ if and only if it also holds with $r' = 1$, $\{s'_n := k_n\}_{n\in\mathbb{N}}$, and GRF $J_X^{(1)}(x) := rJ_X^{(r)}(\sqrt{r}x)$, $x \in [0,\infty)$. Therefore, in essence, one need only consider the cases $r \in \{0,1,\infty\}$ in Assumption C.

Theorem 2.11. *Suppose k_n grows sublinearly, and recall the definition of $Y^n_{q,k}$ given in (2.4).*

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

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

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

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

Note that there is a transition in the form of the LDP depending on the relative growth rates of $\{s_n\}_{n\in\mathbb{N}}$ and $\{k_n\}_{n\in\mathbb{N}}$. The implications of these results for the special case when $X^{(n)}$ is the uniform measure on an ℓ_p^n ball, and their relation to the work of [3], is discussed in Section 3.2.2; see Theorem 3.5 and Remark 3.6.

2.3.2. LDP for an alternative scaling of q-norms of projections

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

$$\Lambda_q(t) := \log \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(t\,|x|^q - \frac{1}{2}x^2\right) dx, \quad t \in \mathbb{R}, \tag{2.7}$$

and let Λ_q^* be the Legendre transform of Λ_q. Moreover, set $\mathcal{M}_q$ to be the q-th absolute moment of a standard Gaussian random variable,

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

Theorem 2.12. *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 $m > 0$. Also, recall the definition of $Y^n_{q,k}$ given in (2.4). Then $\{k_n^{-1/q}Y^n_{q,k_n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $s_n \wedge k_n$ with GRF $\widehat{\mathbb{J}}_q^{\mathsf{sub}} : \mathbb{R} \to [0,\infty]$ defined by*

$$\widehat{\mathbb{J}}_q^{\mathsf{sub}}(x) := \begin{cases} \Lambda_q^*\left(x^q/m^q\right), & \text{if } s_n \gg k_n,\\ \inf_{c>0}\left\{\Lambda_q^*(c^q) + J_X\left(x/c\right)\right\}, & \text{if } s_n = k_n,\\ J_X(x/\mathcal{M}_q^{1/q}), & \text{if } s_n \ll k_n, \end{cases} \tag{2.9}$$

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

The proof of Theorem 2.12 is given in Section 5. When $q \in [1,2)$, the LDP for $\{k_n^{-1/q}Y^n_{q,k_n}\}_{n\in\mathbb{N}}$ is an immediate consequence of Theorem 2.9 and the contraction principle. For $q = 2$, the contraction principle no longer applies since the moment map $M_2(\cdot)$ is not continuous in $\mathcal{P}_q(\mathbb{R})$ for $q < 2$. We take a different approach to the proof for all cases $q \in [1, 2]$, by looking directly at the norm, instead of using the LDP of empirical measures. In the special case when $q = 2$ and $X^{(n)}$ is uniformly distributed on the scaled ℓ_p^n ball of radius $n^{1/p}$, with $p \in [2, \infty]$ (or admits a slightly more general representation), the LDP for $\{\widetilde{Y}^n_{2,k_n}\}_{n\in\mathbb{N}}$ was also obtained in [4, Theorem B].

2.4. Results in the linear regime

Recall the definition of the second moment map $M_2(\cdot)$ from (1.1).

Definition 2.13. For $\lambda \in (0, 1]$, define $\mathcal{H}_\lambda : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ as

$$\mathcal{H}_\lambda(\nu) := \begin{cases} -\lambda\, h(\nu) + \frac{\lambda}{2}\log(2\pi e) + \frac{1-\lambda}{2}\log\left(\frac{1-\lambda}{1-\lambda\, M_2(\nu)}\right), & \text{if } M_2(\nu) \le 1/\lambda, \\ +\infty, & \text{otherwise,} \end{cases} \tag{2.10}$$

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

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

Remark 2.14. Since h is strictly concave and M_2 is linear, from the definition in (2.10), $\mathcal{H}_\lambda$ is strictly convex. A direct verification shows that $\mathcal{H}_\lambda(\gamma_1) = 0$, and the strict convexity of $\mathcal{H}_\lambda$ shows that γ_1 is the unique minimizer of $\mathcal{H}_\lambda$.

Theorem 2.15 *(linear, $k_n \sim \lambda n$). Fix $q \in [1, 2)$. Suppose $\{k_n\}_{n\in\mathbb{N}}$ grows linearly with rate $\lambda \in (0, 1]$ and Assumption 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 $\mathcal{P}_q(\mathbb{R})$ at speed s_n with GRF $\mathbb{I}_{L,\lambda} : \mathcal{P}_q(\mathbb{R}) \to [0, \infty]$, where*

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

$$\begin{aligned} \mathbb{I}_{L,\lambda}(\mu) = \inf_{c > \sqrt{\lambda M_2(\mu)}} &\left\{ J_X(c) - \tfrac{1-\lambda}{2}\log\left(1 - \tfrac{\lambda M_2(\mu)}{c^2}\right) + \lambda \log(c) \right\} \\ &- \lambda h(\mu) + \tfrac{\lambda}{2}\log(2\pi e) + \tfrac{1-\lambda}{2}\log(1-\lambda), \end{aligned} \tag{2.11}$$

   where we use the convention that $0 \log 0 = 0$.
2. *If $s_n \ll n$, then for $\mu \in \mathcal{P}_q(\mathbb{R})$*

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

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

Theorem 2.16. *Suppose $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption A, and $k_n \sim \lambda n$ for some $\lambda \in (0,1]$. Then, for $q \in [1,2]$ the sequence $\{n^{-1/q}Y^n_{q,k_n}\}_{n\in\mathbb{N}}$ defined in (2.4) satisfies an LDP at speed s_n with GRF $\mathbb{J}^{\mathsf{lin}}_{q,\lambda}$, where for $x \in \mathbb{R}_+$,*

$$\mathbb{J}^{\mathsf{lin}}_{q,\lambda}(x) := \begin{cases} \inf_{\nu\in\mathcal{P}(\mathbb{R}),c\in\mathbb{R}_+} \left\{\mathcal{H}_\lambda(\nu) + J_X(c) : \lambda M_q(\nu) = \left(\frac{x}{c}\right)^q\right\}, & \text{if } s_n = n, \\ J_X\left(\frac{x}{(\lambda\mathcal{M}_q)^{1/q}}\right), & \text{if } s_n \ll n, \end{cases} \tag{2.12}$$

with M_q the q-th moment map as in (1.1) and $\mathcal{M}_q$ the q-th absolute moment of a standard Gaussian random variable defined in (2.8).

The proof of Theorem 2.16 is deferred to Section 6.2. It is shown there that when $q \in [1,2)$, the result is an immediate consequence of Theorem 2.15 and an application of the contraction principle. However, even though the rate function still takes an analogous form when $q = 2$, the contraction principle is no longer applicable in that setting because the LDP of Theorem 2.15 only holds in the q-Wasserstein topology for $q \in [1,2)$. Despite this apparent gap, using a different argument in Section 6.2 we show that the result nevertheless holds. In the process, we provide an alternative representation for the rate function (2.12) for all $q \in [1,2]$ (see Proposition 6.1). This is a manifestation of a somewhat nuanced technical issue, which is elaborated upon in Remark 6.4.

3. Examples satisfying the main assumptions

In this section, we present several examples of sequences of random vectors $\{X^{(n)}\}_{n\in\mathbb{N}}$ that satisfy the assumptions introduced in Section 2.

3.1. Product measures

Lemma 3.1 *(i.i.d. case). Let $X_1, X_2, \ldots$ be a sequence of i.i.d. real-valued random variables, and let $X^{(n)} := (X_1, \ldots, X_n)$. Suppose that we have*

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

for t in some open ball of non-zero radius about 0, and let Λ^* be the Legendre transform of Λ defined in (1.6). Then, $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption A with $J_X(x) := \Lambda^*(x^2)$ for $x \in \mathbb{R}_+$ and $J_X(x) := +\infty$ otherwise. Moreover, $m = \sqrt{\mathbb{E}[|X_1|^2]}$ is the unique minimizer of J_X.*

Proof. By Cramér's theorem for sums of i.i.d. random variables [14, Theorem 2.2.1], the sequence $\{\|X^{(n)}\|_2^2/n\}_{n\in\mathbb{N}}$, satisfies an LDP at speed n with GRF Λ^*. By the contraction principle applied to the square root function, Assumption A* holds. As for the unique minimizer, this follows from the law of large numbers, with limit $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{\text{a.s.}} \mathbb{E}[X_1^2] = m^2$. □

3.2. Scaled ℓ_p^n balls

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

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

We introduce some preliminaries in Section 3.2.1, then verify our main assumptions for $\{X^{(n,p)}\}_{n\in\mathbb{N}}$ in the case when $p \in [1, 2)$ in Section 3.2.2 and the easier case when $p > 2$ in Section 3.2.3. When combined with the results of Section 2, these yield LDPs for random projections of these ℓ_p^n balls.

3.2.1. Preliminaries

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

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

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

$$2\exp\left(-\frac{1}{p}t^{p/2} - \frac{p-1}{2}\log t\right) \geq \mathbb{P}(|\xi^{(p)}|^2 > t) \geq \frac{t^{p/2}}{t^{p/2}+1}\exp\left(-\frac{1}{p}t^{p/2} - \frac{p-1}{2}\log t\right). \tag{3.1}$$

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

$$F_p^*(y) := \sup_{t_1,t_2\in\mathbb{R}} \left[t_1 y + t_2 - \log\left(\int_{\mathbb{R}} e^{t_1 x^2 + t_2|x|^p} f_p(x)dx \right) \right], \quad y \in \mathbb{R}, \tag{3.2}$$

and

$$m(p) := \left(\frac{p^{2/p}\Gamma\left(1 + \frac{3}{p}\right)}{3\Gamma\left(1 + \frac{1}{p}\right)} \right)^{1/2}. \tag{3.3}$$

Also, let U be a uniform random variable on $[0, 1]$ and $\{\xi_i^{(p)}\}_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables with density f_p independent of U. For $n \in \mathbb{N}$ and $p \in [1, \infty)$, denote $\xi^{(n,p)} = (\xi_1^{(p)}, \ldots, \xi_n^{(p)})$. In this section, we take advantage of the following useful representation of $X^{(n,p)}$ obtained in [51, Lemma 1]:

$$X^{(n,p)} \stackrel{(d)}{=} n^{1/p} U^{1/n} \frac{\xi^{(n,p)}}{\left\|\xi^{(n,p)}\right\|_p}. \tag{3.4}$$

In view of the representation (3.4), we first establish a property of p-generalized normal random variables which we strengthen from the result in [46].

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

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

The proof is deferred to Appendix B.

3.2.2. Verification of Assumptions B and C when $p \in [1,2)$

When $p \in [1,2)$, it should be noted that annealed LDPs associated with the random vectors $\{X^{(n,p)}\}_{n\in\mathbb{N}}$ defined in Proposition 3.7 occur at different speeds than in the case $p > 2$; see Theorem 2.3 of [21]. In particular, from Theorem 1.3 of [28] it follows that in this case the sequence of scaled norms $\{\|X^{(n,p)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $n^{p/2}$. Thus, in this case $\{X^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies Assumption A with $s_n = n^{p/2}$ and unique minimizer $m = m(p)$ defined in (3.3). In particular, Assumption A* is not satisfied. We show that nevertheless, Assumption B does hold. In what follows, let $U, \{\xi_i^{(p)}\}_{i\in\mathbb{N}}$, and $\xi^{(n,p)}$ be as in the representation of $X^{(n,p)}$ in (3.4), and note that then

$$\frac{\sqrt{s_n}\left\|X^{(n,p)}\right\|_2}{n} \stackrel{(d)}{=} \frac{U^{1/n}}{\left\|\xi^{(n,p)}\right\|_p/n^{1/p}} \frac{\sqrt{s_n}\left\|\xi^{(n,p)}\right\|_2}{n}, \quad n \in \mathbb{N}. \tag{3.5}$$

Proposition 3.3. *For $p \in [1,2)$, $\{X^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies Assumption B with speed $s_n := n^{2p/(2+p)}$ and GRF*

$$J_{X,p}(x) := \begin{cases} \dfrac{x^p}{p}, & x \geq 0, \\ +\infty, & x < 0. \end{cases} \tag{3.6}$$

Proof. Fix $p \in [1,2)$. It follows from [21, Lemma 3.3] that $\{U^{1/n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n, and from Cramér's theorem and the contraction principle that $\{\|\xi^{(n,p)}\|_p/n^{1/p}\}_{n\in\mathbb{N}}$ also satisfies an LDP at speed n. Moreover, it is easy to see that both sequences converge almost surely to 1. Since, in addition, $\xi^{(n,p)}$ and U are independent, appealing again to the contraction principle it follows that their ratio $W_n := U^{1/n}n^{1/p}/\|\xi^{(n,p)}\|_p$ also satisfies an LDP at speed n. Furthermore, since $W_n \to 1$ almost surely as $n \to \infty$, the LDP rate function for $\{W_n\}_{n\in\mathbb{N}}$ has the unique minimizer 1. Given $s_n \ll n$ and the representation in (3.5), an application of (the last assertion of) Lemma 1.9, with W_n as above, $m = 1$, $\gamma_n = n$, $V_n := \sqrt{s_n}\|\xi^{(n,p)}\|_2/n$, and $\beta_n = s_n$ shows that to establish the proposition it suffices to prove that $\{V_n\}_{n\in\mathbb{N}}$ satisfies an LDP with speed s_n and GRF $J_{X,p}$. To this end, using the relation $\frac{\sqrt{s_n}\left\|\xi^{(n,p)}\right\|_2}{n} = \left(\frac{s_n}{n^2}\sum_{i=1}^n(\xi_i^{(p)})^2\right)^{1/2}$, the contraction principle, Lemma 3.2 with $k_n = n$ and $b_n = n^2/s_n$ therein, and the property that $s_n \ll n$, we can conclude that $\{\sqrt{s_n}\|\xi^{(n,p)}\|_2/n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $n^p s_n^{-p/2}$ with GRF $J_{X,p}$ given in (3.6). Finally, since $s_n = n^{2p/(2+p)}$, we obtain $n^p s_n^{-p/2} = s_n$. $\square$

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

Proposition 3.4. *When $p \in [1, 2)$, let $J_{X,p}$ be defined as in (3.6). Then $\{X^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies Assumption C with GRF $J_X^{(r)} = J_{X,p}$ for all $r \in [0,\infty]$, where s_n is defined by*

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

Proof. Let s_n be as defined in the proposition. As in Assumption C, let $r \in [0,\infty]$ be the limit of s_n/k_n as $n \to \infty$. It is easy to see that $r = 0$ if $k_n \gg n^{2p/(2+p)}$, $r = 1$ if $k_n = n^{2p/(2+p)}$, and $r = \infty$ when $k_n \ll n^{2p/(2+p)}$. Now, set $c_n := \sqrt{n/k_n}$ if $r \in [0,\infty)$ and $c_n := \sqrt{n/s_n}$ if $r = \infty$ and note that $c_n \to \infty$ as $n \to \infty$. By (3.4), we have the following representation

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

We first claim that $\{\|\xi^{(n,p)}\|_2/(c_n\sqrt{n})\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $s_n \ll n$. Indeed, by Lemma 3.2 with $k_n = n$ and $b_n = c_n^2 n$ there, the fact that $b_n \to \infty$ as $n \to \infty$ and the contraction principle with the mapping $x \mapsto \sqrt{x}$, $\left\{\|\xi^{(n,p)}\|_2/(c_n\sqrt{n})\right\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $n^{p/2}c_n^p$ with GRF $J_{X,p}$ given in (3.6). By [21, Lemma 3.3], Cramér's theorem and the contraction principle, the independence of $\xi^{(n,p)}$ and U, and the contraction principle, we see that $W_n := U^{1/n}n^{1/p}/\|\xi^{(n,p)}\|_p$ satisfies an LDP at speed n and further, it also converges almost surely to 1. The lemma then follows upon applying Lemma 1.9 with $V_n := \|\xi^{(n,p)}\|_2/(c_n\sqrt{n})$ and $W_n := U^{1/n}n^{1/p}/\|\xi^{(n,p)}\|_p$, $m = 1$, $\gamma_n = n$ and $\beta_n := s_n \ll n$. $\square$

When combined with the results of Section 2.3.1, the last two propositions imply the following LDPs for norms of projections in the constant and sublinear regimes. Here, we focus on the norm $Y_{2,k_n}^{(n,p)}$, which is defined as in (2.4) but with $X^{(n)}$ replaced with

$X^{(n,p)}$. The corresponding results for the differently scaled norms $\{k_n^{-1/2} Y_{2,k_n}^{(n,p)}\}_{n\in\mathbb{N}}$ were obtained in Theorem B of [4].

Theorem 3.5. *Fix $p \in [1, 2)$.*

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

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

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

$$\mathbb{J}_{Y_{2,k}^{(p)}}(x) := \begin{cases} \frac{p+2}{2p} x^{2p/(p+2)}, & x \geq 0, \\ +\infty, & x < 0. \end{cases} \tag{3.7}$$

2. *Suppose k_n grows sublinearly.*
   (a) *If $k_n \ll n^{2p/(2+p)}$. Then $\{n^{-1/2} Y_{2,k}^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ at speed $n^{2p/(2+p)}$ with GRF $\mathbb{J}_{Y_{2,k_n}^{(p)}} : \mathbb{R} \to [0,\infty]$, which is equal to $\mathbb{J}_{Y_{2,k}^{(p)}}$ defined in (3.7).*
   (b) *If $k_n = n^{2p/(2+p)}$. Then $\{n^{-1/2} Y_{2,k}^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ at speed $n^{2p/(2+p)}$ with GRF $\mathbb{J}_{Y_{2,k_n}^{(p)}} : \mathbb{R} \to [0,\infty]$ defined by*

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

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

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

   which is equal to $J_{X,p}$ in (3.6) of Proposition 3.3.

Proof. In the constant regime, Proposition 3.3 shows that $\{X^{(n,p)}\}$ satisfies Assumption B with rate function $J_{X,p}$ taking the explicit form given in (3.6). Substituting this into Theorem 2.7, we see that $\{n^{-1/2}\mathbf{A}_{n,k}^T X^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^k$ at speed $s_n = n^{2p/(2+p)}$ with GRF

$$I_{\mathbf{A}X^{(p)},k}(x) = \inf_{c>0} \left\{ \frac{\|x\|_2^p}{pc^p} + \frac{c^2}{2} \right\} = \frac{p+2}{2p} \|x\|_2^{2p/(p+2)}, \quad x \in \mathbb{R}^k. \tag{3.8}$$

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

In the sublinear regime, Proposition 3.4 shows that $\{X^{(n,p)}\}$ satisfies Assumption C with GRF $J_{X,p}$ (for all values of $r \in [0,\infty]$) and at speed $s_n = n^{2p/(2+p)}$ in case (a), which corresponds to $r = \infty$, and in case (b), which corresponds to $r = 1$, and speed $s_n = n^p k_n^{-p/2}$ in case (c), which corresponds to $r = 0$. Together with Theorem 2.11(ii), this immediately implies case (c) and case (a) follows from the second equality in (3.8) above. On the other hand, if case (b) holds, Theorem 2.11(ii) and the expression for $J_{X,p}$ in (3.6) show that for $x \geq 0$,

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

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

In a similar fashion, applying Theorem 2.9 and Theorem 2.15, one can also deduce LDPs for the empirical measures of the coordinates of the sequences of projections of $X^{(n)} = X^{(n,p)}$ in the sublinear and linear regimes. Furthermore, in the linear regime, similar LDP results for the norms of these projections can be deduced from Theorem 2.16 on noting that $\{X^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies Assumption A. But, we do not state these results since they were already obtained in Theorem 1.2 of [3].

Remark 3.6. Theorem 3.5 shows that when $p \in [1,2)$, the speed of the LDP for Euclidean norms of k_n-dimensional projections of ℓ_p^n balls exhibits an interesting phase transition depending on whether the ratio $k_n/n^{2p/(2+p)}$ is asymptotically finite or infinite, and the form of the rate function also differs depending on whether the limit is zero, strictly positive or infinite. Indeed, note that $x^p \geq x^p/\bar{c}(x)^p = \bar{c}(x)^2 - 1 \geq x^{2p/(p+2)}$. Hence, a comparison of cases (b) and (c) of the last theorem above shows that the faster the growth of the speed k_n, the faster the growth of the rate function $\mathbb{J}_{Y_{2,k_n}^{(p)}}(x)$, as $x \to \infty$. This disproves a conjecture in [3] (see the statement after Theorem 1.2 therein), which states that the LDP (that is, speed and rate function) of these Euclidean norms should be the same in the constant and sublinear cases. This behavior is in contrast to the case of ℓ_p balls, with $p > 2$, where the rate functions of the (Euclidean) norms of the projections in the sublinear and constant cases coincide (see Proposition 3.7 and Remark 3.9).

3.2.3. Verification of Assumption A when $p \in [2,\infty)$*

The verification of the asymptotic thin shell condition is much easier in the case $p \in [2,\infty)$ than when $p \in [1,2)$.

Proposition 3.7. *For $n \in \mathbb{N}$ and $p \in [2, \infty)$, $\{X^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies Assumption A* with $s_n = n$ and GRF $J_{X,p}$ where*

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

and for $p > 2$,

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

Moreover, $J_{X,p}$ has a unique minimizer at $m := m(p)$ defined in (3.3).

Proof. In the case $p = 2$, by (3.4), clearly $\|X^{(n,p)}\|_2/\sqrt{n} \stackrel{(d)}{=} U^{1/n}$. Thus, Lemma 3.3 of [21] shows that $\{X^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies Assumption A* with GRF $J_{X,2}$, and the explicit form of $J_{X,2}$ implies it has a unique minimum at $m(2) = 1$. On the other hand, when $p > 2$, the LDP follows from [28, Theorem 1.2] and uniqueness of the minimizer from [28, Theorem 1.1 (a)], with $d = 1$ and $q_i = 2$ (note that the ℓ_p^n ball is scaled differently there but the results nevertheless follow). □

The last result can also be deduced from Proposition 3.11 on Orlicz balls in Section 3.3 by choosing $V(x) = |x|^p$. However, we presented the self-contained proof above since the special case of ℓ_p balls with $p \geq 2$ is easier due to the representation (3.4). Finally, Proposition 3.7 and Theorem 2.7 together yield the following corollary.

Corollary 3.8. *Fix $p \in [2, \infty)$, let k_n be constant at $k \in \mathbb{N}$ and let $J_{X,p}$ be defined as in Proposition 3.7. Then $\{n^{-1/2}\mathbf{A}_{n,k}^T X^{(n,p)}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^k$ at speed n with GRF $I_{\mathbf{A}X^{(p)},k} : \mathbb{R}^k \to [0,\infty]$ defined by*

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

The particular case when $k_n \equiv 1$, for all range of $p \in [1, \infty)$ was studied in [21]; see Theorem 2.2 therein. Our method is different in that we take advantage of the equivalent representation in Lemma 4.2 and hence a different form of the rate function is obtained here.

Remark 3.9. One may combine Proposition 3.7, Theorem 2.11 and Theorem 2.16 to obtain LDPs for $\{n^{-1/2}\|\mathbf{A}_{n,k_n}^T X^{(n,p)}\|_2\}_{n\in\mathbb{N}}$ in the sublinear and linear regimes when $p \in [2, \infty)$. We omit these results, since they were already obtained in Theorem 1.1 of [3]. LDPs for empirical measures of the coordinates of the projections of the ℓ_p^n balls with $p \in [2, \infty)$ in the sublinear and linear regimes can also be deduced on combining Proposition 3.7 with Theorem 2.9 or Theorem 2.15.

3.3. Orlicz and generalized Orlicz balls

We now consider the class of *Orlicz and generalized Orlicz balls*. These form a natural class of examples because, like the uniform measure on any convex set, the uniform measure on a generalized Orlicz ball is logconcave, and much like ℓ_p^n balls, the coordinates of a uniformly distributed random vector satisfy a certain "subindependence" property known as "negative association" [49], and (under suitable conditions) the Orlicz ball satisfies the KLS conjecture [37].

Definition 3.10. We say V is an *Orlicz function* if $V : \mathbb{R} \to \mathbb{R}_+ \cup \{\infty\}$ is convex and satisfies $V(0) = 0$ and $V(x) = V(-x)$ for $x \in \mathbb{R}$. Let the domain of V be $D_V := \{x \in \mathbb{R} : V(x) < \infty\}$. We say V is *superquadratic* if

$$V(x)/x^2 \to \infty, \qquad \text{as } x \to \infty. \tag{3.9}$$

In particular, 2-convex Orlicz functions are superquadratic in the sense of Definition 3.10, and this includes the case $V(x) = |x|^p$, with $p > 2$. Fix a superquadratic Orlicz function V, and denote the associated symmetric Orlicz ball by

$$\mathbb{B}_V^n := \left\{ x \in \mathbb{R}^n : \sum_{i=1}^n V(x_i) \le n \right\}. \tag{3.10}$$

We also discuss so-called *generalized* Orlicz balls, associated with a sequence of superquadratic Orlicz functions $V_i : \mathbb{R} \to \mathbb{R}_+$, $i \in \mathbb{N}$. These generalized Orlicz balls, which are induced by a norm defined by functions that vary with dimension, are also known in the literature as *Musielak-Orlicz balls* [45]. For $n \in \mathbb{N}$, define the set

$$\mathbb{B}_{V_1,\ldots,V_n}^n := \left\{ x \in \mathbb{R}^n : \sum_{i=1}^n V_i(x_i) \le n \right\}. \tag{3.11}$$

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

We verify Assumption A* for symmetric and generalized Orlicz balls in Sections 3.3.1 and 3.3.2, respectively. As mentioned above, the special case $\mathbb{B}_V^n = n^{1/p}\mathbb{B}_p^n$, with $p > 2$, was analyzed in Section 3.2.3. However, unlike $\mathbb{B}_p^n$, whose study is facilitated by the probabilistic representation (3.4), there is no analogous representation for general $\mathbb{B}_V^n$, making its analysis significantly more complicated. Instead, our proof shows that one can represent $\mathbb{B}_V^n$ in terms of an exponential Gibbs measure; this idea has been subsequently further exploited in [27]. Conditions like 2-convexity (and 2-concavity) often arise in the local theory of Banach spaces. In our case this condition arises naturally because we are also considering 2-norms in this section. However, it would be interesting to investigate whether an analogous analysis can be carried out for Orlicz balls defined via superlinear but subquadratic Orlicz functions V (it is worth noting that we have been able to cover the special case of ℓ_p^n balls with $p \in [1,2)$) or whether there is a more fundamental obstruction. In the following, by a slight abuse of notation, we use $|A|$ to denote the volume of a measurable set $A \subset \mathbb{R}^n$ and use dx to denote the integral over Lebesgue measure in $\mathbb{R}$ or $\mathbb{R}^n$.

3.3.1. Symmetric Orlicz balls

In analogy with (1.1), for $\nu \in \mathcal{P}(\mathbb{R})$, define $M_V : \mathcal{P}(\mathbb{R}) \to \mathbb{R}_+$ as

$$M_V(\nu) := \int_{\mathbb{R}} V(x)\nu(dx), \tag{3.12}$$

for ν such that the integral in the last display is finite. Also, for $b > 0$, define $\mu_{V,b} \in \mathcal{P}(\mathbb{R})$ as follows: for $x \in D_V$, where D_V is the domain of V as specified in Definition 3.10,

$$\mu_{V,b}(dx) := \frac{1}{Z_{V,b}} e^{-b\,V(x)} dx, \tag{3.13}$$

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

$$\mathcal{J}(u,v) := \sup_{s<0, t\in\mathbb{R}} \left\{ su + tv - \log\left(\int_{D_V} e^{sV(x)+tx^2} dx \right) \right\} \quad \text{for} \quad u, v \in \mathbb{R}_+, \tag{3.14}$$

We now state the main result of this section.

Proposition 3.11 *(Annealed example, Orlicz). Suppose V is a superquadratic Orlicz function and for $n \in \mathbb{N}$, let $X^{(n)} \sim \mathrm{Unif}(\mathbb{B}_V^n)$. Then $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption A* with $J_X = J_{X,V}$, where*

$$J_{X,V}(z) := \mathcal{J}(1, z^2) - \sup_{s<0} \left\{ s - \log\left(\int_{D_V} e^{sV(x)} dx \right) \right\}, \quad z \in \mathbb{R}_+. \tag{3.15}$$

Moreover, there exists a unique $b^* > 0$ such that $M_V(\mu_{V,b^*}) = 1$ and $J_{X,V}$ has a unique minimizer $m := \sqrt{M_2(\mu_{V,b^*})}$, with M_2 as defined in (1.1).

The proof of Proposition 3.11, which is given at the end of the section, relies on certain properties of the function $\mathcal{J}$ specified in (3.14), which we state in the lemma below, whose proof is relegated to Appendix C. For $s < 0$, $t \in \mathbb{R}$, define $\nu_{s,t} \in \mathcal{P}(\mathbb{R})$ as

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

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

Lemma 3.12. *Suppose V is a superquadratic Orlicz function and let $\mathfrak{J}$ be defined in (3.14).*

1. *For $u, v \in \mathbb{R}_+$ and (u, v) lying in the interior of the domain of $\mathfrak{J}$, there exists a unique $(s, t) \in \mathbb{R}_- \times \mathbb{R}$ such that the supremum in the definition of (3.14) of $J(u, v)$ is attained. Moreover, this (s, t) satisfies $M_V(\nu_{s,t}) \leq u$ and $M_2(\nu_{s,t}) = v$.*
2. *There exists a unique constant $b^* > 0$ such that $M_V(\mu_{V,b^*}) = 1$. Moreover, $v \mapsto \mathfrak{J}(1, v)$ is a convex function with minimizer $m = M_2(\mu_{V,b^*})$, and furthermore,*

$$\min_{x\in\mathbb{R}_+} \mathfrak{J}(1, x) = \mathfrak{J}(1, m) = \sup_{s<0} \left\{ s - \log\left(\int_{D_V} e^{sV(x)}dx \right) \right\}. \tag{3.16}$$

3. *For $v > m$, the supremum in the definition (3.14) of $\mathfrak{J}(1, v)$ is attained at $(s, t) \in \mathbb{R}_- \times \mathbb{R}_+$ while for $0 < v < m$, the supremum is attained at $(s, t) \in \mathbb{R}_- \times \mathbb{R}_-$.*
4. *For $u, v \in \mathbb{R}_+$ such that $\mathfrak{J}(u, v) < \infty$, $\partial_u\mathfrak{J}(u, v) \leq 0$.*

Remark 3.13. Let $V, \tilde{V}$ be Orlicz functions, and suppose now that the definition of $\mathfrak{J}$ in (3.14) is replaced with

$$\tilde{\mathfrak{J}}(u, v) := \sup_{s<0,t\in\mathbb{R}} \left\{ su + tv - \log\left(\int_{D_V} e^{sV(x)+t\tilde{V}(x)}dx \right) \right\} \quad \text{for} \quad u, v \in \mathbb{R}_+. \tag{3.17}$$

Then, as can be seen from the proof in Appendix C, the conclusions of Lemma 3.12 continue to hold, with $M_{\tilde{V}}$ and $\tilde{\mathfrak{J}}$ in place of M_2 and $\mathfrak{J}$, respectively, as long as $V(x)/\tilde{V}(x) \to \infty$ as $x \to \infty$.

Proof of Proposition 3.11. Fix a superquadratic Orlicz function V (see Definition 3.10). For a measurable set $A \subset \mathbb{R}$, define $\mathbb{B}^n_{2,V}[A] := \mathbb{B}^n_V \cap \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 \in nA\}$, and note that

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n (X_i^n)^2 \in A \right) = \frac{|\mathbb{B}^n_{2,V}[A]|}{|\mathbb{B}^n_V|}, \tag{3.18}$$

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

Step 1. For any closed set $F \subset \mathbb{R}_+$, we will establish an upper bound on the numerator in (3.18),

$$\limsup_{n\to\infty}\frac{1}{n}\log\left|\mathbb{B}^n_{2,V}[F]\right| \le -\inf_{x\in F}\mathcal{J}(1,x). \tag{3.19}$$

With $b_* > 0$ as defined in property 2 of Lemma 3.12, and M_2 as defined in (1.1), set $m := M_2(\mu_{V,b^*})$. Next, set $\alpha_+ := \min\{x \in [m,+\infty)\cap F\}$ and $\alpha_- := \max\{x \in [0,m]\cap F\}$. Assume $\alpha_- < m < \alpha_+$. Then

$$\begin{aligned}\left|\mathbb{B}^n_{2,V}[F]\right| &\le \left|\mathbb{B}^n_V \cap \left\{x\in\mathbb{R}^n : \sum_{i=1}^n x_i^2 \ge n\alpha_+\right\}\right| + \left|\mathbb{B}^n_V \cap \left\{x\in\mathbb{R}^n : \sum_{i=1}^n x_i^2 \le n\alpha_-\right\}\right| \\ &= \left|\mathbb{B}^n_{2,V}[[\alpha_+,\infty)]\right| + \left|\mathbb{B}^n_{2,V}[[0,\alpha_-]]\right|. \end{aligned} \tag{3.20}$$

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

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

and therefore, it follows that

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

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

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

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

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

where, since $\alpha_+ > m$, the last equality follows by property 3 of Lemma 3.12.

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

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

where, noting that $\alpha_- < m$, the last inequality once again follows by property 3 of Lemma 3.12. Together with (3.20), the last two displays show that

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

where the last equality follows from the convexity of $x \to \mathcal{J}(1,x)$ and the fact that m is the unique minimizer of $\mathcal{J}(1,\cdot)$, as established in property 2 of Lemma 3.12, and the definitions of α_- and α_+. This proves the claim (3.19) in this case.

On the other hand, if $\alpha_+ = m$ or $\alpha_- = m$, then by property 2 of Lemma 3.12, $\inf_{x\in F} \mathcal{J}(1,x) = \mathcal{J}(1,m)$. We first obtain an estimate of the volume of the Orlicz ball $\mathbb{B}^n_V$. For $s < 0$,

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

The claim (3.19) then follows in this case as well since

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

where the second inequality follows on taking the infimum over $s < 0$ in the last display and the equality on the third line follows from (3.16).

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

$$\liminf_{n\to\infty} \frac{1}{n} \log \left|\mathbb{B}^n_{2,V}[U]\right| \ge - \inf_{x\in U} \mathcal{J}(1,x). \tag{3.21}$$

Since $\inf_{x\in U} \mathfrak{J}(1,x) = \inf_{x\in U\cap D_{\mathfrak{J}(1,\cdot)}} \mathfrak{J}(1,x)$, with the latter being infinity if the intersection is empty, it suffices to only consider U that lies in the interior of the domain of $\mathfrak{J}(1,\cdot)$. By property 4 of Lemma 3.12, $\partial_u \mathfrak{J}(u,v) \le 0$, and so for each $v \in \mathbb{R}_+$, $u \mapsto \mathfrak{J}(u,v)$ is decreasing. By the convexity, and hence continuity, of $(u,v) \mapsto \mathfrak{J}(u,v)$ in $\mathbb{R}_+ \times \mathbb{R}_+$, for every $\varepsilon > 0$, there exist $0 < y < 1$ and $z \in U$ such that

$$\inf_{x\in U} \mathfrak{J}(1,x) > \mathfrak{J}(y,z) - \varepsilon. \tag{3.22}$$

Pick $\delta > 0$ small such that $(z-\delta, z+\delta) \subset U$. Let (s_y, t_z) be the value that attains the supremum in the expression (3.14) for $\mathfrak{J}(y,z)$, which exists by property 1 of Lemma 3.12 since U lies in the interior of the domain of $\mathfrak{J}(1,\cdot)$. Define

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

and $\bar{Z}_{y,z} := Z_{s_y,t_z} = \int_{D_V} e^{s_y V(x) + t_z x^2} dx$, which is finite by the definition of (s_y, t_z), the definition of $\mathfrak{J}(y,z)$ in (3.14) and the finiteness of $\mathfrak{J}(y,z)$. Suppose $t_z < 0$. We then obtain the following estimate for the lower bound:

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

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

Let $\{\Xi_i\}_{i=1}^\infty$ be a sequence of i.i.d. continuous random variables with density $\frac{1}{\bar{Z}_{y,z}} e^{s_y V(x) + t_z x^2}$. Since (s_y, t_z) attains the supremum in (3.14) of $\mathfrak{J}(y,z)$, by property 1 of Lemma 3.12, $\mathbb{E}[V(\Xi_i)] \le y < 1$ and $\mathbb{E}[\Xi_i^2] = z$. Then the integral in the last display can be rewritten as

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

which converges to 1 by the weak law of large numbers and finiteness of the respective moments. Thus, combining the last two displays with the expression for $\mathfrak{J}(y,z)$ in (3.14) and the definition of (s_y, t_z), we obtain

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

where the last inequality invokes (3.22). Since ε and δ are arbitrary, this implies the desired lower bound in (3.21).

Step 3. We claim that the denominator of (3.18) satisfies

$$\lim_{n\to\infty} \frac{1}{n} \log |\mathbb{B}^n_V| = -\sup_{s<0} \left\{ s - \log \left(\int_{D_V} e^{sV(x)} dx \right) \right\}. \tag{3.23}$$

Since $\mathbb{B}^n_V = \mathbb{B}^n_{2,V}[\mathbb{R}_+]$, applying (3.19) with $F = \mathbb{R}_+$ and (3.21) with $U = (0,\infty)$, we obtain

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

The claim then follows from (3.16) of Lemma 3.12.

Finally, the combination of (3.23), (3.19) and (3.21) with (3.18) yields an LDP for $\{\frac{1}{n}\sum_{i=1}^n (X_i^n)^2\}_{n\in\mathbb{N}}$ at speed n with rate function

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

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

Remark 3.14. The relation (3.23) provides the asymptotic logarithmic volume of an Orlicz ball. A careful inspection of the argument shows that this limit result does not require the super-quadratic restriction on the Orlicz function; it is the LDP for the norms that uses this restriction. Indeed, let V and $\tilde{V}$ be Orlicz functions such that $V(x)/\tilde{V}(x) \to \infty$ as $x \to \infty$, let $\tilde{\mathcal{J}}$ be defined as in (3.17). and for any measurable set $A \subset \mathbb{R}$ define $\mathbb{B}^n_{\tilde{V}}[A] := \{x \in \mathbb{R}^n : \sum_{i=1}^n \tilde{V}(x_i) \in nA\}$. Then, we see from Step 1 in the proof of Proposition 3.11 that the following more general inequality holds:

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

Remark 3.15. After the first version of this paper was posted on arXiv, the asymptotic logarithmic volume of an Orlicz ball was also obtained in [27], where they also established further refined estimates using ideas from sharp large deviations estimates (see, e.g., [7,40]).

3.3.2. Generalized Orlicz balls

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

$$\frac{1}{n} \log \frac{\left|\mathbb{B}^n_{\bar{V}} \,\Delta\, \mathbb{B}^n_{V_1,\dots,V_n}\right|}{\left|\mathbb{B}^n_{\bar{V}} \cup \mathbb{B}^n_{V_1,\dots,V_n}\right|} \to -\infty, \tag{3.24}$$

where $\mathbb{B}^n_{\bar{V}}$ is the Orlicz ball as defined in (3.10), and for any sets $A, B \subset \mathbb{R}^n$, $A\Delta B = [A \setminus B] \cup [B \setminus A]$ represents the symmetric difference of A and B.

Lemma 3.16 *(Generalized Orlicz). Fix Orlicz functions $\{V_i\}_{i\in\mathbb{N}}$ and $\bar{V}$ that satisfy (3.24), and are uniformly superquadratic in the sense that there exists a constant $c_V \in (0,\infty)$ such that for $x > c_V$, $V_i(x) > x^2$ for $i \in \mathbb{N}$ and $\bar{V}(x) > x^2$. Suppose $X^{(n)} \sim \mathrm{Unif}(\mathbb{B}^n_{V_1,\dots,V_n})$. Then, $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption A* with $J_{X,\bar{V}}$, defined as in (3.15), but with V replaced with $\bar{V}$.*

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

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

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

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

Hence, $W^{(n)} \overset{(\mathrm{d})}{=} X^{(n)}$. A similar argument can be used to show that $\bar{W}^{(n)} \overset{(\mathrm{d})}{=} \bar{X}^{(n)}$, thus providing a useful coupling between the uniform measures on $\mathbb{B}^n_{V_1,\dots,V_n}$ and $\mathbb{B}^n_{\bar{V}}$. Due to the contraction principle applied to $x \mapsto \sqrt{x}$, Proposition 3.11 and Remark 1.6, it suffices to show that $\{\|W^{(n)}\|_2^2/n\}_{n\in\mathbb{N}}$ is exponentially equivalent (see Definition 1.5) to $\{\|\bar{W}^{(n)}\|_2^2/n\}_{n\in\mathbb{N}}$.

To this end, let $\kappa := \sqrt{c_V^2+1}$, where c_V is the positive finite constant stated in the lemma, and note that $\mathbb{B}^n_{V_1,\dots,V_n} \subset \kappa\sqrt{n}\mathbb{B}^n_2$ and $\mathbb{B}^n_{\bar{V}} \subset \kappa\sqrt{n}\mathbb{B}^n_2$ since for $x \in \mathbb{B}^n_{V_1,\dots,V_n}$, $\sum_{i=1}^n x_i^2 \le \sum_{i=1}^n (c_V^2 + V_i(x_i)) \le (c_V^2+1)n = \kappa^2 n$, where the last inequality uses (3.11), and for $x \in \mathbb{B}^n_{\bar{V}}$, $\sum_{i=1}^n x_i^2 \le \sum_{i=1}^n (c_V^2 + \bar{V}(x_i)) \le (c_V^2+1)n = \kappa^2 n$, where the last inequality uses (3.10). Thus,

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

Therefore, for every $\delta > 0$,

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

which converges to $-\infty$ as $n \to \infty$ by (3.24). This establishes the desired exponential equivalence and thus completes the proof. □

3.4. Gibbs measures

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

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

As in [17], we assume that there exists a non-atomic, σ-finite measure ℓ on $\mathbb{R}$ which, along with the potentials F and G, satisfies the following conditions:

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

$$\int_{A\times A} [F(x)+F(y)+G(x,y)]\,\ell(dx)\ell(dy) < \infty;$$

4. for all $\lambda \in \mathbb{R}$, we have

$$\int_{\mathbb{R}\times\mathbb{R}} \exp\left[\lambda(x^2+y^2) - F(x) - F(y) - G(x,y)\right] \ell(dx)\ell(dy) < \infty.$$

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

$$P_n(dx) := \frac{1}{Z_n} e^{-n\mathbf{H}_n(x)} \ell^{\otimes n}(dx), \quad x \in \mathbb{R}^n. \tag{3.25}$$

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

Theorem 3.17 *(Theorem 2.7 and Lemma 2.6 of [17]). Under the conditions stated above, $\{Q_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}_2(\mathbb{R})$ equipped with the 2-Wasserstein topology, at speed n, with GRF $\mathcal{I}_* : \mathcal{P}_2(\mathbb{R}) \to [0,\infty]$ defined by*

$$\mathcal{I}_*(\mu) := \mathcal{I}(\mu) - \inf_{\mu\in\mathcal{P}_2(\mathbb{R})} \mathcal{I}(\mu), \tag{3.26}$$

$$\mathcal{I}(\mu) := H(\mu|\ell) + \frac{1}{2}\int_{\mathbb{R}\times\mathbb{R}} G(x,y)\mu(dx)\mu(dy) + \int_{\mathbb{R}} F(x)\mu(dx),$$

with H being the relative entropy functional defined in (1.5).

Proposition 3.18 *(Gibbs measures). Suppose F, G and ℓ satisfy the conditions stated above, and for $n \in \mathbb{N}$, suppose $X^{(n)}$ is drawn from P_n of (3.25). Then $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption A* with GRF*

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

with $\mathcal{I}_*$ as defined in (3.26).

Proof. The LDP for $\{\|X^{(n)}\|_2^2/n\}_{n\in\mathbb{N}}$, follows from an application of the contraction principle to the empirical measure LDP of Theorem 3.17 and the second moment map M_2 of (1.1), which is continuous with respect to the 2-Wasserstein topology. Assumption A* then follows on applying the contraction principle with the map $x \mapsto \sqrt{x}$. □

Remark 3.19. In a similar fashion, the large deviation results in the recent work of [41, Theorem 2.8] can be combined with the contraction principle to show that Assumption A* is also satisfied by sequences of Gibbs measures associated with a class of interaction potentials $G : \mathbb{R}^k \to (-\infty, \infty]$ that capture k-tuple interactions, for fixed $k \in \mathbb{N}$.

4. Proofs of LDPs for the random projections

In this section, we prove the main large deviation results, namely, Theorems 2.7, 2.9 and 2.15, stated in Section 2, which deal with the constant, sublinear and linear regimes, respectively. At many points, we will refer to certain properties of the top row of $\mathbf{A}_{n,k_n}$, which we first establish in Section 4.1. Sections 4.2, 4.3 and 4.4 consider the regimes of $\{k_n\}$ constant, sublinear and linear, respectively. Throughout, let $\zeta_1, \zeta_2, \ldots,$ denote a sequence of i.i.d. standard Gaussian random variables, let $\zeta^{(n)} := (\zeta_1, \ldots, \zeta_n) \in \mathbb{R}^n$, let $e_1 := (1, 0, \ldots, 0) \in \mathbb{R}^n$.

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

Lemma 4.1. *Fix $k, n \in \mathbb{N}$ such that $k \le n$. Then the following relation holds:*

$$\mathbf{A}_{n,k}(1, \cdot\,) = (\mathbf{A}_{n,k}(1,1), \ldots, \mathbf{A}_{n,k}(1,k)) \overset{(\mathrm{d})}{=} \frac{(\zeta_1, \ldots, \zeta_k)}{\|\zeta^{(n)}\|_2}. \tag{4.1}$$

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

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

which implies that $\mathbf{A}_{n,k}(1, \cdot\,)$ is equal in distribution to the vector of the first k elements in the top row of $\mathbf{O}_n$. The marginal distribution of the top row of $\mathbf{O}_n$ is the uniform measure on the unit sphere of $\mathbb{R}^n$, which establishes the identity (4.1), due to the classical fact that $\zeta^{(n)}/\|\zeta^{(n)}\|_2$ is uniformly distributed on the unit sphere in $\mathbb{R}^n$ (see, e.g., [44]). □

Lemma 4.2. *Fix $n \in \mathbb{N}$ and $k \le n$. Suppose $X^{(n)}$ is an n-dimensional random vector independent of $\mathbf{A}_{n,k}$. Then the following relation holds:*

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

Proof. As in Lemma 6.3 of [21], which considered the case $k = 1$, this result is easily deduced from the fact that the distribution of $\mathbf{A}_{n,k}$ is invariant under orthogonal transformations and independent of $X^{(n)}$. In particular, fix $n \in \mathbb{N}$ and given any $n \times n$ orthogonal matrix O_n and $x \in \mathbb{R}^n$, we have

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

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

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

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

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

$$\hat{\mu}_{\mathbf{A}}^n := \frac{1}{k_n}\sum_{j=1}^{k_n} \delta_{\sqrt{n}\mathbf{A}_{n,k_n}(1,j)}, \quad n \in \mathbb{N}. \tag{4.2}$$

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

Lemma 4.3. *For $n \in \mathbb{N}$, let $X^{(n)}$ be independent of $\mathbf{A}_{n,k_n}$ and recall the definition of L^n given in (2.1). Then we have*

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

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

Proof. For each $n \in \mathbb{N}$, using the definition of L^n from (2.1) and Lemma 4.2, we have

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

from which the first assertion of the lemma follows by (4.2). Continuity of the map $(\nu, c) \mapsto \nu(\,\cdot \times c^{-1})$ follows if and only if as $n \to \infty$ a sequence of random variables $\{B_n\}_{n\in\mathbb{N}}$ converging in distribution to B and a sequence of constants $\{c_n\}$ converging to c implies that $c_n B_n$ converges to cB in distribution as $n \to \infty$, which follows from Slutsky's theorem [36, Theorem 13.18]. □

Lemma 4.4. *Suppose $\{\hat{\mu}_{\mathbf{A}}^n\}_{n\in\mathbb{N}}$ and $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfy LDPs, both at speed s_n, and with GRFs $\mathbb{I}_1 : \mathcal{P}_q(\mathbb{R}) \to [0,\infty]$ and $\mathbb{I}_2 : \mathbb{R}_+ \to [0,\infty]$, respectively. Then $\{L^n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF defined by*

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

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

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

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

Lemma 4.5. *Let $\{\zeta_j\}_{j\in\mathbb{N}}$ be i.i.d. $\mathcal{N}(0,1)$ random variables, $\zeta^{(n)} = (\zeta_1, \dots, \zeta_n)$ and consider the sequence*

$$\nu_n := \frac{1}{k_n}\sum_{j=1}^{k_n} \delta_{\zeta_j}, \quad n \in \mathbb{N}. \tag{4.3}$$

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

$$\chi_{\gamma_1}(\nu) := \begin{cases} 0 & \text{if } \nu = \gamma_1 \\ +\infty & \text{else} \end{cases}, \quad \nu \in \mathcal{P}(\mathbb{R}). \tag{4.4}$$

Proof. The first claim is a direct conclusion of Sanov's theorem [14, Theorem 2.1.10] while the second claim follows since $H(\cdot|\gamma_1) : \mathcal{P}(\mathbb{R}) \to [0,\infty]$ is convex with a unique minimizer at γ_1, as in the observation of Remark 1.3. □

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

Remark 4.6. Let $\{\zeta_j\}_{j\in\mathbb{N}}$ be i.i.d. $\mathcal{N}(0,1)$ random variables, let $\{k_n\}_{n\in\mathbb{N}}$ be a sequence of positive integers that converges to infinity and let $C_n := \frac{1}{k_n}\sum_{j=1}^{k_n}\zeta_j^2$, $n \in \mathbb{N}$. It is easy to see that the log moment generating function of ζ_1^2 is given by $-\frac{1}{2}\log(1-2t)$ for $t < 1/2$ and infinity, otherwise. Hence, it is an immediate consequence of Cramér's theorem [14, Theorem 2.2.1] that $\{C_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ at speed k_n with GRF $J_{\zeta^2} : \mathbb{R} \to [0,\infty]$ of the form

$$J_{\zeta^2}(t) := \sup_{s\in\mathbb{R}}\left\{st + \frac{1}{2}\log(1-2s)\right\} = \begin{cases} \frac{t-1}{2} - \frac{1}{2}\log t, & t > 0, \\ +\infty, & t \le 0. \end{cases} \tag{4.5}$$

4.2. Proof of the LDP for random projections in the constant regime

This section is devoted to the proof of Theorem 2.7. Throughout this section, fix $k \in \mathbb{N}$ and suppose $k_n = k$ for each $n \in \mathbb{N}$. We first give the main idea behind the proof. Due to Lemma 4.2 we have

$$\frac{1}{\sqrt{n}}\mathbf{A}_{n,k}^T X^{(n)} \overset{\text{(d)}}{=} \frac{1}{\sqrt{n}}\mathbf{A}_{n,k}^T e_1 \|X^{(n)}\|_2 = \mathbf{A}_{n,k}(1,\cdot)\frac{\|X^{(n)}\|_2}{\sqrt{n}} \tag{4.6}$$

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

Proof of Theorem 2.7. We consider two cases.

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

First, note that by Lemma 4.1, we can apply the LDP in [8, Theorem 3.4] to find that $\{\mathbf{A}_{n,k}(1,\cdot)\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^k$ at speed n with GRF $\mathcal{J}_k(y) := -\frac{1}{2}\log(1-\|y\|_2^2)$, when $\|y\|_2 \le 1$, and $\mathcal{J}_k(y) := +\infty$ otherwise. Since $\{X^{(n)}\}_{n\in\mathbb{N}}$ is independent of $\{\mathbf{A}_{n,k}\}_{n\in\mathbb{N}}$, and the case assumption shows that $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP with speed n and GRF J_X, Lemma 1.9 implies that $\{\mathbf{A}_{n,k}(1,\cdot), \|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n with GRF $\mathcal{J}_k(y) + J_X(\alpha)$ for $y \in \mathbb{R}^k$ and $\alpha \in \mathbb{R}_+$. The contraction principle applied to the mapping $\mathbb{R}^k \times \mathbb{R}_+ \ni (y,\alpha) \mapsto \alpha y \in \mathbb{R}^k$ implies that $\{n^{-1/2}\mathbf{A}_{n,k}^T X^{(n)}\}_{n\in\mathbb{N}}$ satisfies an LDP with GRF

$$I_{\mathbf{A}X,k}(x) := \inf_{y\in\mathbb{R}^k, z\in\mathbb{R}} \left\{-\tfrac{1}{2}\log(1-\|y\|_2^2) + J_X(z) : x = yz,\ \|y\|_2 \le 1,\ z \ge 0\right\}, \quad x \in \mathbb{R}^k.$$

We can without loss of generality restrict the range of z in the infimum to $z > 0$; then, substituting $y = x/z$ and noting that the constraint $\|y\|_2 \le 1$ is equivalent to $\|x\|_2 \le z$, it follows that

$$\begin{aligned} I_{\mathbf{A}X,k}(x) &= \inf_{z \ge \|x\|_2} \left\{-\tfrac{1}{2}\log\left(1 - \frac{\|x\|_2^2}{z^2}\right) + J_X(z)\right\} \\ &= \inf_{z > \|x\|_2} \left\{-\tfrac{1}{2}\log\left(1 - \frac{\|x\|_2^2}{z^2}\right) + J_X(z)\right\},\ x \in \mathbb{R}^k. \end{aligned}$$

On rewriting the above in terms of $c = \|x\|_2/z$, we obtain the form (2.3) for the rate function $I_{\mathbf{A}X,k}$.

Case 2. Suppose Assumption B holds with sequence $\{s_n\}_{n\in\mathbb{N}}$ and GRF J_X.

From Lemma 4.2 and (4.1), denoting $\zeta^{(n)} := (\zeta_1, \dots, \zeta_n)$ with $\{\zeta_i\}_{i\in\mathbb{N}}$ i.i.d. $\mathcal{N}(0,1)$ random variables, we have

$$\frac{1}{\sqrt{n}}\mathbf{A}_{n,k}^T X^{(n)} \stackrel{(d)}{=} \mathbf{A}_{n,k}(1,\cdot)\frac{\|X^{(n)}\|_2}{\sqrt{n}} \stackrel{(d)}{=} \frac{\zeta^{(k)}/\sqrt{s_n}}{\|\zeta^{(n)}\|_2/\sqrt{n}}\frac{\sqrt{s_n}\,\|X^{(n)}\|_2}{n}, \tag{4.7}$$

Now, consider the sequence of vectors

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

which are almost surely well-defined. Then $\zeta_i/\sqrt{s_n}$ is a $\mathcal{N}(0, 1/s_n)$ random variable with $s_n \to \infty$ as $n \to \infty$. Hence, by the Gärtner-Ellis theorem, the sequence $\{\zeta^{(k)}/\sqrt{s_n}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^k$ at speed s_n with GRF $x \mapsto \|x\|^2/2$. Assumption B implies that $\{\sqrt{s_n}\|X^{(n)}\|_2/n\}_{n\in\mathbb{N}}$ satisfies an LDP, also at speed s_n, and with GRF J_X. Finally, by Remark 4.6, the contraction principle applied to $x \mapsto 1/\sqrt{x}$ and the strong law of large numbers, $\{\sqrt{n}/\|\zeta^{(n)}\|_2\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n and converges almost surely to 1, which implies that the associated GRF has a unique minimum at 1.

Together with the independence of $\{\zeta_i\}_{i=1,\dots,k}$ and $X^{(n)}$, and Lemma 1.9 (with $U_n = \zeta^{(k)}/\sqrt{s_n}$, $V_n = \sqrt{s_n}\|X^{(n)}\|_2/n$, $W_n = \sqrt{n}/\|\zeta^{(n)}\|_2$ and $m = 1$), this implies that R_n satisfies an LDP with GRF given by $\mathbb{R}^k \times \mathbb{R}_+ \times \mathbb{R}_+ \ni (r_1, r_2, r_3) \mapsto \frac{\|r_1\|_2^2}{2} + J_X(r_3)$ if $r_2 = 1$ (and $+\infty$ otherwise). Combining this with (4.7) and the contraction principle for the continuous mapping $\mathbb{R}^k \times \mathbb{R}_+^2 \ni (r_1, r_2, r_3) \mapsto r_1 r_2 r_3 \in \mathbb{R}^k$, it follows that the sequence of k-dimensional vectors, $\{n^{-1/2}\mathbf{A}_{n,k}^T X^{(n)}\}_{n\in\mathbb{N}}$, satisfies an LDP at speed s_n with GRF

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

which coincides with the expression given in (2.3). This completes the proof of Theorem 2.7. □

4.3. *Proof of the empirical measure LDP in the sublinear regime* $1 \ll k_n \ll n$

We now present the proof of Theorem 2.9. We first start with an auxiliary result.

Proposition 4.7. *Fix $q \in [1,2)$. Suppose $\{k_n\}_{n\in\mathbb{N}}$ grows sublinearly. Then, $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ defined in (4.2) satisfies an LDP in $\mathcal{P}_q(\mathbb{R})$ at speed k_n with GRF $H(\cdot|\gamma_1) : \mathcal{P}(\mathbb{R}) \to [0,\infty]$.*

Proof. Recall that $\{\zeta_j\}_{j\in\mathbb{N}}$ are i.i.d. $\mathcal{N}(0,1)$ random variables, $\zeta^{(n)} = (\zeta_1,\dots,\zeta_n)$. Due to (4.2) and Lemma 4.1, $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}} \overset{(d)}{=} \{\tilde{\nu}_n\}_{n\in\mathbb{N}}$, where

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

We now claim (and justify below) that $\{\nu_n\}_{n\in\mathbb{N}}$ defined in Lemma 4.5 is exponentially equivalent (at speed k_n) to the sequence $\{\tilde{\nu}_n\}_{n\in\mathbb{N}}$ defined in (4.3).

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

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

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

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

Since ζ_1 has a finite exponential moment, Cramér's theorem [14, Theorem 2.2.1] implies

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

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

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

Combining the last three displays, we find that

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

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

In light of Remark 1.6, the last observation taken together with Lemma 4.5 implies that $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\mathbb{R})$ at speed k_n with GRF $H(\cdot|\gamma_1)$. In order to strengthen the LDP for $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$, by [14, Corollary 4.2.6], it suffices to show that $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ is exponentially tight in $\mathcal{P}_q(\mathbb{R})$. Since $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}\overset{(d)}{=}\{\tilde{\nu}_n\}_{n\in\mathbb{N}}$ and for each $j>0$, the set $K_{2,j}$ defined in (1.2) is compact with respect to $\mathcal{P}_q(\mathbb{R})$ due to Lemma 1.1, it suffices to show that for $\varepsilon>0$, there exists $M<\infty$ such that

$$\lim_{n\to\infty}\frac{1}{k_n}\log\mathbb{P}(\tilde{\nu}_n\in K^c_{2,M})=\lim_{n\to\infty}\frac{1}{k_n}\log\mathbb{P}(M_2(\tilde{\nu}_n)>M)<-\varepsilon. \tag{4.8}$$

Now note that $\{M_2(\tilde{\nu}_n)=\frac{n}{k_n}\sum_{j=1}^{k_n}\frac{\zeta_j^2}{\|\zeta^{(n)}\|_2}\}_{n\in\mathbb{N}}$ and by Remark 4.6, $\{\|\zeta^{(k_n)}\|_2^2/k_n\}_{n\in\mathbb{N}}$ and $\{\|\zeta^{(n)}\|_2^2/n\}_{n\in\mathbb{N}}$ satisfy LDPs at speed k_n and n, respectively, both with the same GRF $J_{\zeta^2}(x)$ defined in (4.5). Thus, $\{n/\|\zeta^{(n)}\|_2^2\}_{n\in\mathbb{N}}$ also satisfies an LDP at speed n by the contraction principle. Together with Lemma 1.9 and the fact that J_{ζ^2} has a unique minimizer at 1, this implies the sequence $\{M_2(\tilde{\nu}_n)\}_{n\in\mathbb{N}}$ satisfies an LDP at speed k_n with GRF J_{ζ^2}. Now, for $\varepsilon>0$, pick M such that $J_{\zeta^2}(M)>\varepsilon$. We then obtain

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

which proves (4.8), and thus concludes the proof of the lemma. □

Proof of Theorem 2.9. By Assumption A, $\|X^{(n)}\|_2/\sqrt{n}$ satisfies an LDP at speed s_n with GRF J_X and when $s_n\gg k_n$ by Remark 1.3 (and the additional assumption that J_X has a unique minimizer m), $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ also satisfies an LDP at speed k_n with (degenerate) rate function χ_m. Also, by Proposition 4.7 and Remark 1.3, $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed k_n with GRF $H(\cdot|\gamma_1)$ and, when $s_n\ll k_n$, $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ also satisfies an LDP at speed s_n with (degenerate) GRF χ_{γ_1} defined in Remark 1.3 (see also Lemma 4.5).

Combining the above observations with Lemma 4.4, we see that $\{L^n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $s_n\wedge k_n$ with GRF $\mathbb{I}_{L,k_n}$, where when $s_n\gg k_n$,

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

$$= H(\mu|\gamma_m),$$

when $s_n = k_n$,

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

and when $s_n \ll k_n$,

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

This proves Theorem 2.9. □

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

Throughout this section, suppose $\{k_n\}_{n\in\mathbb{N}}$ grows linearly with rate $\lambda \in (0, 1]$. As in the sublinear regime, we first analyze the sequence of empirical measures $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ of (4.2) as a precursor to the analysis of $\{L^n\}_{n\in\mathbb{N}}$ of (2.1).

Proposition 4.8. *Fix $q \in [1, 2)$. Suppose k_n grows linearly with rate $\lambda \in (0, 1]$. Then, the sequence $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}_q(\mathbb{R})$ at speed n with GRF $\mathcal{H}_\lambda$ of (2.10).*

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

Proof of Theorem 2.15. Suppose Assumption A is satisfied with corresponding sequence $\{s_n\}_{n\in\mathbb{N}}$. First, suppose $s_n = n$. Then, along with Proposition 4.8 and Lemma 4.4, this implies that $\{L^n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n with GRF $\hat{\mathbb{I}}_{L,\lambda}$, defined by

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

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

$$\hat{\mathbb{I}}_{L,\lambda}(\mu) = \inf_{c>\sqrt{\lambda M_2(\mu)}} \{\mathcal{H}_\lambda(\mu(\cdot \times c)) + J_X(c)\}, \quad \mu \in \mathcal{P}(\mathbb{R}). \tag{4.9}$$

Now for $\mu \in \mathcal{P}(\mathbb{R})$ and $c > \sqrt{\lambda M_2(\mu)}$, the definitions of $\mathcal{H}_\lambda$ and h in (2.10) and (1.4), respectively, imply

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

When substituted back into (4.9), this shows that for every $\mu \in \mathcal{P}(\mathbb{R})$, $\hat{\mathbb{I}}_{L,\lambda}(\mu) = \mathbb{I}_{L,\lambda}(\mu)$, with the latter defined as in (2.11).

Next, consider the case $s_n \ll n$. By Lemma 4.8, $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF $\mathcal{H}_\lambda$. Therefore, by Remark 1.3 and Remark 2.14, $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ also satisfies an LDP at speed s_n with GRF χ_{γ_1}. Then, by Assumption A and Lemma 4.4, $\{L^n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF

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

which coincides with the expression for $I_{L,\lambda}(\mu)$ given in Theorem 2.15 for the case $s_n \ll n$. $\square$

Remark 4.9. Note that the above proof in particular shows that, under Assumption A, the rate function $\mathbb{I}_{L,\lambda}(\mu)$ from Theorem 2.15 in the linear regime in the case $s_n = n$, satisfies, for $\lambda > 0$,

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

with $\mathcal{H}_\lambda$ defined as in (2.10).

The rest of this section is devoted to the proof of Proposition 4.8, which is broken down into the intermediate steps given by Lemmas 4.10 and 4.11 below.

Lemma 4.10. *Suppose $k_n/n \to \lambda \in (0,1]$ as $n\to\infty$, and $\{\zeta_j\}_{j\in\mathbb{N}}$ is an i.i.d. sequence with common law γ_1 (the standard normal distribution). If $\lambda\in(0,1)$, then the sequence*

$$\left(\frac{1}{k_n}\sum_{j=1}^{k_n}\delta_{\zeta_j}\,,\quad \frac{1}{n-k_n}\sum_{j=k_n+1}^{n}\delta_{\zeta_j}\,,\quad \frac{1}{n}\sum_{j=1}^{n}\zeta_j^2\right),\quad n\in\mathbb{N}, \tag{4.10}$$

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

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

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

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

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

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

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

Proof. Our approach is to apply the approximate contraction principle of Proposition A.1 and Corollary A.2 of the appendix, with the following parameters:

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

First, let $\lambda \in (0,1)$, consider $(\mathscr{L}_n, \mathscr{C}_n) = (\mathscr{L}_n^{(1)}, \mathscr{C}_n^{(1)})$. Then the domain $\mathcal{D}$ of (A.3) takes the form $\mathcal{D} = \{\alpha \in \mathbb{R} : \log \mathbb{E}[e^{\lambda^{-1}\alpha\,\zeta_1^2}] < \infty\} = (-\infty, \frac{\lambda}{2})$, and so $0 \in \mathcal{D}^\circ = \mathcal{D}$. Thus, $F(x) = \sup_{\alpha < \lambda/2} \alpha x$, which is equal to $\lambda x/2$ if $x \ge 0$, and equal to ∞ otherwise, and $\int \mathbf{c} d\mu = M_2(\mu)$. Thus, by Corollary A.2, the sequence $\{\mathscr{L}_n^{(1)}, \mathscr{C}_n^{(1)}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n and with GRF $J^{(1)} = J_\lambda^{(1)}$, defined by

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

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

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

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

$$\left\{ \mathscr{L}_n^{(1)},\, \mathscr{L}_n^{(2)},\, \lambda \int \mathbf{c}\, d\mathscr{L}_n^{(1)} + (1-\lambda) \int \mathbf{c}\, d\mathscr{L}_n^{(2)} \right\}_{n\in\mathbb{N}} \tag{4.11}$$

satisfies an LDP at speed n with GRF

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

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

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

are exponentially equivalent (recall Definition 1.5). Hence, the sequence in (4.11) is also exponentially equivalent to (4.10), and the latter thus satisfies an LDP at speed n with GRF $I_{1,\lambda}$.

Now, let $\lambda=1$. We see from the derivation above that the sequence $\{\mathscr{L}_n^{(1)},\mathscr{C}_n^{(1)}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n with GRF $J^{(1)}$, and that the two sequences $\{\frac{1}{k_n}\sum_{j=1}^{k_n}\zeta_j^2\}_{n\in\mathbb{N}}$ and $\{\frac{1}{n}\sum_{j=1}^{n}\zeta_j^2\}_{n\in\mathbb{N}}$ are exponentially equivalent. Hence, from Remark 1.6, we obtain the desired LDP. □

Lemma 4.11. *Suppose $k_n/n\to\lambda\in(0,1]$ and let $\{\zeta_j\}_{j\in\mathbb{N}}$ be as in Lemma 4.10. If $\lambda\in(0,1)$, the sequence of pairs of measures*

$$\left(\frac{1}{k_n}\sum_{j=1}^{k_n}\delta_{\sqrt{n}\zeta_j/\|\zeta^{(n)}\|_2},\quad \frac{1}{n-k_n}\sum_{j=k_n+1}^{n}\delta_{\sqrt{n}\zeta_j/\|\zeta^{(n)}\|_2}\right),\quad n\in\mathbb{N},\tag{4.12}$$

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

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

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

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

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

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

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

is continuous. Then, applying the contraction principle with this map to the LDP of Lemma 4.10, we find that the sequence in (4.12) satisfies an LDP with GRF:

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

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

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

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

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

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

We now turn to the proof of Proposition 4.8. First, for $\lambda \in (0,1]$, define

$$
\mathbf{J}_{2,\lambda}(z) := \begin{cases} \frac{\lambda}{2}\log\left(\frac{\lambda}{z^2}\right) + \frac{1-\lambda}{2}\log\left(\frac{1-\lambda}{1-z^2}\right), & 0 < z < 1, \\ \infty, & \text{otherwise}, \end{cases} \tag{4.13}
$$

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

Proof of Proposition 4.8. We first prove that $\{\hat\mu^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\mathbb{R})$ with respect to the weak topology, with GRF $\mathcal{H}_\lambda$ of (2.10). Due to the representation for $\mathbf{A}_{n,k_n}(1,\cdot)$ established in Lemma 4.1, it suffices to show that the sequence

$$\hat{\nu}_n := \frac{1}{k_n}\sum_{j=1}^{k_n} \delta_{\sqrt{n}\zeta_j/\|\zeta^{(n)}\|_2}, \quad n \in \mathbb{N}, \tag{4.14}$$

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

We deduce in the following the case for $\lambda \in (0,1)$. The case for $\lambda = 1$ can be shown using a similar calculation and is hence omitted. Combining the LDP established in Lemma 4.11 with the contraction principle applied to the projection mapping: $(\mathcal{P}(\mathbb{R}))^2 \ni (\pi_1, \pi_2) \mapsto \pi_1 \in \mathcal{P}(\mathbb{R})$, it follows that $\{\hat{\nu}_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n with GRF given by $\mu \mapsto \inf_{\nu\in\mathcal{P}(\mathbb{R})} I_{2,\lambda}(\mu,\nu)$. Now, note that for $\mu, \nu \in \mathcal{P}(\mathbb{R})$ such that $\lambda M_2(\mu) + (1-\lambda)M_2(\nu) \le 1$, we have

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

where h is the entropy functional defined in (1.4). Thus,

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

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

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

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

To strengthen the LDP on $\mathcal{P}(\mathbb{R})$ to an LDP on $\mathcal{P}_q(\mathbb{R})$ (i.e., with respect to the q-Wasserstein topology), via an appeal to [14, Corollary 4.2.6] it suffices to show that $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ is exponentially tight in $\mathcal{P}_q(\mathbb{R})$. Since for each $j > 0$, the set $K_{2,j}$ defined in (1.2) is compact in the q-Wasserstein topology by Lemma 1.1. By the definition of

$K_{2,j}$, the definition of M_2 in (1.1), and the identity $\hat{\mu}_A^n \overset{(\mathrm{d})}{=} \hat{\nu}_n$ established in (4.2) and Lemma 4.1, and the definition of $\hat{\nu}_n$ in (4.14),

$$\{\hat{\mu}_{\mathbf{A}}^n \in K_{2,M}^c\} = \{\hat{\nu}_n \in K_{2,M}^c\} = \{M_2(\hat{\nu}_n) > M\} = \left\{\sqrt{\frac{n}{k_n}} \frac{\|\zeta^{(k_n)}\|_2}{\|\zeta^{(n)}\|_2} > \sqrt{M}\right\}. \quad (4.15)$$

When $k_n/n \to \lambda \in (0,1]$, by [3, Lemma 4.2], $\{\|\zeta^{(k_n)}\|_2/\|\zeta^{(n)}\|_2\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n with GRF $\mathbf{J}_{2,\lambda}$ defined in (4.13), and $\{\sqrt{M_2(\hat{\nu}_n)}\}_{n\in\mathbb{N}}$ is exponentially equivalent to $\{\|\zeta^{(k_n)}\|_2/(\sqrt{\lambda}\|\zeta^{(n)}\|_2)\}_{n\in\mathbb{N}}$ at speed n by Remark 1.7. Combining this with (4.15) and Remark 1.6, when $M = 4/\lambda$, we obtain

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

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

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

In this section we prove Theorems 2.11 and 2.12. Recall from (2.4) that $Y_{2,k_n}^n = \|\mathbf{A}_{n,k_n}^\top X^{(n)}\|_2$.

Proof of Theorem 2.11. Fix $\{k_n\}_{n\in\mathbb{N}}$ that grows sublinearly. From Lemma 4.2 and (4.1), we have the following distributional identity,

$$n^{-1/2} Y_{2,k_n}^n \overset{(d)}{=} \frac{\left\|\zeta^{(k_n)}\right\|_2}{\left\|\zeta^{(n)}\right\|_2} \frac{\left\|X^{(n)}\right\|_2}{\sqrt{n}} = \frac{\left\|\zeta^{(k_n)}\right\|_2 / \sqrt{k_n}}{\left\|\zeta^{(n)}\right\|_2 / \sqrt{n}} \frac{\sqrt{k_n}\left\|X^{(n)}\right\|_2}{n}. \quad (5.1)$$

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

$$J_\zeta(x) := \begin{cases} -\frac{1}{2}\log(1-x^2), & x \in [0,1), \\ +\infty, & \text{otherwise.} \end{cases} \quad (5.2)$$

Due to the independence of $\zeta^{(n)}$ and $X^{(n)}$, the result follows by Lemma 1.9.

Now, suppose Assumption C holds with sequence $\{s_n\}_{n\in\mathbb{N}}$, $r \in [0,\infty]$ and GRF $J_X^{(r)}$. We will make repeated use of the following simple observation: by Remark 4.6 and the contraction principle applied to the map $x \mapsto \sqrt{x}$, $\{\|\zeta^{(k_n)}\|_2/\sqrt{k_n}\}_{n\in\mathbb{N}}$ and $\{\|\zeta^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfy LDPs at speed k_n and n, respectively, with the same GRF $J_{\zeta^2}(x^2)$, and hence, $\{\sqrt{n}/\|\zeta^{(n)}\|_2\}_{n\in\mathbb{N}}$ also satisfies an LDP at speed n, and all three sequences converge almost surely to 1 by the strong law of large numbers. We now consider three cases.

Case 1. Suppose $r = 0$. Then $s_n \ll k_n \ll n$. By the case assumption, $\{\sqrt{k_n}\|X^{(n)}\|_2/n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF $J_X^{(0)}$. Hence, the result in this case follows by (5.1), the observation above and a dual application of Lemma 1.9 with (V_n, W_n) therein first being $(\sqrt{k_n}\|X^{(n)}\|_2/n, \|\zeta^{(k_n)}\|_2/\sqrt{k_n})$ and then $(\|\zeta^{(k_n)}\|_2\|X^{(n)}\|_2/n, \sqrt{n}/\|\zeta^{(n)}\|_2)$.

Case 2. Suppose $r \in (0,\infty)$. By Remark 2.10, $\{\sqrt{k_n}\|X^{(n)}\|_2/n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed k_n and with GRF $rJ_X^{(r)}(\sqrt{r}x)$. Then (5.1), the independence of $X^{(n)}$ and $\zeta^{(n)}$, the observation above and Lemma 1.9 together show that $\{n^{-1/2}Y^n_{2,k_n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed k_n with GRF

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

Case 3. Suppose $r = \infty$. Again from the reformulation (5.1), we have

$$n^{-1/2}Y^n_{2,k_n} \overset{(d)}{=} \frac{\left\|\zeta^{(k_n)}\right\|_2/\sqrt{s_n}}{\left\|\zeta^{(n)}\right\|_2/\sqrt{n}} \frac{\sqrt{s_n}\left\|X^{(n)}\right\|_2}{n}. \tag{5.3}$$

Since

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

and (because $r = \infty$) $k_n/s_n \to 0$ as $n \to \infty$, by Lemma 3.2 (with $p = 2$) and the contraction principle applied to $t \mapsto \sqrt{t}$, $\{\|\zeta^{(k_n)}\|_2/\sqrt{s_n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF $\mathcal{J}_\zeta(t) = t^2/2$ if $t \ge 0$, and $\mathcal{J}_\zeta(t) = \infty$ otherwise. By the case assumption, $\{\sqrt{s_n}\|X^{(n)}\|_2/n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF $J_X^{(\infty)}$. The independence of $\{\zeta_i\}_{i\in\mathbb{N}}$ and $X^{(n)}$, together with the contraction principle applied to the product mapping $(x,y) \mapsto xy$, then implies that $\{V_n := \|\zeta^{(k_n)}\|_2\|X^{(n)}\|_2/n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF $\inf_{c>0}\{c^2/2 + J_X^\infty(x/c)\}$. The lemma follows on applying Lemma 1.9 with $\{V_n\}_{n\in\mathbb{N}}$ defined above, $W_n := (\|\zeta^{(n)}\|_2/\sqrt{n})^{-1}$, $n \in \mathbb{N}$, and $m = 1$. □

Proof of Theorem 2.12. Fix $q \in [1,2]$ and suppose k_n grows sublinearly. From Lemma 4.2 and (4.1), we have the following reformulation

$$k_n^{-1/q}Y^n_{q,k_n} \overset{(d)}{=} \frac{\left\|\zeta^{(k_n)}\right\|_q/k_n^{1/q}}{\left\|\zeta^{(n)}\right\|_2/\sqrt{n}} \frac{\left\|X^{(n)}\right\|_2}{\sqrt{n}}, \tag{5.4}$$

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

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

By Assumption A, $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF J_X. By Cramér's theorem and the contraction principle, $\{\|\zeta^{(k_n)}\|_q/k_n^{1/q}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed k_n with GRF $J_{\zeta,q}(x) := \Lambda_q^*(x^q)$ if $x \geq 0$, with Λ_q defined in (2.7), and $J_{\zeta,q}(x) = \infty$ if $x < 0$. Note that Λ_q^* is strictly convex with unique minimizer $\mathcal{M}_q^{1/q}$. Similarly, $\{\|\zeta^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n with GRF $J_{\zeta,2}(x) := \Lambda_2^*(x^2)$ if $x \geq 0$ (and is equal to infinity otherwise) with unique minimizer 1.

To prove the theorem, we will use Lemma 1.9 to show that in each of the three regimes, $\{R_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $s_n \wedge k_n$ with a GRF J_R that we identify in each case. In view of (5.4) and the contraction principle, this would imply that $\{k_n^{-1/q} Y_{q,k_n}^n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $s_n \wedge k_n$ with GRF

$$\widehat{\mathbb{J}}_q^{\mathsf{sub}}(u) := \inf\left\{ J_R(x,y,z) : u = \frac{xz}{y}, x, y, z > 0 \right\}. \tag{5.5}$$

Case 1. Suppose $k_n \ll s_n$. Then we also have $k_n \ll n$, and by Lemma 1.9 and the additional assumption that J_X has a unique minimizer m, $\{R_n\}_{n\in\mathbb{N}}$ is exponentially equivalent at speed k_n to $(\|\zeta^{(k_n)}\|_q/k_n^{1/q}, 1, m)$, which satisfies an LDP at speed k_n with GRF $J_R(x,y,z) = J_{\zeta,q}(x)$ when $y = 1$ and $z = m$, and $J_R(x,y,z) = +\infty$ otherwise. Then (5.5) shows that $\widehat{\mathbb{J}}_q^{\mathsf{sub}}(u) = J_{\zeta,q}(u/m) = \Lambda_q^*(u^q/m^q)$ for $u \geq 0$, and is equal to positive infinity otherwise.

Case 2. Suppose $s_n = k_n$. Since $k_n \ll n$, again invoking Lemma 1.9 and the additional assumption that J_X has a unique minimizer m, we have the exponential equivalence at speed k_n between $\{R_n\}_{n\in\mathbb{N}}$ and $\{(\|\zeta^{(k_n)}\|_q/k_n^{1/q}, 1, \|X^{(n)}\|_2/\sqrt{n})\}_{n\in\mathbb{N}}$. Hence, $\{R_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed k_n with GRF $J_R(x,y,z) = J_{\zeta,q}(x) + J_X(z)$ when $y = 1$, and $J_R(x,y,z) = +\infty$ otherwise. Moreover, (5.5) shows that $\widehat{\mathbb{J}}_q^{\mathsf{sub}}(u) = \inf_{x>0}\{\Lambda_q^*(x^q) + J_X(u/x)\}$ if $u > 0$, (and is equal to $+\infty$ otherwise).

Case 3. Suppose $s_n \ll k_n$. Then $k_n \ll n$ implies $s_n \ll n$, and so Lemma 1.9 and the observation that the GRF $J_{\zeta,q}$ has $M_q^{1/q}$ as its unique minimizer imply that $\{R_n\}_{n\in\mathbb{N}}$ is exponentially equivalent at speed s_n to the sequence $\{(\mathcal{M}_q^{1/q}, 1, \|X^{(n)}\|_2/\sqrt{n})\}_{n\in\mathbb{N}}$, and therefore satisfies an LDP at speed s_n with GRF $J_R(x,y,z) = J_X(z)$ when $x = \mathcal{M}_q^{1/q}$ and $y = 1$, and $J_R(x,y,z) = +\infty$ otherwise. When combined with (5.5) this implies that $\widehat{\mathbb{J}}_q^{\mathsf{sub}}(u) = J_X(u/\mathcal{M}_q^{1/q})$ if $u \geq 0$ (and is equal to $+\infty$ otherwise). □

6. Proofs of q-norm LDPs in the linear regime

The goal of this section is to prove Theorem 2.16. Throughout, fix $\lambda \in (0,1]$, and assume $k_n \sim \lambda n$. Also, for $q \in [1,2]$ and $n, k \in \mathbb{N}, k \leq n$, recall the definition $n^{-1/q} Y_{q,k_n}^n = n^{-1/q}\|\mathbf{A}_{n,k_n}^T X^{(n)}\|_q$ given in (2.4). Section 6.1 contains a simple proof that is valid when $q \in [1,2)$. Section 6.2 is devoted to the more involved case of $q = 2$ in the linear regime, which also then provides an alternative proof and alternative form of the rate function in the case $q \in [1,2)$.

6.1. The case $q \in [1, 2)$

Proof of Theorem 2.16 when $q \in [1, 2)$**.** Fix $q \in [1, 2)$, and observe that with M_q, L^n, Y^n_{q,k_n} and $\mathbf{A}_{n,k_n}$ defined as in (1.1), (2.1), (2.4) and Section 2, respectively, it follows that

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

Then, the LDP for $\{L^n\}_{n\in\mathbb{N}}$ in Theorem 2.15 and the contraction principle applied to the continuous map $\mathcal{P}_q(\mathbb{R}) \ni \nu \mapsto (\lambda M_q)^{1/q}(\nu) \in \mathbb{R}$ imply that $\{\lambda M_q(L^n)\}_{n\in\mathbb{N}}$ satisfies an LDP with GRF

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

Since $\frac{k_n}{n} M_q(L^n)$ is exponentially equivalent to $\lambda M_q(L^n)$ at speed n by Remark 1.7, this implies (by Remark 1.6) that $\{n^{-1/q} Y^n_{q,k_n}\}_{n\in\mathbb{N}}$ also satisfies an LDP at speed n with GRF $\mathbb{J}^{\mathsf{lin}}_{q,\lambda}$.

If $x < 0$, we infimize over an empty set, and so $\mathbb{J}^{\mathsf{lin}}_{q,\lambda}(x) = -\infty$. Now, if $s_n = n$, using the expression for $\mathbb{I}_{L,\lambda}$ given in Remark 4.9, we see that for $x \geq 0$,

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

which coincides with (2.12).

On the other hand, if $s_n \ll n$, using the identity $\mathbb{I}_{L,\lambda}(\mu) = J_X(c)$ if $\mu = \gamma_c$ (and infinity, otherwise) established in Theorem 2.15, we have for $x \geq 0$,

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

where $\mathcal{M}_q$ is as defined in (2.8). This proves (2.12). $\square$

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

We now provide an alternative proof of Theorem 2.16 for $q \in [1, 2)$, which also extends to the case $q = 2$. This yields an alternative representation for the rate function $\mathbb{J}^{\mathsf{lin}}_{q,\lambda}$ of (2.12). To introduce this representation, fix $q \in [1, 2]$ and define the following functions. Let

$$\Lambda_{\mathsf{A},q}(t_1, t_2) := \log \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(t_1|x|^q + (t_2 - \tfrac{1}{2})x^2\right) dx, \quad (t_1, t_2) \in \mathbb{R}^2. \tag{6.1}$$

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

$$\Lambda_{\mathsf{B}}(t_3) := \log \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left((t_3 - \tfrac{1}{2})x^2\right) dx, \quad t_3 \in \mathbb{R}. \tag{6.2}$$

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

$$\mathbf{J}_{q,\lambda}(z) := \inf_{(x_1,x_2,x_3)\in\mathbb{R}^3} \left\{ \lambda \Lambda^*_{\mathsf{A},q}\left(\frac{(x_1, x_2)}{\lambda}\right) + (1-\lambda)\Lambda^*_{\mathsf{B}}\left(\frac{x_3}{1-\lambda}\right) \,:\, z = \frac{x_1^{1/q}}{(x_2 + x_3)^{1/2}} \right\}, \tag{6.3}$$

and define for $\lambda = 1$,

$$\mathbf{J}_{q,1}(z) := \inf_{(x_1,x_2)\in\mathbb{R}^3} \left\{ \Lambda^*_{\mathsf{A},q}\left((x_1, x_2)\right) \,:\, z = \frac{x_1^{1/q}}{(x_2)^{1/2}} \right\}. \tag{6.4}$$

Also, recall the definition of $\mathbf{J}_{2,\lambda}$, $\lambda \in (0, 1]$, from (4.13).

We then have the following result.

Proposition 6.1. *Fix $q \in [1, 2]$. Suppose $\{k_n\}_{n\in\mathbb{N}}$ grows linearly with rate $\lambda \in (0, 1]$, and that Assumption A holds with associated GRF J_X. Then, the sequence of scaled ℓ^n_q norms of random projections, $\{n^{-1/q}Y^n_{q,k_n} = n^{-1/q}\|\mathbf{A}^T_{n,k_n}X^{(n)}\|_q\}_{n\in\mathbb{N}}$, satisfies an LDP at speed n with GRF*

$$\bar{\mathbb{J}}^{\mathsf{lin}}_{q,\lambda}(x) := \inf_{y,z\in\mathbb{R}} \left\{ \mathbf{J}_{q,\lambda}(z) + J_X(y) \,:\, x = yz \right\}, \quad x \in \mathbb{R}_+. \tag{6.5}$$

The proof of Proposition 6.1 relies on an auxiliary result stated in Lemma 6.2, which concerns an LDP related to the top row of the matrix $\mathbf{A}_{n,k_n}$. As in Section 4, let $\zeta_1, \zeta_2, \ldots$ denote a sequence of i.i.d. standard Gaussian random variables, independent of $X^{(n)}$, and let $\zeta^{(n)} := (\zeta_1, \ldots, \zeta_n) \in \mathbb{R}^n$. Due to Lemmas 4.2 and 4.1, we have

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

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

$$n^{-1/q} Y^n_{q,k_n} = n^{-1/q} \|\mathbf{A}^T_{n,k_n} X^{(n)}\|_q \overset{(\mathrm{d})}{=} n^{1/2-1/q} \frac{\|\zeta^{(k_n)}\|_q}{\|\zeta^{(n)}\|_2} \frac{\|X^{(n)}\|_2}{\sqrt{n}}. \tag{6.6}$$

Given (6.6) and Assumption A, a natural step to proving Proposition 6.1 is to establish an LDP for the sequence $\{n^{1/2-1/q} \|\zeta^{(k_n)}\|_q / \|\zeta^{(n)}\|_2\}_{n\in\mathbb{N}}$.

Lemma 6.2. *Suppose $\{k_n\}_{n\in\mathbb{N}}$ grows linearly with rate $\lambda \in [0, 1]$. For $q \in [1, 2]$, the sequence*

$$n^{1/2-1/q} \frac{\|\zeta^{(k_n)}\|_q}{\|\zeta^{(n)}\|_2} = n^{1/2-1/q} \frac{\|(\zeta_1, \dots, \zeta_{k_n})\|_q}{\|(\zeta_1, \dots, \zeta_n)\|_2}, \quad n \in \mathbb{N}, \tag{6.7}$$

satisfies an LDP in $\mathbb{R}$ at speed n with GRF $\mathbf{J}_{q,\lambda}$ of (6.3) and (6.4) for $q \in [1, 2)$ and (4.13) for $q = 2$.

Proof. Fix $q \in [1, 2)$. For $n \in \mathbb{N}$. We start with the observation that the quantity in (6.7) can be represented in terms of a continuous mapping of a vector of scaled i.i.d. sums: for each $n \in \mathbb{N}$,

$$n^{1/2-1/q} \frac{\|\zeta^{(k_n)}\|_q}{\|\zeta^{(n)}\|_2} = T(Z_n), \quad Z_n := \left(\frac{1}{n} \sum_{i=1}^{k_n} |\zeta_i|^q, \frac{1}{n} \sum_{i=1}^{k_n} \zeta_i^2, \frac{1}{n} \sum_{j=k_n+1}^{n} \zeta_j^2 \right), \tag{6.8}$$

where T is the continuous map

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

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

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

Note that for $q \in [1, 2]$, the origin $(0, 0)$ lies in the interior of the domain of $\Lambda_{\mathsf{A},q}$, which is the logarithmic moment generating function of $(|\zeta_1|^q, \zeta_1^2)$. Since the $\{\zeta_i\}_{i\in\mathbb{N}}$ are i.i.d., by Cramér's theorem [14, Theorem 2.2.1] the sequence $\{\mathsf{A}_{n,q}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^2$ at speed k_n with GRF $\Lambda^*_{\mathsf{A},q}(\cdot/\lambda)$, and hence (by Remark 1.4 and the fact that $k_n/n \to \lambda$) also satisfies an LDP in $\mathbb{R}^2$ at speed n with GRF $\lambda\Lambda^*_{\mathsf{A},q}(\cdot/\lambda)$. Similarly, since 0 belongs to the interior of the domain of Λ_{B}, the logarithmic moment generating function of ζ_1^2, the sequence $\{\mathsf{B}_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ at speed n with GRF $(1-\lambda)\Lambda^*_{\mathsf{B}}(\cdot/(1-\lambda))$. Since $k_n/n \to \lambda$ and $(n-k_n)/n \to 1-\lambda$ deterministically, by Remarks 1.6 and 1.7, $\{\frac{1}{n}\sum_{i=1}^{k_n}(|\zeta_i|^q, \zeta_i^2)\}_{n\in\mathbb{N}}$ and $\{\frac{1}{n}\sum_{j=k_n+1}^{n}\zeta_j^2\}_{n\in\mathbb{N}}$ satisfy LDPs at speed n with the same GRFs as $\{\mathsf{A}_{n,q}\}_{n\in\mathbb{N}}$ and $\{\mathsf{B}_n\}_{n\in\mathbb{N}}$, respectively. Combined with the independence of $\mathsf{A}_{n,q}$ and B_n, and Lemma 1.9, it follows that the sequence $\{Z_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^3$ at speed n with GRF

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

and with the GRF equal to $-\infty$ otherwise. Finally, use (6.8), the LDP for $\{Z_n\}_{n\in\mathbb{N}}$ and the contraction principle for the map T defined above to show that the sequence in (6.7) satisfies an LDP with the GRF $\mathbf{J}_{q,\lambda}$ of (6.3).

Next, suppose $\lambda = 1$. We see that the sequence $\{\mathsf{A}_{n,q}\}_{n\in\mathbb{N}}$ is exponentially equivalent to $\{(\frac{1}{n}\sum_{i=1}^{k_n}|\zeta_i|^q, \frac{1}{n}\sum_{i=1}^{n}\zeta^2)\}_{n\in\mathbb{N}}$ by Remark 1.7. An application of the contraction principle to the map $(x_1,x_2) \mapsto x_1^{1/q}/x_2^{1/2}$ yields the LDP for the sequence in (6.7) with speed n and GRF (6.4).

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

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

Noting that Λ_B is the log moment generating function of the chi-squared distribution, it follows from Remark 4.6 that $\Lambda^*_B(x) = J_{\zeta^2}(x) = \frac{1}{2}(x - 1 - \log x)$, for $x > 0$ and infinity, otherwise. Substituting this into the expression for $\mathbf{J}_{2,\lambda}(z)$ in the last display and solving the optimization problem yields (4.13). □

Proof of Proposition 6.1. Recall from Lemma 6.2 that the sequence $\{n^{1/2-1/q}\|\zeta^{(k_n)}\|_q/\|\zeta^{(n)}\|_2\}_{n\in\mathbb{N}}$ satisfies an LDP with GRF $\mathbf{J}_{q,\lambda}$. In addition, Assumption A states that the sequence $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP with GRF J_X. Given the equality in distribution of (6.6), the contraction principle applied to the continuous function $(y,z) \mapsto yz$ yields that the sequence of scaled ℓ_q^n norms of random projections, $\{Y^n_{q,k_n}\}_{n\in\mathbb{N}}$, satisfies an LDP at speed n with GRF $\bar{\mathbb{J}}^{\mathsf{lin}}_{q,\lambda}$ of (6.5). □

Remark 6.3. Note that it follows from Proposition 6.1 and the proof of Theorem 2.16 for $q \in [1,2)$ in Section 6.1 that $\bar{\mathbb{J}}^{\mathsf{lin}}_{q,\lambda} = \mathbb{J}^{\mathsf{lin}}_{q,\lambda}$ for all $q \in [1,2)$.

To complete the proof of Theorem 2.16 we show that this relation also holds for $q = 2$.

Proof of Theorem 2.16 in the case $q = 2$. First, consider the case $s_n = n$. Fix $\lambda \in (0, 1)$. The result for $\lambda = 1$ follows on using the convention that $0 \log 0 = 0 \log 0/0 = 0$ everywhere in the derivation below. By Proposition 6.1, it suffices to show that $\bar{\mathbb{J}}^{\mathsf{lin}}_{2,\lambda} = \mathbb{J}^{\mathsf{lin}}_{2,\lambda}$. In view of (2.12), (6.5), and (4.13), it clearly suffices to show that for $z^2 \in (0, 1)$,

$$\mathbf{J}_{2,\lambda}(z) = \frac{\lambda}{2} \log\left(\frac{\lambda}{z^2}\right) + \frac{1-\lambda}{2} \log\left(\frac{1-\lambda}{1-z^2}\right) = \inf_{\nu \in \mathcal{P}(\mathbb{R}): z^2 = \lambda M_2(\nu)} \mathcal{H}_\lambda(\nu). \tag{6.9}$$

since when $z^2 \notin (0, 1)$ both $\mathbf{J}_{2,\lambda}(z)$ and the right-hand side above are infinite. For $z^2 \in (0, 1)$, substituting the expression for $\mathcal{H}_\lambda$ from (2.10), we obtain

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

Recall that the maximum entropy probability measure constrained to have second moment z^2/λ is the Gaussian measure with mean zero and variance z^2/λ (see, e.g., Ex. 12.2.1 of [12]). Since the entropy of such a Gaussian equals $\frac{1}{2} \log(2\pi e z^2/\lambda)$, we have

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

which proves (6.9). This completes the proof of the equality $\bar{\mathbb{J}}^{\mathsf{lin}}_{2,\lambda} = \mathbb{J}^{\mathsf{lin}}_{2,\lambda}$ in this case.

Now, suppose $s_n \ll n$. By (6.6), it follows that $n^{-1/2} Y^n_{2,k_n} \overset{(d)}{=} V_n W_n W'_n$, where $V_n = \|X^{(n)}\|_2/\sqrt{n}$, $W_n := \sqrt{k_n/n}\|\zeta^{(k_n)}\|_2/\sqrt{k_n}$ and $W'_n = \sqrt{n}/\|\zeta^{(n)}\|_2$. Noting that $k_n/n \to \lambda$ as $n \to \infty$, we see that $\{W_n\}_{n\in\mathbb{N}}$ and $\{W'_n\}_{n\in\mathbb{N}}$ converge almost surely to $(\lambda \mathcal{M}_2)^{1/2}$ and 1, respectively (by the strong law of large numbers), and satisfy LDPs at speed n (by Remark 4.6 and the contraction principle). Since by assumption, $\{V_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed $s_n \ll n$ with GRF J_X, by Lemma 1.9 the sequence $\{V'_n := V_n W'_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF J_X. Since the GRF of $\{W'_n\}_{n\in\mathbb{N}}$ has a unique minimum at $m := (\lambda \mathcal{M}_2)^{1/2}$, by Lemma 1.9 $\{V'_n W_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed s_n with GRF $J_X(\cdot/m)$, which coincides with the expression in (2.12). □

Remark 6.4. Recall that in the sublinear/linear regime, when $q = 2$ the contraction principle cannot be applied because the LDP of Theorem 2.15 holds with respect to the q-Wasserstein topology only for $q \in [1, 2)$. Moreover, the LDP for $\{\hat{\mu}^n_{\mathbf{A}}\}_{n\in\mathbb{N}}$ of (4.2) as stated in Proposition 4.8 also holds with respect to the q-Wasserstein topology only for $q < 2$. However, in the case $q = 2$, (6.9) still establishes a variational problem that explicitly relates the rate function $\mathbf{J}_{2,\lambda}$ of (4.13) to the rate function $\mathcal{H}_\lambda$ of (2.10), in the same manner as if the contraction principle were applicable. We claim that this inconvenient gap is related to a more fundamental obstacle. To illustrate this concretely, consider the case where $X_1, X_2, \dots$ are i.i.d. exponential random variables with mean 1. It is known that:

1. Due to Sanov's theorem, the sequence of empirical measures $\{\frac{1}{n}\sum_{i=1}^n \delta_{X_i}\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\mathbb{R})$ with GRF $H(\cdot\|\operatorname{Exp}(1))$, where we write Exp(1) to denote the exponential distribution with mean 1.
2. Due to Cramér's theorem [14, Theorem 2.2.1], the sequence of empirical means $\{\frac{1}{n}\sum_{i=1}^n X_i\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ with GRF
$$L(\beta) := \beta - \log\beta - 1.$$
3. An explicit calculation establishes the expression
$$L(\beta) = \inf_{\mu\in\mathcal{P}(\mathbb{R})} \left\{ H(\mu\|\operatorname{Exp}(1)) : \int_{\mathbb{R}} x\,d\mu = \beta \right\}.$$

Note that if the map $\mu \mapsto \int_{\mathbb{R}} x\,d\mu$ were continuous, then the LDP of point 2. above would follow from point 1., point 3., and an application of the contraction principle. However, the map $\mu \mapsto \int_{\mathbb{R}} x\,d\mu$ is *not* continuous with respect to the weak topology on probability measures. Moreover, the result of [55, Theorem 1.1] applied to the exponential distribution indicates that the LDP of point 1. *does not* hold with respect to the 1-Wasserstein topology. This suggests that the apparently cryptic transition at $q = 2$ in the nature of the proof of Theorem 2.16 and the result of Theorem 2.15 is in a sense a manifestation of a more common sticking point in large deviations theory. In other words, even in the simple setting of i.i.d. random variables, the continuity required by the contraction principle fails to hold, but the consequences (i.e., a large deviation principle and a variational formula for the rate function) *do* still hold in many instances.

Acknowledgments

The authors would like to thank Keith Ball for a question that led to Remark 2.5, and thank Joscha Prochno for feedback on the first version of this paper, which helped improve the exposition. This paper is comprised of some results of the PhD thesis of the first author [32] under the supervision of the third author, namely the results on LDPs for the projection (or their empirical measures) in all regimes, and norms of the projections in the constant and linear regimes under Assumption A*. All other results contained in the paper, including the formulation of Assumptions A, B and C, proofs of related theorems, as well as the verification of these assumptions for ℓ_p^n balls, $p \in [1, 2)$ and Orlicz balls, will be a part of the second author's PhD thesis, under the supervision of the third author. We would also like to thank anonymous referees for many suggestions that greatly improved the exposition of the paper.

Appendix A. An approximate contraction principle

In this appendix, we recall the approximate contraction principle established in Section 6.2 of [10], and establish a corollary of it that we use in the proof of the LDP in the linear regime in Section 4.4. Recall that given $a, b \in \mathbb{R}$, we will use $a \vee b$ and $a \wedge b$ to denote $\max(a, b)$ and $\min(a, b)$, respectively.

Let Σ be a Polish space. Let $\mathcal{X}$ be a separable Banach space with topological dual space $\mathcal{X}^*$, and let $\langle \cdot, \cdot \rangle : \mathcal{X}^* \times \mathcal{X} \to \mathbb{R}$ denote the associated dual pairing. Fix a continuous map $\mathbf{c} : \Sigma \to \mathcal{X}$, and let $\{\mathscr{L}_n\}_{n\in\mathbb{N}}$ be a sequence of $\mathcal{P}(\Sigma)$-valued random elements. For $r \in (0, \infty]$ and a continuous function $W : \Sigma \to \mathbb{R}$ such that $\mathbb{P}$-a.s., $\int_\Sigma (W(x) \vee 0) \mathscr{L}_n(dx) < \infty$ for all $n \in \mathbb{N}$, let

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

and also let

$$\bar{\Lambda}(W) := \sup_{r>0} \Lambda_r(W). \tag{A.2}$$

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

$$\begin{aligned} \mathcal{D} &:= \{\alpha \in \mathcal{X}^* : \bar{\Lambda}(\langle \alpha, \mathbf{c}(\cdot)\rangle) < \infty\}, \\ \mathcal{D}_\circ &:= \{\alpha \in \mathcal{X}^* : \exists p > 1, p\alpha \in \mathcal{D}\}. \end{aligned} \tag{A.3}$$

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

$$F(x) := \sup_{\alpha\in\mathcal{D}_\circ} \langle \alpha, x \rangle. \tag{A.4}$$

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

Proposition A.1 *(Proposition 6.4 of [10]). Suppose that:*

1. *$\{\mathscr{L}_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\Sigma)$ at speed n with GRF I_0;*
2. *$\{\mathscr{L}_n, \mathscr{C}_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\Sigma) \times \mathcal{X}$ at speed n with some convex GRF $\mathbb{I}$;*
3. *for any sequence $\{W_n\}_{n\in\mathbb{N}}$, in the set*

$$\{V + \langle \alpha, \mathbf{c}(\cdot)\rangle : V : \Sigma \to \mathbb{R} \text{ continuous and bounded}, \alpha \in \mathcal{D}_\circ\} \tag{A.5}$$

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

$$\limsup_{n\to\infty} \bar{\Lambda}(W_n) \le \bar{\Lambda}(W_\infty). \tag{A.6}$$

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

$$\mathbb{I}(\mu, s) := \begin{cases} I_0(\mu) + F\left(s - \int_\Sigma \mathbf{c}\, d\mu\right) & \text{if } I_0(\mu) < \infty, \\ +\infty & \text{else,} \end{cases} \tag{A.7}$$

with F *as defined in* (A.4)*.*

The following corollary considers a special case where the conditions of Proposition A.1 can be easily verified.

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

$$\widehat{\Lambda}(W) := \log \mathbb{E}[e^{\lambda^{-1} W(\eta_1)}] \tag{A.8}$$

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

$$\mathscr{D} := \left\{\alpha \in \mathcal{X}^* : \widehat{\Lambda}(\langle \alpha, \mathbf{c}(\cdot)\rangle) < \infty\right\}, \tag{A.9}$$

and let $\mathscr{C}_n := \int_\Sigma \mathbf{c}(x)\mathscr{L}_n(dx)$*. Then* $\{\mathscr{L}_n, \mathscr{C}_n\}$ *satisfies an LDP with GRF given by* (A.7)*, where* $I_0(\nu) := \lambda H(\nu|\mu)$ *and* $F(x) = \sup_{\alpha\in\mathscr{D}^\circ} \langle \alpha, x\rangle$*.*

Proof. We start by verifying the conditions of Proposition A.1.

The fact that $\{\mathscr{L}_n\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\Sigma)$ at speed k_n with GRF $H(\cdot|\mu)$ follows from Sanov's theorem on the Polish space $\mathcal{P}(\Sigma)$ (see, e.g., Theorem 6.6.9 of [30]). Since $k_n/n \to \lambda$, this immediately implies that $\{\mathscr{L}_n\}_{n\in\mathbb{N}}$ satisfies an LDP at speed n with GRF $I_0(\cdot) := \lambda H(\cdot|\mu)$, so condition 1. of Proposition A.1 is satisfied.

Next, note that $(\mathscr{L}_n, \mathscr{C}_n) = \frac{1}{k_n}\sum_{j=1}^{k_n}(\delta_{\eta_j}, \mathbf{c}(\eta_j)) \in \mathcal{P}(\Sigma)\times\mathcal{X}$. Therefore, it follows from Cramér's theorem on any locally convex topological space (see, e.g., Theorem 6.1.3 and Corollary 6.16 of [14]), with an appeal to the assumption that 0 lies in $\mathscr{D}^\circ$, the interior of the set $\mathscr{D}$ of (A.9), that $\{(\mathscr{L}_n, \mathscr{C}_n)\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\Sigma)\times\mathcal{X}$ at speed k_n with a convex GRF. Since $k_n/n \to \lambda \in (0, 1]$, condition 2. of Proposition A.1 is satisfied.

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

$$\lim_{n\to\infty} \widehat{\Lambda}(W_n) = \widehat{\Lambda}(W_\infty). \tag{A.10}$$

To complete the proof, it clearly suffices to show that $\mathscr{D} = \mathcal{D}$, for the domains $\mathscr{D}$ and $\mathcal{D}$ defined in (A.9) and (A.3), respectively, and that relation (A.10) holds when $\widehat{\Lambda}$ is replaced with $\bar{\Lambda}$. In turn, to show the latter, it suffices to prove that $\bar{\Lambda} = \lambda\widehat{\Lambda}$. Indeed, note that for any continuous $W : \Sigma \to \mathbb{R}$ such that $\mathbb{P}$-a.s., $\int_\Sigma (W(x) \vee 0)\mathscr{L}_n(dx) < \infty$ for all $n \in \mathbb{N}$, by the definitions of Λ_∞ and $\bar{\Lambda}$ from (A.1) and (A.2), respectively, the i.i.d. assumption on $\{\eta_i\}_{i=1}^{k_n}$ and the fact that $k_n/n \to \lambda$, we have

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

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

The corollary then follows from Proposition A.1 and the identity $\mathscr{D}^\circ = \mathfrak{D}_0$. □

Appendix B. LDP for generalized normal random variables

The aim of this section is to prove Lemma 3.2, which concerns an LDP for weighted sums of stretched exponential random variables.

Proof. Fix $p \in [1, 2)$, $t > 0$, and denote $\xi := \xi_i^{(p)}$. Define $m := \mathbb{E}[\xi_1^2]$ and for $n \in \mathbb{N}$ define $S_n := \sum_{i=1}^{k_n}(\xi_i^2 - m)$. Since $k_n/b_n \to 0$ as $n \to \infty$, by Theorem 3 in [46], with n, x_n therein and ε replaced by k_n, $b_n t$ and $1 - p/2$, we see that

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

Hence, by (3.1) and the convergence $k_n/b_n \to 0$ as $n \to \infty$,

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

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

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

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

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

where the second equality follows from Stirling's approximation.

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

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

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

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

which completes the proof of the upper bound.

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

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{b_n^{p/2}} \log \mathbb{P}(T_n \in U) \\
&\quad \ge \liminf_{n\to\infty} \frac{1}{b_n^{p/2}} \log \mathbb{P}(S_n \in (\beta - \varepsilon, \beta + \varepsilon - \tau)) \\
&\quad = \liminf_{n\to\infty} \frac{1}{b_n^{p/2}} \log \left[\mathbb{P}(S_n \in (\beta - \varepsilon, \infty)) - \mathbb{P}(S_n \in [\beta + \varepsilon - \tau, \infty))\right]
\end{aligned}$$

$$= \liminf_{n\to\infty} \frac{1}{b_n^{p/2}} \log \mathbb{P}(S_n \in (\beta - \varepsilon, \infty))$$
$$= -\mathcal{J}_{\xi,p}(\beta - \varepsilon)$$
$$\geq - \inf_{x\in U} \mathcal{J}_{\xi,p}(x) - \delta,$$

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

Appendix C. Properties of the rate function $\mathcal{J}$ of (3.14)

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

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

Note that when f is differentiable at x, the subdifferential $\partial f(x)$ has only one element, which is the gradient $\nabla f(x)$. For further discussion of subdifferentials, the reader is referred to Chapter X in [23].

Proof of Lemma 3.12. Before proving the individual properties, we make a general observation. Recall that D_V is the domain of V. For u, $v \in \mathbb{R}_+$, define $H : \mathbb{R}_- \times \mathbb{R} \to \mathbb{R}$ to be

$$H(s,t) := su + tv - \log \int\limits_{D_V} e^{sV(x)+tx^2} dx. \tag{C.1}$$

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

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

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

We now turn to the proof of property 1. Let $(u, v) \in \mathbb{R}_+^2$ be in the interior of the domain of $\mathcal{J}$. Then by [23, Theorem 1.4.2], since $\mathcal{J}$ is convex, the subdifferential at (u, v) is nonempty, $\partial\mathcal{J}(u, v) \neq \varnothing$. Therefore there exists $s \in \mathbb{R}_-$, $t \in \mathbb{R}$ such that $(s, t) \in \partial\mathcal{J}(u, v)$. Since $\mathcal{J}$ and $(s, t) \mapsto \log \int_{D_V} e^{sV(x)+tx^2} dx$ are both convex, both functions satisfy the condition in [23, Theorem 1.4.1]. Hence, we see that there exist $s \in \mathbb{R}_-$ and $t \in \mathbb{R}$ such that $\mathcal{J}(u, v) = H(s, t)$. If H is strictly concave, then automatically (s, t) is the unique maximizer in the supremum of $\mathcal{J}(u, v)$ in (3.14). If H is not strictly concave, then $V \equiv 0$ in its domain, and so by (3.14), $\mathcal{J}(v) = \sup_{s<0, t\in\mathbb{R}_+} \left\{su + tv - \log(\int_{D_V} e^{tx^2} dx)\right\}$, from which it clearly follows that $(0, t)$ is the unique maximizer (since $v \geq 0$). Next, again by the last display and [23, Theorem 1.4.1], we see that $(u, v) \in \partial \log \int_{D_V} e^{sV(x)+tx^2} dx$. By the definition of a subdifferential, we conclude that

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

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

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

To see why property 2 holds, note that by the duality of the Legendre transform [56, Equation (12)], the minimizer of $\mathcal{J}(1, \cdot)$ is obtained at m such that

$$\frac{d}{dv}\mathcal{J}(1, v)\bigg|_{v=m} = t(m) = 0. \tag{C.3}$$

Substituting this relation back into (C.2), we obtain $1 = M_V(\nu_{s(m),0})$ and $m = M_2(\nu_{s(m),0})$. Setting $b^* = -s(m) > 0$ and observing that $\nu_{s,0} = \mu_{V,-s}$ for any s, we conclude that $1 = M_V(\nu_{s(m),0}) = M_V(\mu_{V,b^*})$ and $m = M_2(\nu_{s(m),0}) = M_2(\mu_{V,b^*})$. Now, (3.16) follows since the supremum on the right-hand side of (3.16) is attained at $1 = M_V(\nu_{s(m),0})$ and (C.2) is uniquely solvable. This proves property 2.

For the remaining properties, first note that by the duality of the Legendre transform, the supremum in the definition (3.14) of $\mathcal{J}(u, v)$, when finite, is attained at $(\partial_u\mathcal{J}(u, v), \partial_v\mathcal{J}(u, v)) \in \mathbb{R}_- \times \mathbb{R}$. This proves property 4. By the convexity of $v \mapsto \mathcal{J}(1, v)$ and the fact that it uniquely attains its minimum at the value m defined above, we see that $\partial_v\mathcal{J}(1, v) > 0$ if $v > m$ and $\partial_v\mathcal{J}(1, v) < 0$ if $0 < v < m$. This proves property 3, and completes the proof of the lemma. $\square$

References

[1] D. Alonso-Gutiérrez, J. Bastero, Approaching the Kannan-Lovász-Simonovits and Variance Conjectures, vol. 2131, Springer, 2015.

[2] D. Alonso-Gutiérrez, J. Prochno, Thin-shell concentration for random vectors in Orlicz balls via moderate deviations and Gibbs measures, J. Funct. Anal. (2021) 109291.
[3] D. Alonso-Gutiérrez, J. Prochno, C. Thäle, Large deviations for high-dimensional random projections of ℓ_p^n-balls, Adv. Appl. Math. 99 (2018) 1–35.
[4] D. Alonso-Gutiérrez, J. Prochno, C. Thäle, Large deviations, moderate deviations, and the KLS conjecture, J. Funct. Anal. 280 (1) (2021) 108779.
[5] B. Anderson, J. Jackson, M. Sitharam, Descartes' rule of signs revisited, Am. Math. Mon. 105 (5) (1998) 447–451.
[6] M. Anttila, K. Ball, I. Perissinaki, The central limit problem for convex bodies, Trans. Am. Math. Soc. 355 (12) (2003) 4723–4735.
[7] R. Bahadur, R. s Rao, On deviations of the sample mean, Ann. Math. Stat. 31 (4) (1960) 1015–1027.
[8] F. Barthe, F. Gamboa, L. Lozada-Chang, A. Rouault, Generalized Dirichlet distributions on the ball and moments, ALEA Lat. Am. J. Probab. Math. Stat. 7 (2010) 319–340.
[9] F. Barthe, O. Guédon, S. Mendelson, A. Naor, A probabilistic approach to the geometry of the ℓ_p^n-ball, Ann. Probab. 33 (2) (2005) 480–513.
[10] G. Ben Arous, A. Dembo, A. Guionnet, Aging of spherical spin glasses, Probab. Theory Relat. Fields 120 (1) (2001) 1–67.
[11] U. Brehm, J. Voigt, Asymptotics of cross sections for convex bodies, Beitr. Algebra Geom. 41 (2) (2000) 437–454.
[12] T.M. Cover, J.A. Thomas, Elements of Information Theory, 2nd ed., Wiley, 2001.
[13] I. Csiszár, Sanov property, generalized I-projection and a conditional limit theorem, Ann. Probab. (1984) 768–793.
[14] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, vol. 38, 2nd ed., Springer Science & Business Media, 2009.
[15] P. Diaconis, D. Freedman, Asymptotics of graphical projection pursuit, Ann. Stat. 12 (3) (1984) 793–815.
[16] I.H. Dinwoodie, S.L. Zabell, Large deviations for exchangeable random vectors, Ann. Probab. 20 (3) (1992) 1147–1166.
[17] P. Dupuis, V. Laschos, 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).
[18] J. Freedman, J.W.T. Tukey, A projection pursuit algorithm for exploratory data analysis, IEEE Trans. Comput. 9 (1974) 881–890.
[19] D.J. Fresen, A simplified proof of CLT for convex bodies, Proc. Am. Math. Soc. 149 (1) (2021) 345–354.
[20] N. Gantert, S.S. Kim, K. Ramanan, Cramér's theorem is atypical, Adv. Math. Sci. (2016) 253–270, Research from the 2015 Association for Women in Mathematics Symposium.
[21] N. Gantert, S.S. Kim, K. Ramanan, Large deviations for random projections of ℓ^p balls, Ann. Probab. 45 (2017) 4419–4476.
[22] O. Guédon, E. Milman, Interpolating thin-shell and sharp large-deviation estimates for lsotropic log-concave measures, Geom. Funct. Anal. 21 (5) (2011) 1043.
[23] J. Hiriart-Urruty, C. Lemaréchal, Convex Analysis and Minimization Algorithms II, vol. 305, Springer Science & Business Media, 2013.
[24] P. Indyk, R. Motwani, Approximate nearest neighbors: towards removing the curse of dimensionality, in: STOC '98, Dallas, TX, 1999, pp. 604–613.
[25] W.B. Johnson, J. Lindenstrauss, Extensions of Lipschitz mappings into a Hilbert space, in: Conference in Modern Analysis and Probability, New Haven, Conn., 1982, 1984, pp. 189–206.
[26] S. Johnston, J. Prochno, A Maxwell principle for generalized Orlicz balls, arXiv:e-prints, available at arXiv:2012.11568, 2020.
[27] Z. Kabluchko, J. Prochno, The maximum entropy principle and volumetric properties of Orlicz balls, J. Appl. Math. Anal. Appl. 495 (1) (2021) 124687.
[28] Z. Kabluchko, J. Prochno, C. Thäle, High-dimensional limit theorems for random vectors in ℓ_p^n-balls, Commun. Contemp. Math. 21 (1) (2019) 1750092.
[29] Z. Kabluchko, J. Prochno, C. Thäle, High-dimensional limit theorems for random vectors in ℓ_p^n-balls. II, Commun. Contemp. Math. 23 (3) (2021) 1950073.
[30] D. Kannan, V. Lakshmikanthan, Handbook of Stochastic Analysis and Applications, First, Statistics: a Series of Textbooks and Monographs, CRC Press, 2001.
[31] R. Kannan, L. Lovász, M. Simonovits, Isoperimetric problems for convex bodies and a localization lemma, Discrete Comput. Geom. 13 (3–4) (1995) 541–559.

[32] S.S. Kim, Problems at the interface of probability and convex geometry: Random projections and constrained processes, Ph.D. Thesis, 2017.
[33] S.S. Kim, K. Ramanan, A conditional limit theorem for high-dimensional ℓ^p-spheres, J. Appl. Probab. 55 (2018) 1060–1077.
[34] B. Klartag, A central limit theorem for convex sets, Invent. Math. 168 (2007) 91–131.
[35] B. Klartag, Power-law estimates for the central limit theorem for convex sets, J. Funct. Anal. 245 (1) (2007) 284–310.
[36] A. Klenke, Probability Theory: A Comprehensive Course, Springer Science & Business Media, 2013.
[37] A.V. Kolesnikov, E. Milman, et al., The KLS isoperimetric conjecture for generalized Orlicz balls, Ann. Probab. 46 (6) (2018) 3578–3615.
[38] Y.-T. Liao, K. Ramanan, Quenched sharp large deviations for multidimensional random projections of high-dimensional measures, preprint, 2021.
[39] Y.-T. Liao, K. Ramanan, Sharp large deviation estimates and their applications to asymptotic convex geometry, preprint, 2021.
[40] Y.-T. Liao, K. Ramanan, Geometric sharp large deviations for random projections of ℓ_p^n spheres, arXiv:e-prints arXiv:2001.04053v3, 2020.
[41] W. Liu, L. Wu, Large deviations for empirical measures of mean-field Gibbs measures, Stoch. Process. Appl. 130 (2) (2020) 503–520.
[42] O. Maillard, R. Munos, Compressed least-squares regression, Adv. Neural Inf. Process. Syst. 22 (2009) 1213–1221.
[43] E. Meckes, Projections of probability distributions: a measure-theoretic Dvoretzky theorem, in: Geometric Aspects of Functional Analysis, 2012, pp. 317–326.
[44] M.E. Muller, A note on a method for generating points uniformly on n-dimensional spheres, Commun. ACM 2 (4) (1959) 19–20.
[45] J. Musielak, Orlicz Spaces and Modular Spaces, Lecture Notes in Mathematics, vol. 1034, Springer, 1983.
[46] A.V. Nagaev, Integral limit theorems taking large deviations into account when Cramér's condition does not hold. I, Theory Probab. Appl. 14 (1) (1969) 51–64.
[47] G. Paouris, Concentration of mass on convex bodies, Geom. Funct. Anal. 16 (5) (2006) 1021–1049.
[48] C.H. Papadimitriou, P. Raghavan, H. Tamaki, S. Vempala, Latent semantic indexing: a probabilistic analysis, J. Comput. Syst. Sci. 61 (2) (2000) 217–235, Special issue on the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (Seattle, WA, 1998), MR1802556.
[49] M. Pilipczuk, J.O. Wojtaszczyk, The negative association property for the absolute values of random variables equidistributed on a generalized Orlicz ball, Positivity 12 (3) (2008) 421–474.
[50] S.T. Rachev, L. Rüschendorf, Approximate independence of distributions on spheres and their stability properties, Ann. Probab. 19 (3) (1991) 1311–1337.
[51] G. Schechtman, J. Zinn, On the volume of the intersection of two L_p^n balls, Proc. Am. Math. Soc. 110 (1) (1990) 217–224.
[52] V.N. Sudakov, Typical distributions of linear functionals in finite-dimensional spaces of high dimension, Dokl. Akad. Nauk SSSR 243 (6) (1978) 1402–1405.
[53] C. Villani, Optimal Transport: Old and New, vol. 338, Springer Science & Business Media, 2008.
[54] H. von Weizsäcker, Sudakov's typical marginals, random linear functionals and a conditional central limit theorem, Probab. Theory Relat. Fields 107 (3) (1997) 313–324.
[55] R. Wang, X. Wang, L. Wu, Sanov's theorem in the Wasserstein distance: a necessary and sufficient condition, Stat. Probab. Lett. 80 (5) (2010) 505–512.
[56] R. Zia, E. Redish, S. McKay, Making sense of the Legendre transform, Am. J. Phys. 77 (7) (2009) 614–622.