QUENCHED LARGE DEVIATION PRINCIPLES FOR PROJECTIONS OF ℓ_p^n BALLS

PATRICK LOPATTO, KAVITA RAMANAN, AND XIAOYU XIE

Contents

1. Introduction

High-dimensional measures are ubiquitous in mathematics, and are often profitably studied through their lower-dimensional projections. This approach has been successfully applied to problems in numerous fields, including statistics [9, 10, 23], asymptotic functional analysis [14] and convex geometry [3, 20]. In the case of the uniform measure on a high-dimensional convex body (a compact convex set with non-empty interior), low-dimensional projections are known to satisfy a central limit theorem. This theorem states that most k-dimensional projections of an n-dimensional isotropic convex body are approximately Gaussian in total variation norm, if n is sufficiently large and k is sufficiently small relative to n [19]. The typical behavior of a low-dimensional projection is therefore uninformative about the high-dimensional convex body from which it originated. However, it was recently discovered that the tail behavior of these projections retains interesting information about the original measure. This tail behavior is quantified through large deviations principles (LDPs), of which there are two main types: *quenched* LDPs, which consider a deterministic projection or sequence of projections (independent from the projected measure), and *annealed* LDPs, which consider an average over the direction of the projection. In this article, we focus on quenched LDPs for projections of the unit ball in ℓ_p^n, one of the most fundamental examples of a convex body. Denoting this ball by $\mathbb{B}_p^n$, we have

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

We begin by reviewing previous work on LDPs for projections of $\mathbb{B}_p^n$, starting with the annealed case. Annealed LDPs are typically easier to analyze than quenched LDPs, since averaging over the projection renders the entries of the projected vector exchangeable. For $\mathbb{B}_p^n$, annealed LDPs for one-dimensional projections were first shown in [12] for all $p \in [1, \infty]$. Later, annealed LDPs for the ℓ_2 norm of k_n-dimensional projections were established in [1] for all $p \in [1, \infty]$ when $\lim_{n\to\infty} k_n/n = \lambda \in [0, 1]$, with the additional requirement that $\lambda > 0$ when $p \leq 2$. The restriction that $\lambda > 0$ was removed in [17], which also proved a phase transition in the speed of the LDP as k_n varies (see [17, Remark 3.6]). Further, [17] established LDPs for the empirical measures of the projections, in addition to the ℓ_2 norm. This work also goes beyond the balls $\mathbb{B}_p^n$ and establishes results for general high-dimensional measures satisfying an asymptotic thin shell condition, including uniform measures on Orlicz balls and certain Gibbs measures.

Compared to the annealed case, much less is known about quenched LDPs. Previous works have focused exclusively on k-dimensional projections for k independent of n. For one-dimensional projections of $\mathbb{B}_p^n$, quenched LDPs were established for all $p \in [1, \infty)$ in [12]. In the case $p \in (1, \infty)$, the speed and rate function of the LDP are insensitive to the choice of projection, except for a measure zero set of so-called *atypical* projections, which was studied in [11]. Later, *sharp* LDPs with an optimal constant prefactor were obtained for one-dimensional projections of $\mathbb{B}_p^n$ in [22]. Quenched LDPs for k-dimensional projections were established for $p \geq 2$ in [18], for any positive integer k. As in the one-dimensional case, these are almost surely independent of the choice of projection. Quenched LDPs for projections of various radially symmetric measure on $\mathbb{B}_p^n$ were recently studied in [16]. We also remark that LDPs for the ℓ_q norm of a random element of $\mathbb{B}_p^n$ (with $q \neq p$) were shown in [1, 15]

In this work, we prove a quenched LDP for sequences of k_n-dimensional projections of ℓ_p^n balls when k_n grows in n. Our main result is an LDP for all $p \in [2, \infty)$ and all sequences $(k_n)_{n=1}^\infty$ such that $\lim_{n\to\infty} k_n = \infty$ and $\lim_{n\to\infty} k_n/n = 0$ (meaning k_n grows sublinearly). As in previous work, our LDP is almost surely insensitive to the sequence of projections. We show that our LDP also implies LDPs for the sequences of ℓ_2 norms and ℓ_∞ norms of the projections. The extension to LDPs for the ℓ_q norms for all $q \in [1, \infty)$ is also possible using our methods; see Remark 1.6 below. Our results generalize to projections of the ℓ_p^n sphere and a broad class of product measures.

There are two primary technical challenges involved in establishing our results. First, LDPs are defined relative to a sequence of measures on a single probability space, but each element of a sequence of k_n-dimensional projections has a different codomain. To remedy this problem, each projection must be embedded into a suitable parent space. In the annealed case, the empirical measure completely determines the distribution of projected vector, due to the exchangeability of the coordinates. Previous work on annealed LDPs for projections of growing dimension encoded these projections as probability measures on $\mathbb{R}$, and studied their large deviations [17]. However, this device is not available in the quenched case. Instead, we take inspiration from the work of Comets and Dembo on large deviations for mean-field spin glass models [7], and embed our projections into a space consisting of infinite sequences whose coordinates appear in descending order (in absolute value) paired with their ℓ_2 norms. While [7] worked under the assumption of bounded spins, no similar boundedness hypothesis is made in this work, which creates some additional technical complications. However, this approach captures a great deal of information about the sequence of projections, allowing us to also prove the additional LDPs for the ℓ_∞ and ℓ_2 norms mentioned previously. We remark that the previous works [17, 18] on multidimensional projections did not address LDPs for the ℓ_∞ norm.

The second technical obstacle is that, in the course of the proof, we must show that a projection matrix sampled uniformly from the appropriate Haar measure can be well approximated by a matrix of independent Gaussians, even on events of exponentially small size. This claim is made precise in Lemma 3.6 below. Previously, in [18, Corollary 2.11], a similar Gaussian approximation result was obtained by first establishing an LDP for random projections in the space of probability measures over $\mathbb{R}^k$ for fixed $k \in \mathbb{N}$. However, such a statement is not available if k_n grows in n. Instead, we use the Weingarten calculus (see Appendix B) to calculate the joint moments of random projection matrix entries in an asymptotically precise way, which allows us to deduce the necessary Gaussian approximation estimates using a standard moment convergence argument.

1.1. **Definitions.** For $k \in \mathbb{N}$, let I_k be the $k \times k$ identity matrix, and for $n \geq k$, set

$$\mathbb{V}_{n,k} := \left\{ A \in \mathbb{R}^{n\times k} : A^\mathsf{T} A = I_k \right\}. \tag{1.1}$$

The set $\mathbb{V}_{n,k}$ is called the Stiefel manifold of k-frames in $\mathbb{R}^n$ and consists of k-dimensional orthonormal bases in $\mathbb{R}^n$.

Let ℓ^2 denote the set of infinite sequences $\mathbf{x} = (x_1, x_2, \dots) \in \mathbb{R}^\infty$ of real numbers such that the norm

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^\infty x_i^2 \right)^{1/2} \tag{1.2}$$

is finite. We say that $\mathbf{x} \in \ell^2$ is ordered if $|x_k| \geq |x_{k+1}|$, and $x_k \geq x_{k+1}$ if $|x_k| = |x_{k+1}|$, for all $k \in \mathbb{N}$. Given a sequence $\mathbf{x} \in \ell^2$ with a finite number of nonzero entries, let $\mathfrak{n}(\mathbf{x}) := |\{i \in \mathbb{N} : x_i \neq 0\}|$ denote the number of such entries. Let $[\mathbf{x}] \in \ell^2$ denote the sequence whose first $\mathfrak{n}(\mathbf{x})$ coordinates $[x]_1, \dots, [x]_{\mathfrak{n}(\mathbf{x})}$ are equal to the nonzero entries of $\mathbf{x}$ arranged so that $[\mathbf{x}]$ is ordered, and $[x]_i = 0$ for $i \geq \mathfrak{n}(\mathbf{x}) + 1$. Given $k \in \mathbb{N}$ and $\mathbf{x} \in \mathbb{R}^k$, we define $[\mathbf{x}]$ by first defining $\tilde{\mathbf{x}} \in \ell^2$ by $\tilde{\mathbf{x}}_i = \mathbf{x}_i$ for $i \leq k$ and $\mathbf{x}_i = 0$ for $i > k$, then setting $[\mathbf{x}] = [\tilde{\mathbf{x}}]$.

We define $\mathcal{X}$ as the set of pairs

$$\mathcal{X} := \{(\mathbf{x}, r) : \mathbf{x} \in \ell^2,\, r \in [0, \infty),\, \mathbf{x} \text{ is ordered},\, \|\mathbf{x}\|_2 \leq r\}. \tag{1.3}$$

We equip $\mathcal{X}$ with the distance

$$d\big((\mathbf{x},r),(\mathbf{y},s)\big) \coloneqq \|\mathbf{x}-\mathbf{y}\|_\infty + |r-s|. \tag{1.4}$$

For $\mathbf{x} \in \mathbb{R}^k$, let

$$\Pi(\mathbf{x}) = ([\mathbf{x}], \|\mathbf{x}\|_2) \in \mathcal{X} \tag{1.5}$$

denote the corresponding representative in $\mathcal{X}$.

The function $\log(x)$ denotes the natural logarithm. We let $\mathbb{R}_+$ denote the non-negative reals. We define the ℓ_n^p ball as the set

$$\mathbb{B}_p^n = \left\{ \mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n |x_i|^p \le n \right\} \tag{1.6}$$

and the ℓ_n^p sphere by

$$\mathbb{S}_p^n = \left\{ \mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n |x_i|^p = n \right\}. \tag{1.7}$$

Given a sequence $\{k_n\}_{n=1}^\infty$ of positive integers, we σ_n be the Haar measure on $\mathbb{V}_{n,k_n}$, and let σ be a measure on $\mathbb{V} = \otimes_n \mathbb{V}_{n,k_n}$ whose n-th marginal is σ_n for all $n \in \mathbb{N}$. All statements about probabilities on $\mathbb{V}_{n,k_n}$ and $\mathbb{V}$ in this work are made with respect to the measures σ_n and σ. We say that the sequence $\{k_n\}_{n=1}^\infty$ is increasing if $k_{n+1} \ge k_n$ for all $n \in \mathbb{N}$.

Finally, we recall the definition of a large deviations principle.

Definition 1.1 ([8, Section 1.2]). Let $\mathcal{T}$ be a topological space with Borel σ-algebra $\mathcal{B}$. A sequence $\{\mathbb{P}_n\}_{n=1}^\infty$ of probability measures on $\mathcal{T}$ satisfies a large deviation principle (LDP) with speed $s_n : \mathbb{N} \to \mathbb{R}_+$ and rate function $I \colon \mathcal{T} \to [0,\infty]$ if for all $B \in \mathcal{B}$,

$$-\inf_{x \in B^\circ} I(x) \le \liminf_{n\to\infty} \frac{1}{s_n} \log \mathbb{P}_n(B) \le \limsup_{n\to\infty} \frac{1}{s_n} \log \mathbb{P}_n(B) \le -\inf_{x \in \overline{B}} I(x), \tag{1.8}$$

where B° and $\overline{B}$ denote the interior and closure of B, respectively. We say that $\{\mathbb{P}_n\}_{n=1}^\infty$ satisfies a weak LDP when these inequalities hold for all $B \in \mathcal{B}$ such that $\overline{B}$ is compact.

A sequence of $\mathcal{T}$-valued random variables $\{Z^{(n)}\}_{n=1}^\infty$ is said to satisfy an LDP with speed s_n and rate function I if the sequence of measures $\{\mathbb{P}(Z^{(n)} \in \cdot)\}_{n=1}^\infty$ does. When the speed is not mentioned explicitly, we use the default $s_n = n$. The function I is said to be a good rate function if it has compact level sets $\Psi_I(\alpha) = \{x \in \mathcal{T} : I(x) \le \alpha\}$ for all $\alpha \in [0,\infty)$.

1.2. **Main Result.** For $p \in [1,\infty)$, let $f_p : \mathbb{R} \to \mathbb{R}$ be the probability density function of the p-generalized normal distribution:

$$f_p(x) \coloneqq \frac{1}{2p^{1/p}\Gamma(1+1/p)} \exp\left(-\frac{|x|^p}{p}\right). \tag{1.9}$$

Let Z_p be a random variable with density (1.9). We define $\overline{\Lambda}_p \colon \mathbb{R}^2 \to \mathbb{R} \cup \{\infty\}$ as

$$\overline{\Lambda}_p(t_1,t_2) = \log \mathbb{E}\left[\exp(t_1 Z_p + t_2 Z_p^2)\right]. \tag{1.10}$$

Further, let $\{g_i\}_{i=0}^{\infty}$ be an infinite sequence of independent Gaussian random variables with mean zero and variance one, denote $\mathbf{g} = (g_1, g_2, \dots)$, and define $\Lambda_p : \ell^2 \times \mathbb{R}^2 \to \mathbb{R} \cup \{\infty\}$ as

$$\Lambda_p(\mathbf{a}, b, c) := \mathbb{E}\big[\overline{\Lambda}_p\big(\langle \mathbf{a}, \mathbf{g}\rangle + bg_0, c\big)\big]. \tag{1.11}$$

where the expectation is taken with respect to g_0 and $\mathbf{g}$. The Legendre transform of $\Lambda_p(\mathbf{a}, b, c)$ is given by

$$\Lambda_p^*(\mathbf{a}', b', c') := \sup_{(\mathbf{a},b,c)\in\ell^2\times\mathbb{R}\times\mathbb{R}} \big\{\langle \mathbf{a}, \mathbf{a}'\rangle + bb' + cc' - \Lambda_p(\mathbf{a}, b, c)\big\}. \tag{1.12}$$

We also define

$$I_p(\mathbf{x}, r_1, r_2) := \Lambda_p^*\left(\mathbf{x}, \sqrt{r_1^2 - \|\mathbf{x}\|_2^2}, r_2\right) \tag{1.13}$$

for $(\mathbf{x}, r_1, r_2) \in \mathcal{X} \times \mathbb{R}_+$,[1] and for $(\mathbf{y}, r) \in \mathcal{X}$ we set

$$\mathcal{I}_p(\mathbf{y}, r) = \inf_{r_2\in\mathbb{R}_+} I_p\left(\frac{\mathbf{y}}{r_2^{1/p}}, \frac{r}{r_2^{1/p}}, r_2\right), \quad \hat{\mathcal{I}}_p(\mathbf{y}, r) = \inf_{r_2\in\mathbb{R}_+} I_p\left(\mathbf{y}, r, r_2\right), \tag{1.14}$$

where we set $I_p\left(\frac{\mathbf{y}}{r_2^{1/p}}, \frac{r}{r_2^{1/p}}, r_2\right) = I(0,0,0)$ for $r_2 = 0$.

The following theorem is our main result. It is proved in Section 5.

Theorem 1.2. *Fix $p \in [2,\infty)$, and let $\{k_n\}_{n=1}^{\infty}$ be an increasing sequence of positive integers such that $\lim_{n\to\infty} k_n = \infty$ and $\lim_{n\to\infty} k_n/n = 0$. For each $n \in \mathbb{N}$, let $Y^{(n)}$ be uniformly distributed on $\mathbb{B}_p^n$. Then for almost every sequence $\mathbf{a} = \{\mathbf{a}^{(n)}\}_{n\in\mathbb{N}} \in \mathbb{V}$, the sequence*

$$\left\{\Pi\left(n^{-1/2}(\mathbf{a}^{(n)})^{\mathsf{T}} Y^{(n)}\right)\right\}_{n=1}^{\infty} \tag{1.15}$$

satisfies an LDP with speed n in $\mathcal{X}$ with good rate function $\mathcal{I}_p$.

From Theorem 1.2, we can deduce LDPs for the Euclidean norm and maximal coordinate of the random projection $n^{-1/2}(\mathbf{a}^{(n)})^{\mathsf{T}}Y^{(n)}$. We state these in the following corollaries, which are also proved in Section 5.

Corollary 1.3. *Let p, $\{k_n\}_{n=1}^{\infty}$, and $Y^{(n)}$ be as in Theorem 1.2. Then for almost every sequence $\mathbf{a} = \left\{\mathbf{a}^{(n)}\right\}_{n\in\mathbb{N}} \in \mathbb{V}$, the sequence*

$$\left\{\left\|n^{-1/2}\left(\mathbf{a}^{(n)}\right)^{\mathsf{T}} Y^{(n)}\right\|_2\right\}_{n=1}^{\infty} \tag{1.16}$$

satisfies an LDP with speed n in $\mathbb{R}_+$ with a good rate function $\mathbb{I}_p$ given by

$$\mathbb{I}_p(r) = \sup_{t_1,t_2\in\mathbb{R}} \left\{t_1 r + t_2 - \mathbb{E}\left[\overline{\Lambda}_p(t_1 g, t_2)\right]\right\}, \tag{1.17}$$

where $\overline{\Lambda}_p$ is defined in (1.10), and g is a mean zero, variance one Gaussian random variable.Furthermore, $\mathbb{I}_p$ is convex.

[1] We sometimes use the notation (x, r_1, r_2) for elements of $\mathcal{X}\times\mathbb{R}$ instead of the more accurate $((x, r_1), r_2)$.

Corollary 1.4. *Let p, $\{k_n\}_{n=1}^{\infty}$, and $Y^{(n)}$ be as in Theorem 1.2. Then for almost every (deterministic) sequence $\mathbf{a} = \{\mathbf{a}^{(n)}\}_{n\in\mathbb{N}} \in \mathbb{V}$, the sequence*

$$\left\{ \left[n^{-1/2} (\mathbf{a}^{(n)})^{\mathsf{T}} Y^{(n)} \right]_1 \right\}_{n=1}^{\infty} \tag{1.18}$$

satisfies an LDP in $\mathbb{R}$ with good rate function

$$\mathcal{I}_{\max}(r) = \mathcal{I}_p((r, 0, 0 \dots), r). \tag{1.19}$$

Remark 1.5. The conclusions of Theorem 1.2, Corollary 1.3, and Corollary 1.4 also hold when $Y^{(n)}$ is uniformly distributed on $\mathbb{S}_p^n$. The proofs of these results also contain, as intermediate steps, proofs of the analogous claims for $\mathbb{S}_p^n$.

Remark 1.6. The space $\mathcal{X}$ is defined in (1.3) using the ℓ_2 norm in one coordinate, so Corollary 1.3 follows Theorem 1.2 by applying the contraction principle. The ℓ_2 norm in this definition could be replaced by ℓ_q for any $1 \le q < \infty$, and a parallel argument would give the analogue of Corollary 1.3 for the ℓ_q norm of the projections. For brevity, we do not take this up here.

Outline. Our main theorem is a consequence of a more general result, which provides LDPs for the projections of a large class of sequences of random vectors. It is given as Theorem 2.5 below, and follows from combining an upper bound and a lower bound, given in the next section as Proposition 2.3 and Proposition 2.4 respectively.

Section 2 states the three results just mentioned, along with their necessary assumptions. In Section 3 we prove Proposition 2.3, assuming certain preliminary lemmas stated in Section 3.1. In Section 4 we prove Proposition 2.4, assuming certain preliminary lemmas stated in Section 4.1. In Section 5, we prove Theorem 2.5 as a consequence Proposition 2.3 and Proposition 2.4, and use it to deduce Theorem 1.2, Corollary 1.3, and Corollary 1.4. In Section 6 we prove the lemmas necessary for the proof of Proposition 2.3, and in Section 7 we prove the preliminary lemmas necessary for the proof of Proposition 2.4. Appendix A contains auxiliary computations for p-Gaussian random variables. Appendix B recalls the Weingarten calculus and proves an auxiliary lemma.

Acknowledgments. P.L. is supported by NSF postdoctoral fellowship DMS-220289. K.R. is supported by NSF grant DMS-1954351 and Vannevar Bush Faculty Fellowship N0014-21-1-2887. X.X. is supported by NSF grant DMS-1954351. The second author would like to thank Amir Dembo for bringing to her attention the reference [7], which suggested the convenient topological space in which the LDP in this paper is established.

2. Main Technical Results

In this section, we list our main technical results and the assumptions they require. We first recall the following concept from convex analysis.

Definition 2.1 (Essential Smoothness). Let $f : \mathbb{R}^d \to (-\infty, \infty]$ be a convex function and let $\mathcal{D}_f = \{x \in \mathbb{R}^d : f(x) < \infty\}$. Then f is said to be essentially smooth if

(1) $\mathcal{D}_f^{\circ}$ is non-empty.

(2) $f(\cdot)$ is differentiable throughout $\mathcal{D}_f^\circ$.
(3) $f(\cdot)$ satisfies $\lim_{n\to\infty} |\nabla f(\lambda_n)| = \infty$ whenever $\{\lambda_n\}$ is a sequence in $\mathcal{D}_f^\circ$ converging to a boundary point of $\mathcal{D}_f^\circ$.

Given a sequence $\{Y^{(n)}\}_{n=1}^\infty$ of random vectors $Y^{(n)} = (Y_1, Y_2, \ldots, Y_n) \in \mathbb{R}^n$, we consider the following assumptions.

Assumption 1. There exists a sequence of independent, identically distributed real-valued random variables $\{X_j\}_{j=1}^\infty$, a continuous function $\eta\colon \mathbb{R} \to \mathbb{R}_+$, and a continuous function $\rho\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$Y_j^{(n)} \overset{d}{=} X_j \cdot \rho\left(\frac{1}{n}\sum_{i=1}^n \eta(X_i)\right) \tag{2.1}$$

for all $n \in \mathbb{N}$ and $j \le n$. We further suppose that $\rho(x) > 0$ for all $x > 0$.

Assumption 2. The variable X_1 has mean zero, is symmetric ($X_1 \overset{d}{=} -X_1$), and there exists a constant $C > 0$ such that

$$\mathbb{P}(|X_1| \ge s) \le 2\exp(-s^2/C^2) \tag{2.2}$$

for all $s \ge 0$.

Assumption 3. Set

$$\Lambda(t_1, t_2) \coloneqq \log \mathbb{E}\big[\exp\big(t_1 X_1 + t_2 \eta(X_1)\big)\big] \tag{2.3}$$

for $t_1, t_2 \in \mathbb{R}$. There exists some $0 < T \le \infty$ such that $\Lambda(t_1, t_2)$ is finite for all $(t_1, t_2) \in \mathbb{R} \times (-\infty, T)$ and $\Lambda(t_1, t_2)$ is essentially smooth. Further, the derivatives $\partial_1^\alpha \partial_2^\beta \Lambda(t_1, t_2)$ exist for all $(t_1, t_2) \in \mathbb{R} \times (-\infty, T)$ and all integers $\alpha, \beta \ge 0$ such that $\alpha + \beta \le 2$.

Assumption 4. There exists a continuous function $C\colon (-\infty, T) \to \mathbb{R}_+$ such that

$$\left|\partial_1^\alpha \partial_2^\beta \Lambda(t_1, t_2)\right| \le C(t_2)\left(1 + |t_1|^{2-\alpha}\right), \tag{2.4}$$

for all $(t_1, t_2) \in \mathbb{R} \times (-\infty, T)$ and all integers $\alpha, \beta \ge 0$ such that $\alpha + \beta \le 2$.

Remark 2.2. By taking $\rho(x) = 1$ for all $x \in \mathbb{R}_+$, we see that these assumptions are satisfied when $Y^{(n)} = (Y_1, \ldots, Y_n)$ is constructed from a sequence of independent, identically distributed random variables $\{Y_j\}_{j=1}^\infty$ satisfying Assumptions 2, 3, and 4.

Recalling g_0, $\mathbf{g}$, and $\Lambda(t_1, t_2)$ from (2.3), we define $\Lambda\colon \ell^2 \times \mathbb{R}^2 \to \mathbb{R} \cup \{\infty\}$ by

$$\Lambda(\mathbf{a}, b, c) \coloneqq \mathbb{E}\big[\Lambda\big(\langle \mathbf{a}, \mathbf{g}\rangle + b g_0, c\big)\big], \tag{2.5}$$

analogously to (1.10). The Legendre transform Λ^* is

$$\Lambda^*(\mathbf{a}', b', c') \coloneqq \sup_{(\mathbf{a},b,c)\in\ell^2\times\mathbb{R}\times\mathbb{R}} \left\{\langle \mathbf{a}, \mathbf{a}'\rangle + bb' + cc' - \Lambda(\mathbf{a}, b, c)\right\}. \tag{2.6}$$

We define

$$I(\mathbf{x}, r_1, r_2) \coloneqq \Lambda^*\left(\mathbf{x}, \sqrt{r_1^2 - \|\mathbf{x}\|_2^2}, r_2\right) \tag{2.7}$$

for $(x, r_1, r_2) \in \mathcal{X} \times \mathbb{R}_+$, and we define $\mathcal{I}_p(\mathbf{y}, r)$ for $(\mathbf{y}, r) \in \mathcal{X}$ by

$$\mathcal{I}(\mathbf{y}, r) = \min\left(I(0,0,0), \inf_{r_2>0} I\left(\frac{\mathbf{y}}{\rho(r_2)}, \frac{r}{\rho(r_2)}, r_2\right)\right). \tag{2.8}$$

The following large deviations upper bound is proved in Section 3.

Proposition 2.3. *Let $\{k_n\}_{n=1}^\infty$ be an increasing sequence of positive integers such that $\lim_{n\to\infty} k_n = \infty$. and $\lim_{n\to\infty} k_n/n = 0$. Suppose that $\{X_j\}_{j=1}^\infty$ is a sequence of i.i.d. random variables satisfying Assumptions 2 through 4. Let $X^{(n)} = (X_1, \dots, X_n) \in \mathbb{R}^n$ for all $n \geq 1$. Then for almost every sequence $\mathbf{a} = (\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \dots) \in \mathbb{V}$, we have*

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{P}\left(\left(\Pi(n^{-1/2}\mathbf{a}^\mathsf{T} X^{(n)}), \frac{1}{n}\sum_{i=1}^n \eta(X_i)\right) \in \mathcal{S}\right) \leq -\inf_{(\mathbf{x},r_1,r_2)\in\mathcal{S}} I(\mathbf{x}, r_1, r_2) \tag{2.9}$$

for all closed sets $\mathcal{S} \subset \mathcal{X} \times \mathbb{R}_+$.

The following large deviations lower bound is proved in Section 4.

Proposition 2.4. *Let $\{k_n\}_{n=1}^\infty$ be an increasing sequence of positive integers such that $\lim_{n\to\infty} k_n = \infty$ and $\lim_{n\to\infty} k_n/n = 0$. Suppose that $\{X_j\}_{j=1}^\infty$ is a sequence of i.i.d. random variables satisfying Assumptions 2 through 4. Let $X^{(n)} = (X_1, \dots, X_n) \in \mathbb{R}^n$ for all $n \geq 1$. Then for almost every sequence $\mathbf{a} = (\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \dots) \in \mathbb{V}$, we have*

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{P}\left(\left(\Pi(n^{-1/2}\mathbf{a}^\mathsf{T} X^{(n)}), \frac{1}{n}\sum_{i=1}^n \eta(X_i)\right) \in \mathcal{O}\right) \geq -\inf_{(x,r_1,r_2)\in\mathcal{O}} I(\mathbf{x}, r_1, r_2) \tag{2.10}$$

for all open sets $\mathcal{O} \subset \mathcal{X} \times \mathbb{R}_+$.

The following theorem is proved in Section 5 using the previous two lemmas, where it is then used to prove the main results stated in the previous section.

Theorem 2.5. *Let $\{k_n\}_{n=1}^\infty$ be an increasing sequence of positive integers such that $\lim_{n\to\infty} k_n = \infty$ and $\lim_{n\to\infty} k_n/n = 0$. Let $\{Y^{(n)}\}_{n=1}^\infty$ be a sequence of random vectors $Y^{(n)} = (Y_1, Y_2, \dots, Y_n)$ satisfying Assumptions 1–4. Then for almost every sequence $\mathbf{a} = \{\mathbf{a}^{(n)}\}_{n\in\mathbb{N}} \in \mathbb{V}$, the sequence*

$$\left\{\Pi\left(n^{-1/2}(\mathbf{a}^{(n)})^\mathsf{T} Y^{(n)}\right)\right\}_{n=1}^\infty \tag{2.11}$$

satisfies an LDP in $\mathcal{X}$ with good rate function $\mathcal{I}(\mathbf{y}, r)$.

3. Upper Bound

3.1. **Preliminary Lemmas.** In this section, we state some preliminary lemmas. The first lemma, which collects several topological properties of $\mathcal{X}$, is proved in Section 6.1.

Lemma 3.1. *(1) The topology on $\mathcal{X}$ is equivalent to the product topology on $\mathbb{R}^\mathbb{N} \times \mathbb{R}_+$, where $\mathbb{R}^\mathbb{N}$ is itself equipped with the product topology.*

(2) The topology on $\mathcal{X}$ is equivalent to the product topology on $\mathbb{R}^\mathbb{N} \times \mathbb{R}_+$, where $\mathbb{R}^\mathbb{N}$ is itself equipped with the weak-ℓ^2 topology.

(3) For any fixed $A < \infty$, the set $\{(\mathbf{x}, r) \in \mathcal{X} : r \le A\}$ is compact.

Out second lemma shows that the rate function in Proposition 2.3 is good. Its proof is deferred to Section 6.2.

Lemma 3.2. *The function I defined in* (2.7) *is a good rate function.*

To state the following lemmas, we recall the definition of exponential tightness.

Definition 3.3. A family of measures $\{\mu_n\}_{n=1}^{\infty}$ on a topological space space $\mathcal{T}$ is exponentially tight with speed $s_n : \mathbb{N} \to \mathbb{R}_+$ if for every $\alpha < \infty$, there exists a compact set $K_\alpha \subset \mathcal{T}$ such that

$$\limsup_{s_n \to \infty} \frac{1}{n} \log \mu_\varepsilon(K_\alpha^c) < -\alpha. \tag{3.1}$$

We say that a sequences of random variables $\{X_n\}_{n=1}^{\infty}$ is exponentially tight with speed s_n if the sequence of measures $\{\mu_n\}_{n=1}^{\infty}$ defined by $\mu_n(A) = \mathbb{P}(X_n \in A)$ is. We default to $s_n = n$ when no speed is explicitly given.

The next two lemmas address exponential tightness of the sequence of the norms of the projections $n^{-1/2}\mathbf{a}^\mathsf{T} X^{(n)}$ and the sums $\frac{1}{n}\sum_{i=1}^n \eta(X_i)$. They are proved in Section 6.3.

Lemma 3.4. *Let $\{X_j\}_{j=1}^{\infty}$ be random variables satisfying Assumption 2, and let $X^{(n)} = (X_1, \dots, X_n)$. Let $\{k_n\}_{n=1}^{\infty}$ be an increasing sequence of positive integers such that*

$$\lim_{n\to\infty} \frac{k_n}{n} = 0. \tag{3.2}$$

Then there exists a constant $c > 0$, depending only on the sequence $\{k_n\}_{n=1}^{\infty}$ and the constant C from Assumption 2, such that

$$\frac{1}{n} \log \mathbb{P}\left(\|n^{-1/2}\mathbf{a}^\mathsf{T} X^{(n)}\|_2^2 \ge t + 1 \right) \le -ct. \tag{3.3}$$

for every $n \in \mathbb{N}$ and $\mathbf{a} \in \mathbb{V}_{n,k_n}$.

Lemma 3.5. *Let $\{X_j\}_{j=1}^{\infty}$ be i.i.d. random variables satisfying Assumption* (2)*. Then this sequence is exponentially tight.*

Let $\mathbf{g} = (g_1, g_2, \dots, g_{k_n}) \in \mathbb{R}^{k_n}$ be a vector of mean zero, variance one Gaussians. For $\mathbf{u} \in \mathbb{R}^{k_n}$, $\mathbf{a} \in \mathbb{V}_{n,k_n}$, and $s \in \mathbb{R}$, define

$$F(\mathbf{u}, \mathbf{a}, s) := \frac{1}{n} \sum_{i=1}^{n} \Lambda\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i \rangle, s \right), \tag{3.4}$$

$$F_g(\mathbf{u}, s) := \mathbb{E}_{\mathbf{g}}\left[\Lambda\left(\langle \mathbf{u}, \mathbf{g} \rangle, s \right) \right], \tag{3.5}$$

where $\mathbf{a}_i$ denotes the i-th row of $\mathbf{a}$. The next lemma shows that $F(\mathbf{u}, \mathbf{a}, s)$ is closely approximated by $F_g(\mathbf{u}, s)$, with exponentially high probability. It is proved in Section 6.5.

Lemma 3.6. *Fix a constant $D > 0$, and let $d = \{d_n\}_{n=1}^{\infty}$ be a deterministic sequence of positive integers such that $\limsup_{n\to\infty} n^{-1}\log(d_n) = 0$, and $\{X_j\}_{j=1}^{\infty}$ are i.i.d. random variables satisfying Assumptions 2–4. For every $n \in \mathbb{N}$, let*

$$\mathcal{V}_n = \{\mathbf{v}^{(n,1)}, \mathbf{v}^{(n,2)}, \dots, \mathbf{v}^{(n,d_n)}\} \tag{3.6}$$

be a finite collection of vectors in ℓ^2 such that $\|\mathbf{v}^{(n,j)}\|_2 = D$ for all $n \geq 1$ and $j \leq d_n$. Fix $s \in (-\infty, T)$, where T is the constant from Assumption 3. Then

$$\lim_{n\to\infty} \sup_{j\leq d_n} \left| F(\mathbf{v}^{(n,j)}, \mathbf{a}, s) - F_g(\mathbf{v}^{(n,j)}, s) \right| = 0 \tag{3.7}$$

or almost every $(\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \dots) \in \mathbb{V}$.

For $\mathbf{a} \in \ell^2$ and a given $m \in \mathbb{N}$, we let $\mathbf{a}_{\leftarrow}$ be the sequence containing the first m coordinates of $\mathbf{a}$ followed by zeros:

$$\mathbf{a}_{\leftarrow} = \mathbf{a}_{\leftarrow m} = (a_1, a_2, \dots, a_m, 0, 0, \dots), \tag{3.8}$$

and $\mathbf{a}_{\rightarrow} = \mathbf{a}_{\rightarrow m} = \mathbf{a} - \mathbf{a}_{\leftarrow}$. We set

$$\Lambda_m(\mathbf{a}, b, c) \coloneqq \Lambda_g(\mathbf{a}_{\leftarrow}, b, c), \tag{3.9}$$

where we recall that Λ_g is defined in (2.5). We observe by Jensen's inequality that[2]

$$\Lambda(\mathbf{a}, b, c) \geq \Lambda_m(\mathbf{a}_{\leftarrow}, b, c), \tag{3.10}$$

and we have the monotonic limit

$$\lim_{m\to\infty} \Lambda_m(\mathbf{a}, b, c) = \Lambda(\mathbf{a}, b, c) \tag{3.11}$$

by the dominated convergence theorem and (2.4).

3.2. **Proof of the Upper Bound.** We first reduce the proof of the upper bound to a statement about closed sets of a simpler form.

Lemma 3.7. *Retain the notation and hypotheses of Proposition 2.3 with condition (1). Suppose that for every closed set $\mathcal{S} \subset \mathcal{X} \times \mathbb{R}$ and constant $A > 0$, the upper bound (2.9) holds for the set*

$$\mathcal{S} \cap \{(\mathbf{x}, r_1, r_2) : r_1 \leq A, r_2 \leq A\} \tag{3.12}$$

for a set $\Omega_{\mathcal{S},A}$ of sequences $\mathbf{a} = \big(\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \dots\big) \in \mathbb{V}$ such that $\mathbb{P}(\Omega_{\mathcal{S},A}) = 1$. Then for almost every sequence $\mathbf{a} \in \mathbb{V}$, the upper bound (2.9) holds for all closed sets $\mathcal{S}$.

Proof. Step 1. We first show that for any closed set $\mathcal{S}$,

$$\limsup_{n\to\infty} \frac{1}{n} \mathbb{P}\left(\left(\Pi(n^{-1/2}\mathbf{a}^{\mathsf{T}} X^{(n)}), \frac{1}{n}\sum_{i=1}^{n} \eta(X_i) \right) \in \mathcal{S} \right) \leq - \inf_{(\mathbf{x}, r_1, r_2)\in\mathcal{S}} I(\mathbf{x}, r_1, r_2) \tag{3.13}$$

holds for almost every sequence $\mathbf{a} \in \mathbb{V}$.

Set

$$K_A \coloneqq \{(\mathbf{x}, r_1, r_2) : r_1 \leq A, r_2 \leq A\}.$$

[2] Note that the function $\Lambda(t_1, t_2)$ is convex in t_1, since X_1 was assumed to be symmetric.

Let $\Omega_{\mathcal{S},A}$ be the set of sequences such that (3.13) holds for the closed set $\mathcal{S} \cap K_A$; by assumption, $\mathbb{P}(\Omega_{\mathcal{S},A}) = 1$. Set $\Omega_{\mathcal{S}} = \cap_{n\geq 1}\Omega_{\mathcal{S},n}$. Then $\Omega_{\mathcal{S}}$ has probability one. Pick any $\alpha < \infty$ such that $\mathcal{S} \subset \{I > \alpha\}$. By Lemma 3.4 and Lemma 3.5, there exists $n_0 \geq 1$ such that for every sequence $\mathbf{a} \in \mathbb{V}$,

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{P}\left(\left(\Pi(n^{-1/2}(\mathbf{a}^{(n)})^{\intercal} X^{(n)}), \frac{1}{n}\sum_{i=1}^{n} \eta(X_i)\right) \in K_{n_0}^c\right) \leq -\alpha.$$

By the assumption that the upper bound (2.9) holds for $\mathcal{S} \cap K_{n_0}$ and the fact that $\mathcal{S} \subset \{I > \alpha\}$,

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{P}\left(\left(\Pi(n^{-1/2}(\mathbf{a}^{(n)})^{\intercal} X^{(n)}), \frac{1}{n}\sum_{i=1}^{n} \eta(X_i)\right) \in \mathcal{S} \cap K_{n_0}\right) \leq -\alpha$$

for all sequences in $\Omega_{\mathcal{S}}$.

The above two equations imply that

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{P}\left(\left(\Pi(n^{-1/2}(\mathbf{a}^{(n)})^{\intercal} X^{(n)}), \frac{1}{n}\sum_{i=1}^{n} \eta(X_i)\right) \in \mathcal{S}\right) \leq -\alpha$$

holds for all sequences in $\Omega_{\mathcal{S}}$. Since $\alpha < \infty$ is arbitrary subject to $\mathcal{S} \subset \{I \leq \alpha\}^c$, the upper bound holds for the closed set $\mathcal{S}$ on $\Omega_{\mathcal{S}}$.

Step 2. We next show that for almost every sequence $\left(\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \ldots\right) \in \mathbb{V}$, the upper bound (2.9) holds for all closed sets $\mathcal{S}$.

Note that the space $\mathcal{X} \times \mathbb{R}_+$ possesses a countable basis $\{\mathcal{O}_i\}_{i=1}^{\infty}$, since it is a separable metric space. Consider the countable family of closed sets $\mathfrak{C} = \{\cap_{i\in\mathcal{F}}\mathcal{O}_i^c : \mathcal{F} \text{ is finite subset of } \mathbb{N}\}$ and define $\widehat{\Omega} = \cap_{\mathcal{S}\in\mathfrak{C}}\Omega_{\mathcal{S}}$. Then $\mathbb{P}(\widehat{\Omega}) = 1$.

For an arbitrary closed set $\mathcal{S}$, we have $\mathcal{S} = \cap_{i\in\mathcal{F}'}\mathcal{O}_i^c$ for some $\mathcal{F}' \subset \mathbb{N}$. For any $\varepsilon > 0$, there exists a finite $\mathcal{F} \subset \mathcal{F}' \subset \mathbb{N}$, such that for $\mathcal{C} = \cap_{i\in\mathcal{F}}\mathcal{O}_i^c$, we have

$$\inf_{(\mathbf{x},r_1,r_2)\in\mathcal{C}} I\left(\mathbf{x}, r_1, r_2\right) > \inf_{(\mathbf{x},r_1,r_2)\in\mathcal{S}} I\left(\mathbf{x}, r_1, r_2\right) - \varepsilon.$$

Therefore, on $\widehat{\Omega}$,

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n}\mathbb{P}\left(\left(\Pi(n^{-1/2}(\mathbf{a}^{(n)})^{\intercal} X^{(n)}), \frac{1}{n}\sum_{i=1}^{n} \eta(X_i)\right) \in \mathcal{S}\right) \\
\leq\ &\limsup_{n\to\infty} \frac{1}{n}\mathbb{P}\left(\left(\Pi(n^{-1/2}(\mathbf{a}^{(n)})^{\intercal} X^{(n)}), \frac{1}{n}\sum_{i=1}^{n} \eta(X_i)\right) \in \mathcal{C}\right) \\
\leq\ &-\inf_{(\mathbf{x},r_1,r_2)\in\mathcal{C}} I\left(\mathbf{x}, r_1, r_2\right) \\
<\ &-\inf_{(\mathbf{x},r_1,r_2)\in\mathcal{S}} I\left(\mathbf{x}, r_1, r_2\right) + \varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, the upper bound (2.9) holds for $\mathcal{S}$ on $\widehat{\Omega}$. The proof is complete. $\square$

Given positive integers $k \geq m$, let $\mathcal{J}_{k,m}$ denote the set of injective mappings from $\{1,\dots,m\}$ to $\{1,\dots,k\}$. For $\mathbf{x} \in \mathbb{R}^k$, we define $\tau(\mathbf{x}) \in \mathbb{R}^k$ as the vector

$$\tau(\mathbf{x}) = (x_{\tau(1)}, x_{\tau(2)}, \dots, x_{\tau(m)}, \dots) \tag{3.14}$$

whose first m coordinates are $x_{\tau(1)}, x_{\tau(2)}, \dots, x_{\tau(m)}$, with the remaining coordinates taken in their original order.

Proof of Proposition 2.3. The proof consists of two steps.
Step 1: Reduction to a finite number of coordinates. By Lemma 3.1, a subset $\mathcal{C} \subset \times\mathbb{R}_+$ is compact if and only if $\mathcal{C}$ is closed and there exists a constant $A \geq 0$ such that $\max(r_1, r_2) < A$ for all $(\mathbf{x}, r_1, r_2) \in \mathcal{C}$. By Lemma 3.7, it suffices to prove (2.9) for compact sets. We fix a compact set $\mathcal{C}$ and the corresponding constant A for the remainder of this proof.

Set

$$I = I(\mathcal{C}) = \inf_{(x,r_1,r_2)\in\mathcal{C}} I(x, r_1, r_2). \tag{3.15}$$

Let $\delta > 0$ be a parameter and define

$$I_\delta = \min(I, \delta^{-1}) - \delta. \tag{3.16}$$

From the definition (2.7), we see that for every $(\mathbf{x}, r_1, r_2) \in \mathcal{C}$, there exists a triple

$$(\mathbf{a}, b, c) = \big(\mathbf{a}(\mathbf{x}, r_1, r_2, \delta), b(\mathbf{x}, r_1, r_2, \delta), c(\mathbf{x}, r_1, r_2, \delta)\big) \in \ell^2 \times \mathbb{R}_+ \times \mathbb{R}_+ \tag{3.17}$$

such that

$$\langle \mathbf{a}, \mathbf{x} \rangle + bt + cr_2 - \Lambda(\mathbf{a}, b, c) > I_\delta, \tag{3.18}$$

with $t = \sqrt{r_1^2 - \|\mathbf{x}\|_2^2}$.

By (3.11), there exists $m(\mathbf{x}, r_1, r_2, \delta) \in \mathbb{N}$ such that

$$\langle \mathbf{a}_\leftarrow, \mathbf{x}_\leftarrow \rangle + bt_m + cr_2 - \Lambda_m(\mathbf{a}_\leftarrow, b, c) > I_\delta, \tag{3.19}$$

where $t_m = \sqrt{r_1^2 - \|\mathbf{x}_\leftarrow\|^2}$.

Consider the open set of near-optimizers

$$B_{\mathbf{x},r_1,r_2} = \{(\mathbf{y}, s_1, s_2) \in \mathbb{R}^m \times \mathbb{R}_+ \times \mathbb{R}_+ : \mathbf{a}_\leftarrow \cdot \mathbf{y} + bs_1 + cs_2 - \Lambda_m(\mathbf{a}_\leftarrow, b, c) > I_\delta\}, \tag{3.20}$$

where $(\mathbf{a}, b, c)$ are chosen as in (3.17), and define for $(\mathbf{x}, r_1, r_2) \in \mathcal{X} \times \mathbb{R}_+$ the set

$$\mathcal{V}_{\mathbf{x},r_1,r_2} = \left\{ (\mathbf{x}', r_1', r_2') \in \mathcal{X} \times \mathbb{R}_+ : \left(\mathbf{x}'_\leftarrow, \sqrt{r_1'^2 - |\mathbf{x}'_\leftarrow|^2}, r_2' \right) \in B_{\mathbf{x},r_1,r_2}, r_1' < A, r_2' < A \right\}. \tag{3.21}$$

This is an open set that contains $(\mathbf{x}, r_1, r_2)$. By compactness we can cover $\mathcal{C}$ by finitely many open sets of the form $\mathcal{V}_{x,r_1,r_2}$. Call these sets $\mathcal{V}^{(1)}, \dots, \mathcal{V}^{(m_0)}$, so that

$$\mathcal{C} \subset \bigcup_{j=1}^{m_0} \mathcal{V}^{(j)}. \tag{3.22}$$

We have $\mathcal{V}^{(1)} = \mathcal{V}_{x,r_1,r_2}$ for some triple (x, r_1, r_2). We write $\mathcal{V}$ for $\mathcal{V}^{(1)}$, B for the corresponding B_{x,r_1,r_2} defined in (3.20), and for the rest of this proof we let $(\mathbf{a}, b, c)$ stand for the associated triple (3.17).

We write $k = k_n$, suppressing the dependence on n. Given $\mathbf{u} \in \mathbb{R}^k$, we consider the map $\tau^{\mathbf{u}} \in \mathcal{J}_{k,m}$ such that $\tau^{\mathbf{u}}(\mathbf{u})_{\leftarrow} = [\mathbf{u}]_{\leftarrow}$. Then we see that

$$\begin{aligned}\Big\{(\mathbf{u}, v) \in \mathbb{R}^k \times \mathbb{R}_+ : ([\mathbf{u}], \|\mathbf{u}\|_2, v) \in \mathcal{V}\Big\} &\subset \{([\mathbf{u}]_{\leftarrow}, \|[\mathbf{u}]_{\rightarrow}\|_2, v) \in B, \|\mathbf{u}\|_2 \leq A\} \\ &= \{\tau^{\mathbf{u}}(\mathbf{u})_{\leftarrow}, \|\tau^{\mathbf{u}}(\mathbf{u})_{\rightarrow}\|_2, v) \in B, \|\mathbf{u}\|_2 \leq A\}.\end{aligned} \tag{3.23}$$

We conclude that the set $\{(\mathbf{u}, v) \in \mathbb{R}^k \times \mathbb{R}_+ : ([\mathbf{u}], \|\mathbf{u}\|_2, v) \in \mathcal{V}\}$ is contained in the set $\widehat{B}_k$ defined by

$$\bigcup_{\tau \in \mathcal{J}_{k,m}} \Big\{(\mathbf{u}, v) \in \mathbb{R}^k \times \mathbb{R}_+ : (\tau(\mathbf{u})_{\leftarrow}, \|\tau(\mathbf{u})_{\rightarrow}\|_2, v) \in B, \|\tau(\mathbf{u})_{\rightarrow}\|_\infty \leq Am^{-1/2}, v \leq A, \|u\|_2 \leq A\Big\}. \tag{3.24}$$

To obtain the condition $\|\tau(\mathbf{u})_{\rightarrow}\|_\infty \leq Am^{-1/2}$, we used that $\|\mathbf{y}_m\|_\infty \leq m^{-1/2}\|\mathbf{y}\|_2$ for any sequence $\mathbf{y}$ whose coordinates are in descending order, and the condition that $r_1 \leq A$ in the definition of $\mathcal{V}$ (so $\|\mathbf{u}\|_2 \leq r_1 \leq A$).

We introduce the notation

$$W = W_n = n^{-1/2}(\mathbf{a}^{(n)})^{\mathsf{T}} X^{(n)}, \qquad L = L_n = \frac{1}{n}\sum_{i=1}^n \eta(X_i). \tag{3.25}$$

By the previous lines, if $([W], \|W\|_2, L) \in \mathcal{V}$, then $(W, L) \in \widehat{B}_k$. Analogous reasoning applies not just to all $\mathcal{V} = \mathcal{V}^{(1)}$, but all sets $\mathcal{V}^{(j)}$ in the finite cover (3.22). We therefore see that if $([W], \|W\|_2, L) \in \mathcal{V}^{(j)}$, then $(W, L) \in \widehat{B}_k^{(j)}$, where $\widehat{B}_k^{(j)}$ is defined analogously to $\widehat{B}_k$ in (3.24), and $\widehat{B}_k^{(j)} = \widehat{B}_k$.

Using (3.22) and a union bound, we find

$$\mathbb{P}\left((\Pi(W), L) \in \mathcal{C}\right) \leq \sum_{j=1}^{m_0} \mathbb{P}\left((\Pi(W), L) \in \mathcal{V}^{(j)}\right) \tag{3.26}$$

$$\leq \sum_{j=1}^{m_0} \mathbb{P}\left((W, L) \in \widehat{B}_k^{(j)}\right). \tag{3.27}$$

It therefore suffices to bound the probabilities $\mathbb{P}\left((W, L) \in \widehat{B}_k^{(j)}\right)$. For concreteness, we focus on $\mathbb{P}\left((W, L) \in \widehat{B}_k^{(1)}\right)$. The argument for the other terms is analogous.

Step 2: Probability bound for $\widehat{B}_k$. We have

$$\mathbb{1}_{(W,L) \in \widehat{B}_k} \leq \sum_{\tau \in \mathcal{J}_{k,m}} \mathbb{1}_{(\tau(W)_{\leftarrow}, \|\tau(W)_{\rightarrow}\|_2, L) \in B} \mathbb{1}_{\|\tau(W)_{\rightarrow}\|_\infty \leq Am^{-1/2}} \mathbb{1}_{\|L\|_\infty \leq A}. \tag{3.28}$$

We define

$$H_{\mathbf{a},b,c}(B) = \inf_{(\mathbf{y}, s_1, s_2) \in B} \{\langle \mathbf{a}_{\leftarrow}, \mathbf{y}\rangle + bs_1 + cs_2\}. \tag{3.29}$$

Then

$$\mathbb{1}_{W\in\widehat{B}_k} \leq \exp(-NH_{\mathbf{a},b,c}(B)) \times \sum_{\tau\in\mathcal{J}_{k,m}} \exp(N(\mathbf{a}_{\leftarrow}\cdot\tau(W)_{\leftarrow} + b\|\tau(W)_{\rightarrow}\|_2 + cL))\mathbb{1}_{\|\tau(W)_{\rightarrow}\|\leq Am^{-1/2}}\mathbb{1}_{|L|\leq A}. \tag{3.30}$$

Let $\kappa > 0$ be a parameter, set $K_m(\kappa) = Ab\kappa^{-1}m^{-1/2}$, and consider a minimal cover $\mathcal{C}$ of the set

$$\mathcal{G} = \left\{\mathbf{u}' \in \mathbb{R}^{k-m} : \|\mathbf{u}'\|_2 = b\right\} \tag{3.31}$$

by open ℓ^2 balls of radius $b\kappa$ with centers in $\mathcal{G}$. Let $\mathcal{U}'$ denote the set of the centers of balls in $\mathcal{C}$, and let

$$\mathcal{U} = \bigcup_{\tau\in\mathcal{J}_{k,m}} \left\{\mathbf{u} \in \mathbb{R}^k : \tau(\mathbf{u})_{\leftarrow} = a_{\leftarrow}, \tau(\mathbf{u})_{\rightarrow} \in \mathcal{U}' \cup \{0\}\right\}. \tag{3.32}$$

A straightforward volume estimate shows that

$$|\mathcal{U}'| \leq (1 + 4/\kappa)^{k-m} \leq (1 + 4/\kappa)^k.$$

Together with the bound $|\mathcal{J}_{k,m}| \leq k^m$, this implies that for every $\hat{\varepsilon} > 0$, there exists $n_0(\{k_n\}_{n=1}^{\infty}, \kappa) \in \mathbb{N}$ such that

$$|\mathcal{U}| = \exp\left(o(n)\right). \tag{3.33}$$

Our next goal is to approximate the sum in (3.30) by a sum over elements in $\mathcal{U}$. Consider $\tau \in \mathcal{J}_{k,m}$ such that $\|\tau(W)_{\rightarrow}\|_\infty \leq Am^{-1/2}$. If $\|\tau(W)_{\rightarrow}\|_2 > \kappa$, then the random vector

$$U = b \cdot \frac{(\tau W)_{\rightarrow}}{\|\tau W_{\rightarrow}\|_2} \tag{3.34}$$

is such that $U \in \mathcal{G}$. Then there exists $\mathbf{u} \in \mathcal{U}$ and some $\tau \in \mathcal{J}_{k,m}$ such that $\tau(\mathbf{u})_{\leftarrow} = \mathbf{a}_{\leftarrow}$ and $\|U - \tau(\mathbf{u})_{\rightarrow}\|_2 < b\kappa$, so

$$\mathbf{a}_{\leftarrow}\cdot\tau(W)_{\leftarrow} + b\|\tau(W)_{\rightarrow}\|_2 = \tau(\mathbf{u})_{\leftarrow}\cdot\tau(W)_{\leftarrow} + \tau(W)_{\rightarrow}\cdot U \tag{3.35}$$

$$\leq \mathbf{u}\cdot W + Ab\kappa. \tag{3.36}$$

In the other case, if $\|\tau(W)_{\rightarrow}\|_2 \leq \kappa$, then a direct bound shows that

$$\mathbf{a}_{\leftarrow}\cdot\tau(W)_{\leftarrow} + b\|\tau(X)_{\rightarrow}\|_2 \leq \mathbf{u}\cdot W + 2b\kappa. \tag{3.37}$$

for $\mathbf{u}$ such that $\tau(\mathbf{u})_{\leftarrow} = a$ and $\tau(\mathbf{u})_{\rightarrow} = 0$, which satisfies $u \in \mathcal{U}$ by the definition of $\mathcal{U}$.

This upper bounds (3.35) and (3.37) together show that

$$\mathbb{P}(W \in \widehat{B}_k) \leq \exp(-n(H_{\mathbf{a},b,c}(B) - 2b\kappa - \kappa))\,|\mathcal{J}|\sum_{u\in\mathcal{U}}\mathbb{E}\exp(n(\mathbf{u}\cdot W + cL)). \tag{3.38}$$

Taking logarithms, we find

$$n^{-1}\log\mathbb{P}(W \in \widehat{B}_k) \leq -(H_{\mathbf{a},b,c}(B)-2b\kappa)+n^{-1}\log|\mathcal{J}|\,|\mathcal{U}|+\max_{u\in\mathcal{U}} n^{-1}\log\mathbb{E}\exp(n(\mathbf{u}\cdot W+cL)).$$

Recall $(W_n)_j = n^{-1/2}\sum_{i=1}^n a_{ji}X_i$. We have

$$\begin{aligned}
&n^{-1}\log\mathbb{E}\exp(n(\mathbf{u}\cdot W + ncL)) && (3.39)\\
&\qquad = n^{-1}\log\mathbb{E}\exp\left(n\sum_{i=1}^k u_j(W_n)_j + ncL\right) && (3.40)\\
&\qquad = n^{-1}\log\mathbb{E}\exp\left(n^{1/2}\sum_i^n X_i\left(\sum_{j=1}^k u_j a_{ji}\right) + c\sum_i^n X_i^2\right) && (3.41)\\
&\qquad = n^{-1}\sum_{i=1}^n \log\mathbb{E}\exp\left(n^{1/2}X_i\left(\sum_{j=1}^k u_j a_{ji}\right) + cX_i^2\right) && (3.42)\\
&\qquad = F(\mathbf{u},\mathbf{a}^{(n)},c). && (3.43)
\end{aligned}$$

Further, we have $\|\mathbf{u}\|^2 = \|\mathbf{a}\|_2^2$ or $\|\mathbf{u}\|^2 = \|\mathbf{a}\|_2^2 + \|b\|_2^2$ for $\mathbf{u}\in\mathcal{U}$. In particular, all vectors $\mathbf{u}\in\mathcal{U}$ have one of two ℓ^2 norms. Recalling the previous estimates on $|\mathcal{U}|$ from (3.33), and using Lemma 3.6 to control the maximum in (3.2), we obtain

$$\begin{aligned}
\limsup n^{-1}\log\mathbb{P}(W\in\widehat{B}_k) &\le -H_{\mathbf{a},b,c}(B) + 2b\kappa + \max(\Lambda_m(\mathbf{a},b,c),\Lambda_m(\mathbf{a},0,c)) + \varepsilon && (3.44)\\
&\le -H_{\mathbf{a},b,c}(B) + 2b\kappa + \Lambda_m(\mathbf{a},b,c) + \varepsilon. && (3.45)
\end{aligned}$$

In the last line, we used (3.11) to simplify the maximum. By the definition of B in (3.20), this becomes

$$\begin{aligned}
\limsup n^{-1}\log\mathbb{P}(W\in\widehat{B}_k) &\le -I_\delta - \Lambda_m(\mathbf{a},b,c) + 2b\kappa + \Lambda_m(\mathbf{a},b,c) + \varepsilon && (3.46)\\
&\le -I_\delta + 2b\kappa + \varepsilon. && (3.47)
\end{aligned}$$

Since this bound holds for all $\kappa,\varepsilon>0$, it implies that

$$\limsup n^{-1}\log\mathbb{P}(W\in\widehat{B}_k) \le -I_\delta. \tag{3.48}$$

Finally, since this holds for all $\delta>0$, we obtain

$$\limsup n^{-1}\log\mathbb{P}(W\in\widehat{B}_k) \le -I. \tag{3.49}$$

Inserting this and the analogous bounds for all $B_k^{(j)}$ into (3.27) completes the proof. □

4. Lower Bound

4.1. **Preliminary Lemmas.** We begin by stating some preliminary lemmas, whose proofs are postponed until Section 7.1.

Given a sequence $\{X_i\}_{i=1}^\infty$ of random variables satisfying the assumptions of Proposition 2.4, integers $n,k\ge 1$, and a matrix $\mathbf{a}\in\mathbb{V}_{n,k}$, we define

$$W_j = W_j^{(n)} = \frac{1}{\sqrt{n}}\sum_{i=1}^n a_{ij}X_i, \qquad W = W^{(n)} = (W_1^{(n)}, W_2^{(n)},\dots,W_k^{(n)}),$$

for $j \le k$. Given a continuous function $\eta\colon \mathbb{R} \to \mathbb{R}_+$, we also define

$$L = L_n = \frac{1}{n}\sum_{i=1}^{n} \eta(X_i).$$

Given a metric space $\mathcal{T}$ with metric d, we define the ball $B(t_0, r)$ for $t_0 \in \mathcal{T}$ and $r \ge 0$ by

$$B(t_0, r) = \{t \in \mathcal{T} : d(t_0, t) < r\}. \tag{4.1}$$

Given $m \in \mathbb{N}$ and $\mathbf{v} \in \ell^2$, we recall that $\mathbf{v}_{\to m}$ and $\mathbf{v}_{\leftarrow m}$ were defined in (3.8). Recalling the constant T from Assumption 3, we set

$$\mathcal{D}_m = \{(\mathbf{v}, b, c) : (\mathbf{v}, b) \in \mathbb{R}^m \times R, c \in (-\infty, T)\}. \tag{4.2}$$

Finally, we let $\nabla\Lambda_m \in \mathbb{R}^{m+2}$ denote the gradient of Λ_m:

$$\nabla\Lambda_m(\mathbf{v}, b, c) = \\ \left(\partial_{v_1}\Lambda_m(\mathbf{v}, b, c), \partial_{v_2}\Lambda_m(\mathbf{v}, b, c), \ldots, \partial_{v_m}\Lambda_m(\mathbf{v}, b, c), \partial_b\Lambda_m(\mathbf{v}, b, c), \partial_c\Lambda_m(\mathbf{v}, b, c)\right). \tag{4.3}$$

Lemma 4.1. *Fix $\varepsilon > 0$, an open set $\mathcal{O} \subset \mathcal{X} \times \mathbb{R}$, and a point $(\mathbf{w}, r, s) \in \mathcal{O}$ such that $I(\mathbf{w}, r, s) < \infty$. Then there exist $\bar\kappa > 0$, $m_0 \in \mathbb{N}$, and $(\overline{\mathbf{w}}, \bar t, \bar s) \in \mathbb{R}^{m_0} \times \mathbb{R}_+ \times \mathbb{R}_+$ such that the following claims hold.*

(1) $(\bar{\mathbf{w}}, \bar t, \bar s) = \nabla\Lambda_{m_0}(\bar{\mathbf{v}}, \bar b, \bar c)$ for some $(\bar{\mathbf{v}}, \bar b, \bar c) \in \mathbb{R}^{m_0} \times \mathbb{R} \times \mathbb{R}$.
(2) $|\bar{\mathbf{w}}_j| > |\bar{\mathbf{w}}_{j+1}| > 0$, for all $1 \le j \le m_0 - 1$.
(3) $\Lambda^*_{m_0}(\bar{\mathbf{w}}, \bar t, \bar s) \le I(\mathbf{w}, r, s) + \varepsilon$.
(4) For all $(\mathbf{w}', r', s') \in \mathcal{X} \times \mathbb{R}$, if

$$\left(\mathbf{w}'_{\leftarrow m_0}, \sqrt{r'^2 - \|\mathbf{w}'_{\leftarrow m_0}\|_2^2}, s'\right) \in B\left((\overline{\mathbf{w}}, \bar t, \bar s), \bar\kappa/2\right), \tag{4.4}$$

then $(\mathbf{w}', r', s') \in \mathcal{O}$.

Lemma 4.2. *Fix $m \ge 1$, $r > 0$, and $(\mathbf{v}, b, c) \in \mathcal{D}_m$. Define $(\mathbf{w}, t, s) = \nabla\Lambda_m(\mathbf{v}, b, c)$ and $B = B((\mathbf{w}, t, s), r)$. Then, for σ-a.e. $\mathbf{a} \in \mathbb{V}$,*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left((W_{\leftarrow m}, \|W_{\to m}\|_2, L) \in B, \|W_{\to m}\|_\infty \le 2m^{-1/2}\right) \ge -\Lambda^*_m(\mathbf{w}, t, s).$$

4.2. **Proof of Lower Bound.**

Proof of Proposition 2.4. We first show that for any $\varepsilon > 0$, open set $\mathcal{O} \subset \mathcal{X} \times R$, and point $(\mathbf{w}, r, s) \in \mathcal{O}$, we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(([W], \|W\|_2, L) \in \mathcal{O}\right) \ge -I(\mathbf{w}, r, s) - \varepsilon \tag{4.5}$$

for a set $\Omega_{\mathcal{O},\varepsilon,(\mathbf{w},r,s)} \subset \mathbb{V}$ of sequences $\{\mathbf{a}^{(n)}\}_{n\in\mathbb{N}}$ such that $\sigma(\Omega_{\mathcal{O},\varepsilon,(\mathbf{w},r,s)}) = 1$. We may assume that $I(\mathbf{w}, r, s) < \infty$, since otherwise the claim is trivial.

Let $\bar{\kappa} > 0$, $m_0 \in \mathbb{N}$, and $(\bar{\mathbf{w}}, \bar{t}, \bar{s})$ and be given by Lemma 4.1. Fix $m > m_0$ such that $3m^{-1/2} < |\bar{\mathbf{w}}_{m_0}|$ and set

$$\kappa = \min\left\{ \frac{\bar{\kappa}}{2}, m^{-1/2}, \frac{1}{3} \min_{j \le m_0 - 1} \left(|\bar{\mathbf{w}}_j| - |\bar{\mathbf{w}}_{j+1}| \right) \right\}.$$

Observe that $\kappa > 0$. Set

$$\widehat{\mathbf{w}} = (\bar{\mathbf{w}}, 0, \ldots, 0) \in \mathbb{R}^m, \quad B_1 = B\left((\widehat{\mathbf{w}}, \bar{t}, \bar{s}), \kappa\right), \quad B_2 = B\left((\overline{\mathbf{w}}, \bar{t}, \bar{s}), \bar{\kappa}/2\right).$$

By the definition of κ, and using Lemma 4.1,

$$\begin{aligned} &\{ (\mathbf{w}', t', s') \in \ell^2 \times \mathbb{R}_+ \times \mathbb{R}_+ : (\mathbf{w}'_{\leftarrow m}, t', s') \in B_1 \text{ and } \left\|\mathbf{w}'_{\to m}\right\|_\infty \le 2m^{-1/2} \} \\ &\subset \{ (\mathbf{w}', t', s') \in \ell^2 \times \mathbb{R}_+ \times \mathbb{R}_+ : [\mathbf{w}']_{\leftarrow m_0} = \mathbf{w}'_{\leftarrow m_0} \text{ and } (\mathbf{w}'_{\leftarrow m_0}, t'', s') \in B_2 \}, \end{aligned} \tag{4.6}$$

where $t'' = \sqrt{(t')^2 - \|\mathbf{w}'_{\leftarrow m_0}\|_2^2 + \|\mathbf{w}'_{\leftarrow m}\|_2^2}$.

Combining (4.4), (4.6), the first claim of Lemma 4.1, and Lemma 4.2, we find that the following holds for almost all sequences $\mathbf{a} \in \mathbb{V}$:

$$\begin{aligned} &\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(([W], \|W\|_2, L) \in \mathcal{O}\right) \\ &\quad \ge \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(([W]_{\leftarrow m_0}, \|W_{\to m_0}\|_2, L) \in B_2 \big) \\ &\quad \ge \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left((W_{\leftarrow m}, \|W_{\to m}\|_2, L) \in B_1, \|W_{\to m}\|_\infty \le 2m^{-1/2} \right) \\ &\quad \ge -\Lambda_m^*(\widehat{\mathbf{w}}, \bar{t}, \bar{s}) = -\Lambda_{m_0}^*(\bar{\mathbf{w}}, \bar{t}, \bar{s}) \ge -I(\mathbf{w}, r, s) - \varepsilon. \end{aligned} \tag{4.7}$$

This proves the claim (4.5).

Recall that $\Omega_{\mathcal{O},\varepsilon,(\mathbf{w},r,s)}$ denotes the set in $\mathbb{V}$ such that (4.5) holds. Let $\{\mathbf{w}_l, r_l, s_l\}_{l=1}^\infty$ be a sequence such that

$$\lim_{l\to\infty} I(\mathbf{w}_l, r_l, s_l) = \inf_{(\mathbf{w},r,s)\in\mathcal{O}} I(\mathbf{w}, r, s).$$

Let $\Omega_{\mathcal{O}} = \cap_{l=1}^\infty \Omega_{\mathcal{O},\frac{1}{l},(\mathbf{w}_l,r_l,s_l)}$. We have $\sigma(\Omega_{\mathcal{O}}) = 1$ and

$$\liminf_{n\to\infty} n^{-1} \log P_n\left(([W], \|W\|_2, L) \in \mathcal{O}\right) \ge -\inf_{(\mathbf{w},r,s)\in\mathcal{O}} I(\mathbf{w}, r, s).$$

Finally, observe that since $\mathcal{X} \times \mathbb{R}$ is a separable metric space, it possesses a countable basis $\{\mathcal{O}_i\}_{i=1}^\infty$. Let $\widehat{\Omega} = \cap_{i=1}^\infty \Omega_{\mathcal{O}_i}$. We have $\sigma(\widehat{\Omega}) = 1$. Now consider an arbitrary open set $\mathcal{O} \subset \mathcal{X} \times \mathbb{R}$ and any $(\mathbf{w}, r, s) \in \mathcal{O}$. By the definition of a basis, there exists a basis element $\mathcal{O}_i$ such that $(\mathbf{w}, r, s) \in \mathcal{O}_i \subset \mathcal{O}$. On the set $\widehat{\Omega}$, we have

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(([W], \|W\|_2, L) \in \mathcal{O}\right) &\ge \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(([W], \|W\|_2, L) \in \mathcal{O}_i\right) \\ &\ge -I(\mathbf{w}, r, s). \end{aligned} \tag{4.8}$$

Since $(\mathbf{w}, r, s) \in \mathcal{O}$ is arbitrary, (2.4) follows from (4.8). $\square$

5. Proof of Main Result

Proof of Theorem 2.5. Combining Proposition 2.3 (1) and Proposition 2.4 (1), we conclude that the sequence of random variables

$$\left(\Pi\left(n^{-1/2}\left(\mathbf{a}^{(n)}\right)^T X^{(n)}\right), \frac{1}{n}\sum_{i=1}^{n}\eta(X_i)\right) \tag{5.1}$$

satisfies a large deviation principle with good rate function $I\left(\mathbf{x}, r_1, r_2\right) = \Lambda_g^*\left(\mathbf{x}, \sqrt{r_1^2 - \|\mathbf{x}\|_2^2}, r_2\right)$.

By definition,

$$\Pi\left(n^{-1/2}\left(\mathbf{a}^{(n)}\right)^T Y^{(n)}\right) = \Pi\left(n^{-1/2}\left(\mathbf{a}^{(n)}\right)^T X^{(n)}\right) \times \rho\left(\frac{1}{n}\sum_{i=1}^{n}\eta(X_i)\right).$$

By applying the contraction principle (see [8, Theorem 4.2.1]) to the continuous map

$$f\colon \mathcal{X} \times \mathbb{R}_+ \to \mathcal{X}, \qquad f(\mathbf{x}, r_1, r_2) = (\rho(r_2)\mathbf{x}, \rho(r_2)r_1),$$

we obtain a large deviation principle for $\Pi\left(n^{-1/2}\left(\mathbf{a}^{(n)}\right)^T Y^{(n)}\right)$ with good rate function

$$\mathcal{I}(\mathbf{y}, r) = \inf\{I\left(\mathbf{x}, r_1, r_2\right) : \mathbf{y} = \rho(r_2)\mathbf{x},\ \rho(r_2)r_1 = r\} = \inf_{r_2 \in \mathbb{R}_+} I\left(\frac{\mathbf{y}}{\rho(r_2)}, \frac{r}{\rho(r_2)}, r_2\right).$$

□

Given Theorem 2.5, the next proof closely follows [18, Section 5]. We include the details for completeness.

Proof of Theorem 1.2. We now verify that $Y^{(n)}$ satisfies the assumptions of Theorem 2.5 when uniformly distributed on the ℓ_n^p sphere for some $p \in [2, \infty)$:

(1) By [26, Lemma 1], this condition holds with $\{X_i\}_{i\in\mathbb{N}}$ being the i.i.d. sequence with common distribution equal to the p-Gaussian distribution (the probability measure on $\mathbb{R}$ with density proportional to $e^{-|y|^p/p}$), $\eta(x) = |x|^p$, and $\rho(y) = y^{-1/p}$.
(2) This condition follows directly from (1) and $p \in [2, \infty)$.
(3) It is straightforward to see that domain of Λ is $\mathbb{R} \times (-\infty, \frac{1}{p})$. The essential smoothness is established in [12, Lemma 5.8].
(4) The growth conditions on Λ and the derivatives of Λ are established in Lemma A.2.

By Theorem 2.5, this proves the theorem when $Y^{(n)}$ is uniformly distributed on $\mathbb{S}_n^p$.

Next, we consider the case where $Y^{(n)}$ is uniformly distributed on $\mathbb{D}_p^n$. Let U be a uniform random variable on $[0, 1]$. We recall that if the random vector $X^{(n)} \in \mathbb{R}^n$ is uniformly distributed on the sphere $\mathbb{S}_p^n$ and independent from U, then $U^{1/n}X^{(n)}$ is uniform on $\mathbb{S}_p^n$ [26]. It then suffices to prove the LDP for $Y^{(n)} = U^{1/n}X^{(n)}$.

By [12, Lemma 3.3], the sequence $\{U^{1/n}\}_{n=1}^{\infty}$ satisfies an LDP with good rate function I_U given by $I_U(u) = -\log u$ for $u \in (0, 1]$, and $I_U(u) = +\infty$ otherwise. Recalling that U

and $X^{(n)}$ are independent, and using Theorem 2.5, we find that

$$\left\{\left(\Pi\left(n^{-1/2}(\mathbf{a}^{(n)})^{\intercal}X^{(n)}\right), U^{1/n}\right)\right\}_{n=1}^{\infty} \tag{5.2}$$

satisfies an LDP in $\mathcal{X}\times\mathbb{R}$ with good rate function $\mathcal{I}_U(u)+I(\mathbf{x},r)$. By the contraction principle,

$$\left\{\Pi\left(U^{1/n}n^{-1/2}(\mathbf{a}^{(n)})^{\intercal}X^{(n)}\right)\right\}_{n=1}^{\infty} \tag{5.3}$$

satisfies an LDP on $\mathcal{X}$ with rate function

$$\widehat{I}(\mathbf{x},r) = \inf_{(\mathbf{y},s)\in\mathcal{X},u\in\mathbb{R}}\{I_U(u)+\mathcal{I}(\mathbf{x},r) : (\mathbf{x},r)=(u\mathbf{y},us)\} \tag{5.4}$$

$$= \inf_{u\in(0,1]}\left\{-\log u + \mathcal{I}\left(\frac{\mathbf{x}}{u},\frac{r}{u}\right)\right\}. \tag{5.5}$$

Note the map $u\mapsto u^{-1}$ is monotonically decreasing. It is clear from the definition (1.14) that the map $t\mapsto\mathcal{I}(t\mathbf{x},tr)$ is increasing for $t\in R_+$. Since $u\mapsto u^{-1}$ is monotonically decreasing, we deduce that $u\mapsto\mathcal{I}(u^{-1}\mathbf{x},u^{-1}r)$ is decreasing. Combined with the fact that $-\log u$ is decreasing, we find that the infimum in (5.5) is attained at $u=1$. Then

$$\widehat{I}(\mathbf{x},r)=\mathcal{I}(\mathbf{x},r), \tag{5.6}$$

as claimed. □

Proof of Corollary 1.3. It was shown in the previous proof that $Y^{(n)}$ satisfies the assumptions of Theorem 2.5 when uniformly distributed on $\mathbb{S}_n^p$ for some $p\in[2,\infty)$. By applying contraction principle to Theorem 2.5 with the continuous function

$$f:\mathcal{X}\to\mathbb{R}_+,\quad f(\mathbf{y},r):=r,$$

we have

$$\left\{\left\|n^{-1/2}\left(\mathbf{a}^{(n)}\right)^{\intercal}Y^{(n)}\right\|_2\right\}_{n=1}^{\infty}$$

satisfies an LDP with speed n and good rate function

$$\mathbb{I}(r)=\inf\left\{\mathcal{I}(\mathbf{y},r):(\mathbf{y},r)\in\mathcal{X}\right\}.$$

By definition of $\mathcal{I}$, we have

$$\begin{aligned}\mathbb{I}(r) &= \inf\left\{I\left(\mathbf{x},r_1,r_2\right):(\mathbf{x},r_1,r_2)\in\mathcal{X}\times\mathbb{R}_+,\ r_1|r_2|^{-1/p}=r\right\}\\ &= \inf\left\{I\left(\mathbf{x},r_1,\left(\frac{r_1}{r}\right)^p\right):(\mathbf{x},r_1)\in\mathcal{X}\right\}\\ &= \inf\left\{\Lambda_g^*\left(\mathbf{x},\sqrt{r_1^2-\|\mathbf{x}\|_2^2},\left(\frac{r_1}{r}\right)^p\right):(\mathbf{x},r_1)\in\mathcal{X}\right\}.\end{aligned} \tag{5.7}$$

Note that

$$\begin{aligned}
\Lambda_g^*\left(\mathbf{x}, \sqrt{r_1^2-\|\mathbf{x}\|_2^2}, \left(\frac{r_1}{r}\right)^p\right) &= \sup_{(\mathbf{u},b,c)\in\ell^2\times\mathbb{R}\times\mathbb{R}} \left\{\langle \mathbf{u},\mathbf{x}\rangle + b\sqrt{r_1^2-\|\mathbf{x}\|_2^2} + c\left(\frac{r_1}{r}\right)^p - \Lambda_g\left(\mathbf{u},b,c\right)\right\} \\
&= \sup_{(\mathbf{u},b,c)\in\ell^2\times\mathbb{R}\times\mathbb{R}} \left\{\|(\mathbf{u},b)\|_2\, r_1 + c\left(\frac{r_1}{r}\right)^p - \mathbb{E}\left[\Lambda\left(\|\left(\mathbf{u},b\right)\|_2 g, c\right)\right]\right\} \\
&= \sup_{(v,c)\in\mathbb{R}_+\times\mathbb{R}} \left\{v r_1 + c\left(\frac{r_1}{r}\right)^p - \mathbb{E}\left[\Lambda\left(vg,c\right)\right]\right\} \\
&= \sup_{(v,c)\in\mathbb{R}\times\mathbb{R}} \left\{v r_1 + c\left(\frac{r_1}{r}\right)^p - \mathbb{E}\left[\Lambda\left(vg,c\right)\right]\right\},
\end{aligned} \tag{5.8}$$

where the second equality follows from the Cauchy–Schwarz inequality and the last equality follows from the observation that replacing v by $|v|$ increases the quantity we are taking supremum over.

Combining (5.7) and (5.8), we have

$$\mathbb{I}(r) = \inf_{\tau\in\mathbb{R}_+} \sup_{(v,c)\in\mathbb{R}\times\mathbb{R}} \left\{v\tau + c\left(\frac{\tau}{r}\right)^p - \mathbb{E}\left[\Lambda\left(vg,c\right)\right]\right\}.$$

By [22, Lemma 2.1], we have

$$\mathbb{I}(r) = \sup_{(v,c)\in\mathbb{R}\times\mathbb{R}} \left\{vr + c - \mathbb{E}\left[\Lambda\left(vg,c\right)\right]\right\}.$$

Convexity of $\mathbb{I}$ is now a standard result from convex analysis (see [25, Theorem 11.1]). ☐

Proof of Corollary 1.4. Consider the map $\pi : \mathcal{X} \to \mathbb{R}$ given by $\pi(\mathbf{x}, r) = x_1$, the first element of the ordered sequence $\mathbf{x}$. By Theorem 2.5 and the contraction principle, the sequence (1.18) satisfies an LDP with good rate function

$$\widehat{I}(r) = \inf_{(\mathbf{y},s)\in\mathcal{X}} \{\mathcal{I}(\mathbf{y},s) : r = \pi(\mathbf{y},s)\} = \mathcal{I}((r,0,0\ldots),r), \tag{5.9}$$

where the last inequality follows from the definition (1.14) and the fact that Λ_g is even in its first two arguments.

The proof for $\mathbb{B}_p^n$ is nearly identical, so we omit it. ☐

6. Preliminary Lemmas: Upper Bound

6.1. Topological properties of $\mathcal{X}$.

Proof of Lemma 3.1. Due to the ordering of the first component in $\mathcal{X}$, for all $(\mathbf{x}, r) \in \mathcal{X}$,

$$|x_m| \le m^{-1/2}\|\mathbf{x}\|_2 \le m^{-1/2} r. \tag{6.1}$$

Claim 1. Suppose $(\mathbf{x}_n, r_n) \to (\mathbf{x}, r)$ in $\mathcal{X}$. Then in particular, $\mathbf{x}_n \to \mathbf{x}$ in ℓ_∞. Therefore, for any coordinate projection map $p_i : \mathbf{y} \mapsto y_i$, $p_i(\mathbf{x}_n - \mathbf{x}) \to 0$ as $n \to \infty$. Thus $(\mathbf{x}_n, r_n) \to (\mathbf{x}, r)$ in $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}_+$ when $\mathbb{R}^{\mathbb{N}}$ is equipped with the product topology.

On the other hand, suppose $(\mathbf{x}_n, r_n) \to (\mathbf{x}, r)$ in $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}_+$ when $\mathbb{R}^{\mathbb{N}}$ is equipped with the product topology. Pick any $\varepsilon > 0$. There exists $N \in \mathbb{N}$, such that $\|(\mathbf{x}_n)_{\leftarrow N} - \mathbf{x}_{\leftarrow N}\|_\infty < \varepsilon$, $N^{-1/2} < \varepsilon/(2r+1)$ and for all $n > N$, $r_n < r+1$. Let $n > N$ be arbitrary, by (6.1), we have

$$\begin{aligned}\|\mathbf{x}_n - \mathbf{x}\|_\infty &= \max\left(\|(\mathbf{x}_n)_{\leftarrow N} - \mathbf{x}_{\leftarrow N}\|_\infty, \|(\mathbf{x}_n)_{\to N} - \mathbf{x}_{\to N}\|_\infty\right) \\ &\leq \max\left(\|(\mathbf{x}_n)_{\leftarrow N} - \mathbf{x}_{\leftarrow N}\|_\infty, N^{-1/2} r_n + N^{-1/2} r\right) \\ &\leq \varepsilon.\end{aligned}$$

Therefore, $(\mathbf{x}_n, r_n) \to (\mathbf{x}, r)$ in $\mathcal{X}$.

Claim 2. Suppose $(\mathbf{x}_n, r_n) \to (\mathbf{x}, r)$ in $\mathcal{X}$. Pick any $\mathbf{u} \in \ell^2$ and any $\varepsilon > 0$. It suffices to show that $\langle \mathbf{u}, \mathbf{x}_n - \mathbf{x}\rangle \to 0$. Without loss of generality, assume $\mathbf{u} \neq 0$. Since $\mathbf{u} \in \ell^2$, there exists $m \in \mathbb{N}$, such that $\|\mathbf{u}_{\to m}\|_2 < \varepsilon/(2r+1)$. Let $N \in \mathbb{N}$ be large enough such that $\|\mathbf{x}_n - \mathbf{x}\|_\infty < \varepsilon m^{-1/2}/\|\mathbf{u}\|_2$ and for all $n > N$, $r_n < r+1$. Then for any $n > N$,

$$\begin{aligned}\langle \mathbf{u}, \mathbf{x}_n - \mathbf{x}\rangle &= \langle \mathbf{u}_{\leftarrow m}, (\mathbf{x}_n - \mathbf{x})_{\leftarrow m}\rangle + \langle \mathbf{u}_{\to m}, (\mathbf{x}_n - \mathbf{x})_{\to m}\rangle \\ &\leq \|\mathbf{u}\|_2 \|\left(\mathbf{x}_n - \mathbf{x}\right)_{\leftarrow m}\|_2 + \|\mathbf{u}_{\to m}\|_2 \|\left(\mathbf{x}_n - \mathbf{x}\right)\|_2 \\ &\leq \|\mathbf{u}\|_2 \|\left(\mathbf{x}_n - \mathbf{x}\right)_{\leftarrow m}\|_\infty m^{1/2} + \|\mathbf{u}_{\to m}\|_2 \left(r_n + r\right) \\ &\leq 2\varepsilon.\end{aligned}$$

Since ε is arbitrary, this shows $(\mathbf{x}_n, r_n) \to (\mathbf{x}, r)$ in $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}_+$ when $\mathbb{R}^{\mathbb{N}}$ is equipped with the weak-ℓ^2 topology.

On the other hand, suppose $(\mathbf{x}_n, r_n) \to (\mathbf{x}, r)$ in $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}_+$ when $\mathbb{R}^{\mathbb{N}}$ is equipped with the weak-ℓ^2 topology. Pick any $\varepsilon > 0$. Choose $m \in \mathbb{N}$ large enough such that $m^{-1/2} < \varepsilon/(2r+1)$. Choose $N \in \mathbb{N}$ large enough such that for all $n > N$, $\|\mathbf{x}_n - \mathbf{x}\|_2 < \varepsilon$ and $r_n < r+1$.Then for any $n > N$, by (6.1),

$$\begin{aligned}\|\mathbf{x}_n - \mathbf{x}\|_\infty &= \max\left(\|(\mathbf{x}_n)_{\leftarrow m} - \mathbf{x}_{\leftarrow m}\|_\infty, \|(\mathbf{x}_n)_{\to m} - \mathbf{x}_{\to m}\|_\infty\right) \\ &\leq \max\left(\|(\mathbf{x}_n)_{\leftarrow m} - \mathbf{x}_{\leftarrow m}\|_2, m^{-1/2}\left(r_n + r\right)\right) \\ &< 2\varepsilon.\end{aligned}$$

Since ε is arbitrary, this implies $(\mathbf{x}_n, r_n) \to (\mathbf{x}, r)$ in $\mathcal{X}$.

Claim 3. By definition, for all $(\mathbf{x}, r) \in \mathcal{X}$, $|x_i| \leq \|\mathbf{x}\|_2 \leq r$. Therefore, $\{(\mathbf{x}, r) \in \mathcal{X} : r \leq A\}$ is a closed subset of $\prod_{i=1}^\infty [-A, A] \times [0, A]$. By item (1), the topology of $\mathcal{X}$ is equivalent to $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}_+$ where $\mathbb{R}^{\mathbb{N}}$ is equipped with the product topology, $\prod_{i=1}^\infty [-A, A] \times [0, A]$ is compact by Tychonoff's Theorem. Thus, as a closed subset of $\prod_{i=1}^\infty [-A, A] \times [0, A]$, $\{(\mathbf{x}, r) \in \mathcal{X} : r \leq A\}$ is compact. □

6.2. **Rate Function.**

Proof of Lemma 3.2. Step 1. We first show that the following inequality holds for all $(\mathbf{w}, r, s) \in \mathcal{X}$:

$$I(\mathbf{w}, r, s) \geq \Lambda_m^* \left(\mathbf{w}_{\leftarrow}, \sqrt{r^2 - \|\mathbf{w}_{\leftarrow}\|_2^2}, s \right). \tag{6.2}$$

If $\mathbf{w}_{\rightarrow} = \mathbf{0}$, then the above inequality becomes an equality by the definition of Λ_m^*. Now suppose $r > \|\mathbf{w}_{\leftarrow}\|_2$.

Pick any $\mathbf{v} \in \mathbb{R}^m$ and $b \geq 0$. Let $\mathbf{u} = \left(\mathbf{v}, b\mathbf{w}_{\rightarrow}/\sqrt{r^2 - \|\mathbf{w}_{\leftarrow}\|_2^2}\right)$. Then by the definition of Λ^*,

$$\begin{aligned}
I(\mathbf{w}, r, s) &= \Lambda^* \left(\mathbf{w}, \sqrt{r^2 - \|\mathbf{w}\|_2^2}, s \right) \\
&\geq \langle \mathbf{u}, \mathbf{w} \rangle + \left(\frac{b\sqrt{r^2 - \|\mathbf{w}\|_2^2}}{\sqrt{r^2 - \|\mathbf{w}_{\leftarrow}\|_2^2}} \right) \sqrt{r^2 - \|\mathbf{w}\|_2^2} + cs - \Lambda \left(\mathbf{u}, \frac{b\sqrt{r^2 - \|\mathbf{w}\|_2^2}}{\sqrt{r^2 - \|\mathbf{w}_{\leftarrow}\|_2^2}}, c \right) \\
&= \langle \mathbf{v}, \mathbf{w}_{\leftarrow} \rangle + b\sqrt{r^2 - \|\mathbf{w}_{\leftarrow}\|_2^2} + cs - \Lambda_m(\mathbf{v}, b, c),
\end{aligned}$$

where we used $\Lambda(\mathbf{u}, b', c) = \Lambda_m \left(\mathbf{u}_{\leftarrow}, \sqrt{(b')^2 + \|\mathbf{u}_{\rightarrow}\|_2^2}, c \right)$ in the last step. Since $\mathbf{v}$ and $b \geq 0$ are arbitrary, (6.2) holds by the definition of Λ_m^*.

Step 2. We next claim that I is lower semicontinuous.

First note that Λ_m^* is lower semicontinuous, since it is defined by a supremum over a set of continuous functions. By Lemma 3.1, the topology on $\mathcal{X}$ is equivalent to the topology on $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}_+$, where $\mathbb{R}^{\mathbb{N}}$ is equipped with the weak-ℓ^2 topology. Therefore, Λ_* is also defined by a supremum over a set of continuous functions, which is lower semicontinuous. Let $(\mathbf{w}^{(n)}, r^{(n)}, s^{(n)})$ be a sequence converging to $(\mathbf{w}, r, s)$ in $\mathcal{X} \times \mathbb{R}_+$. By (6.2) and the lower semicontinuity of Λ_m^*, we have

$$\begin{aligned}
\liminf_{n\to\infty} I\left(\mathbf{w}^{(n)}, r^{(n)}, s^{(n)}\right) &\geq \liminf_{n\to\infty} \Lambda_m^* \left(\mathbf{w}_{\leftarrow}^{(n)}, \sqrt{\left(r^{(n)}\right)^2 - \left\|\mathbf{w}_{\leftarrow}^{(n)}\right\|_2^2}, s^{(n)} \right) \\
&\geq \Lambda_m^* \left(\mathbf{w}_{\leftarrow}, \sqrt{r^2 - \|\mathbf{w}_{\leftarrow}\|_2^2}, s \right) \\
&= \Lambda^* \left((\mathbf{w}_{\leftarrow}, \mathbf{0}), \sqrt{r^2 - \|\mathbf{w}_{\leftarrow}\|_2^2}, s \right).
\end{aligned}$$

By passing to the $m \to \infty$ limit in the previous inequality (note that $\mathbf{w}_{\leftarrow}$ is the abbreviation for $\mathbf{w}_{\leftarrow m}$) and using the lower semicontinuity of Λ^*, the lower semicontinuity of I follows.

Step 3. We claim $\{I \leq \alpha\}$ is compact for any $\alpha \geq 0$.

Since I is lower semicontinuous, $\{I \leq \alpha\}$ is closed. We will show that

$$\{I \leq \alpha\} \subset \{r \leq A\} \cap \{s \leq B\} \tag{6.3}$$

for some $A, B > 0$, and that the latter set is compact by Lemma 3.1. This will prove the claim, since closed subsets of compact sets are compact.

To show (6.3) we suppose, towards a contradiction, that there exists a sequence

$$\{(\mathbf{w}^{(n)}, r^{(n)}, s^{(n)})\}_{n=1}^{\infty}, \qquad (\mathbf{w}^{(n)}, r^{(n)}, s^{(n)}) \in \{I \leq \alpha\}$$

such that $r^{(n)} \to \infty$ or $s^{(n)} \to \infty$. Fix $0 < c < T$, where $\mathbb{R} \times (0 - \infty, T)$ is the domain of $\Lambda(t_1, t_2)$ from Assumption 3. By the assumptions on Λ,

$$\Lambda_m(\mathbf{v}, b, 0) = \mathbb{E}\left[\Lambda\left(\sum_{i=1}^{m} v_i g_i + b g_0, 0\right)\right] \leq C\left(1 + \|(\mathbf{v}, b)\|_2^2\right), \tag{6.4}$$

for some absolute constant C.

Suppose first that $r^{(n)} \to \infty$. By (6.2) and (6.4),

$$\begin{aligned} I\left(\mathbf{w}^{(n)}, r^{(n)}, s^{(n)}\right) &\geq \Lambda_m^*\left(\mathbf{w}_{\leftarrow}^{(n)}, \sqrt{\left(r^{(n)}\right)^2 - \|\mathbf{w}_{\leftarrow}^{(n)}\|_2^2}, s^{(n)}\right) \\ &= \sup_{\mathbf{v},b,c}\left\{\langle \mathbf{v}, \mathbf{w}_{\leftarrow}^{(n)}\rangle + b\sqrt{\left(r^{(n)}\right)^2 - \|\mathbf{w}_{\leftarrow}^{(n)}\|_2^2} + c s^{(n)} - \Lambda_m(\mathbf{v}, b, c)\right\} \\ &\geq \sup_{\mathbf{v},b}\left\{\langle \mathbf{v}, \mathbf{w}_{\leftarrow}^{(n)}\rangle + b\sqrt{\left(r^{(n)}\right)^2 - \|\mathbf{w}_{\leftarrow}^{(n)}\|_2^2} - C\left(1 + \|(\mathbf{v}, b)\|_2^2\right)\right\}, \end{aligned}$$

where the second inequality follows by fix $c = 0$. Note that

$$\langle \mathbf{v}, \mathbf{w}_{\leftarrow}^{(n)}\rangle + b\sqrt{\left(r^{(n)}\right)^2 - \|\mathbf{w}_{\leftarrow}^{(n)}\|_2^2} \leq r^{(n)}\|(\mathbf{v}, b)\|_2,$$

and equality holds when $(\mathbf{v}, b) = \alpha\left(\mathbf{w}_{\leftarrow}^{(n)}, \sqrt{\left(r^{(n)}\right)^2 - \|\mathbf{w}_{\leftarrow}^{(n)}\|_2^2}\right)$ for some $\alpha \in \mathbb{R}$. Therefore

$$\begin{aligned} &\sup_{\mathbf{v},b}\, \langle \mathbf{v}, \mathbf{w}_{\leftarrow}^{(n)}\rangle + b\sqrt{\left(r^{(n)}\right)^2 - \|\mathbf{w}_{\leftarrow}^{(n)}\|_2^2} - C\left(1 + \|(\mathbf{v}, b)\|_2^2\right) \\ =&\sup_{\mathbf{v},b}\, r^{(n)}\|(\mathbf{v}, b)\|_2 - C\|(\mathbf{v}, b)\|_2^2 - C = \frac{\left(r^{(n)}\right)^2}{4C} - C \to \infty. \end{aligned}$$

To get the last equality, we compute the supremum directly as a function of $\|(\mathbf{v}, b)\|_2$. This contradicts the fact that $(\mathbf{w}^{(n)}, r^{(n)}, s^{(n)}) \in \{I \leq \alpha\}$.

Next, suppose $s^{(n)} \to \infty$. By (6.2),

$$I\left(\mathbf{w}^{(n)}, r^{(n)}, s^{(n)}\right) \geq \Lambda_m^*\left(\mathbf{w}_{\leftarrow}^{(n)}, \sqrt{\left(r^{(n)}\right)^2 - \|\mathbf{w}_{\leftarrow}^{(n)}\|_2^2}, s^{(n)}\right) \geq c s^{(n)} - C,$$

and $cs^{(n)} \to \infty$ if c is chosen to be positive. Again, this contradicts the assumption that $(\mathbf{w}^{(n)}, r^{(n)}, s^{(n)}) \in \{I \leq \alpha\}$. The proof is complete. □

6.3. **Exponential Tightness.** We start by recalling the Hanson–Wright concentration inequality for quadratic forms. We recall that the sub-Gaussian norm of a random variable X is defined as

$$\|X\|_{\psi_2} = \inf\{t > 0 : \mathbb{E}[\exp(X^2/t^2)] \leq 2\}. \tag{6.5}$$

Theorem 6.1 ([27, Theorem 6.2.1])**.** *Let $X = (X_1, \dots, X_n)$ be a random vector with i.i.d., mean zero, sub-Gaussian entries. Let A be an $n \times n$ matrix. Then for every $t \geq 0$,*

$$\mathbb{P}\left(\left|X^\mathsf{T} AX - \mathbb{E}X^\mathsf{T} AX\right| \geq t\right) \leq 2\exp\left[-c\min\left(\frac{t^2}{K^4\|A\|_F^2}, \frac{t}{K^2\|A\|_2}\right)\right], \tag{6.6}$$

where $K = \|X\|_{\psi_2}$ is the sub-Gaussian norm of X_1.

Proof of Lemma 3.4. We have

$$W_n = \frac{1}{n}X^\mathsf{T}\mathbf{a}\mathbf{a}^\mathsf{T}X = \frac{1}{n}\sum_{i,j=1}^n \beta_{ij}X_iX_j \tag{6.7}$$

with

$$\beta_{ij} = \sum_{\ell=1}^{k_n} a_{j\ell}a_{i\ell}. \tag{6.8}$$

Set $\beta = \frac{1}{n}\mathbf{a}\mathbf{a}^\mathsf{T}$. Then

$$\mathbb{E}\left[X^\mathsf{T}\beta X\right] = \frac{1}{n}\sum_{i=1}^n \beta_{ii}\mathbb{E}[X_i^2] = \frac{1}{n}\sum_{\ell}\sum_{i=1^n} a_{i\ell}^2 = \frac{k_n}{n}. \tag{6.9}$$

We write the matrix β as a sum of rank one matrices, $\beta = n^{-1}\sum_{\ell=1}^{k_n} Y_\ell Y_\ell^\mathsf{T}$, where

$$Y_\ell = \begin{bmatrix} a_{1\ell} \\ a_{2\ell} \\ \vdots \\ a_{n\ell} \end{bmatrix}. \tag{6.10}$$

The, using orthogonality of the Y_ℓ,

$$\|\beta\|_F^2 = \operatorname{Tr}\beta\beta^\mathsf{T} \tag{6.11}$$

$$= n^{-2}\operatorname{Tr}\left(\sum_{\ell=1}^{k_n} Y_\ell Y_\ell^\mathsf{T}\right)\left(\sum_{\ell=1}^{k_n} Y_\ell Y_\ell^\mathsf{T}\right)^\mathsf{T} \tag{6.12}$$

$$= n^{-2}\operatorname{Tr}\sum_{\ell=1}^{k_n} Y_\ell Y_\ell^\mathsf{T} \tag{6.13}$$

$$= n^{-2}\operatorname{Tr}\beta = n^{-2}k_n. \tag{6.14}$$

Further,

$$\|\beta\|_2^2 = \max_{\|v\|=1} \|\beta v\|_2^2 \tag{6.15}$$

$$= \max_{\|v\|=1} (\beta v)^{\mathsf{T}}(\beta v) \tag{6.16}$$

$$= \max_{\|v\|=1} v^{\mathsf{T}}\beta^{\mathsf{T}}\beta v \tag{6.17}$$

$$= \max_{\|v\|=1} n^{-2}\sum_{\ell=1}^{k_n} v^{\mathsf{T}}Y_\ell Y_\ell^{\mathsf{T}} v \tag{6.18}$$

$$= \max_{\|v\|=1} n^{-2}\langle Y_\ell, v\rangle^2. \tag{6.19}$$

The last quantity is equal to n^{-2} because the Y_ℓ are orthonormal. We conclude using Theorem 6.1 that

$$\mathbb{P}\left(\left|\|n^{-1/2}\mathbf{a}^{\mathsf{T}}X^{(n)}\|_2^2 - \frac{k_n}{n}\right| \geq t\right) \leq 2\exp\left[-c\min\left(k_n^{-1}(nt)^2, nt\right)\right]. \tag{6.20}$$

for any $\mathbf{a} \in \mathbb{V}_{n,k_n}$. This implies the conclusion. □

Proof of Lemma 3.5. Pick any $M > 0$ and fix $\lambda > 0$ such that $\Lambda(0,\lambda) < \infty$. We note that such a λ exists by Assumption (3). Fix $\alpha > 0$ satisfying $\alpha > (\Lambda(0,\lambda)+M)/\lambda$. By Markov's inequality, we have

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \eta\left(X_i\right) \geq \alpha\right) \leq e^{-n\lambda\alpha}\left(\mathbb{E}\left[\exp\left(\lambda\eta(X_1)\right)\right]\right)^n.$$

Therefore

$$\frac{1}{n}\log\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \eta\left(X_i\right) \geq \alpha\right) \leq -\lambda\alpha + E\left[\exp\left(\lambda\eta(X_1)\right)\right] < -M,$$

which proves the exponential tightness of $\frac{1}{n}\sum_{i=1}^n \eta\left(X_i\right)$. □

6.4. **Preliminary Lemmas for Gaussian Approximation.** Let μ be the standard Gaussian measure on $\mathbb{R}$, corresponding to a Gaussian random variable with mean zero and variance one. For $k, n \in \mathbb{N}$ and any $\mathbf{a} \in \mathbb{V}_{n,k}$, we let $\mathbf{a}_i$ be the i-th row of $\mathbf{a}$. For any choice of $\mathbf{u} \in \mathbb{R}^k$, let $\mu_n(\mathbf{u})$ be the measure on $\mathbb{R}$ given by

$$\mu_n = \mu_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^n \delta_{\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\rangle}. \tag{6.21}$$

Lemma 6.2. *Fix $k, n \in \mathbb{N}$ and $\mathbf{u} \in \mathbb{R}^k$. Let $\mathbf{a} \in \mathbb{V}_{n,k}$ be random and distributed according to the Haar measure. For every $m \in \mathbb{N}$, we have*

$$\mathbb{E}\left[\int x^m \, d\mu_n\right] = \|\mathbf{u}\|_2^m \left(\int x^m \, d\mu\right)\left(1 + O\left(\frac{1}{n}\right)\right), \tag{6.22}$$

where the implicit constant depends only on m.

Proof. We just treat the even m case. The odd m case is clear by symmetry, as the moments on the left and right side of the equality both vanish. By exchangeability of the rows of $\mathbf{a}$,

$$\int x^m \, d\mu_n = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}\left(\sum_{j=1}^{k} u_j(\sqrt{n}a_{ij})\right)^m\right] = \mathbb{E}\left[\left(\sum_{j=1}^{k} u_j(\sqrt{n}a_{1j})\right)^m\right]. \tag{6.23}$$

We recognize this quantity as

$$\mathbb{E}\left[\langle \mathbf{u}, \sqrt{n}\mathbf{a}_1\rangle^m\right] = \|\mathbf{u}\|_2^m \cdot \mathbb{E}\left[(\sqrt{n}\mathbf{a}_{11})^m\right]. \tag{6.24}$$

where the equality is justified by the rotational invariance of m. By Wick's theorem (Theorem B.5) and Lemma B.6,

$$\mathbb{E}\left[(\sqrt{n}\mathbf{a}_{11})^m\right] = \left(\int x^m \, d\mu\right)\left(1 + O\left(\frac{1}{n}\right)\right). \tag{6.25}$$

This completes the proof. □

The next lemma makes use of a concentration inequality due to Gromov, which we recall as follow:

Theorem 6.3 ([2, Corollary 4.4.28])**.** *Let $f\colon O(n) \to \mathbb{R}$ be a Lipschitz function with Lipschitz constant L in the Hilbert–Schmidt norm, meaning*

$$|g(X) - g(Y)| \le L\|X - Y\|_{\mathrm{HS}} \tag{6.26}$$

for all $X, Y \in O(n)$. Let $\mathbb{P}$ be the normalized Haar measure on $O(n)$ and let $\mathbb{E}$ be the corresponding expectation. There is a constant $c > 0$ such that

$$\mathbb{P}\left(|g(X) - \mathbb{E}[g(X)] \ge \delta L n^{-1/2}\right) \le 2\exp\left(-c\delta^2\right) \tag{6.27}$$

for all $\delta \ge 0$.

Lemma 6.4. *Fix $0 \le r \le 2$ and an increasing sequence of positive integers $\{k_n\}_{n=1}^{\infty}$. For any $\varepsilon > 0$, there exists $R = R(\varepsilon, \|\mathbf{u}\|_2) > 2$ and constants $C_1 = C_1(\varepsilon, \|\mathbf{u}\|_2)$ such that*

$$\int_{|x|>R} |x|^r \, d\mu_n \le \varepsilon \tag{6.28}$$

with probability at least $1 - 2\exp(-C_1 n)$.

Proof. Since $|x|^r \le x^2$ when $|x| > 2$, it suffices to show that there exists $R = R(\varepsilon) > 2$ and constants $C_1 = C_1(\varepsilon)$, such that

$$\int_{|x|>R} x^2 d\mu_n \le \varepsilon$$

holds with the required probability.

Let $\varphi(x)$ be a smooth increasing function such that

$$0 \le \varphi \le 1, \quad \varphi(x) = 1 \text{ if } |x| \ge 1, \quad \varphi(x) = 0 \text{ if } |x| \le 0, \quad \varphi'(x) \le 2$$

Let $R>2$ be chosen later, define

$$\varphi_R(x)=\varphi(x-R+1).$$

By invariance of Haar measure under orthogonal transformations, we have

$$\begin{aligned}\sigma_n\left(\int_{|x|>R}x^2d\mu_n>\varepsilon\right)&=\sigma_n\left(\frac{1}{n}\sum_{i=1}^n\left\langle\mathbf{u},\sqrt{n}\mathbf{a}_i\right\rangle^2\mathbf{1}_{|\langle\mathbf{u},\sqrt{n}\mathbf{a}_i\rangle|>R}>\varepsilon\right)\\&=\sigma_n\left(\frac{\|\mathbf{u}\|_2^2}{n}\sum_{i=1}^n\left(\sqrt{n}a_{i1}\right)^2\mathbf{1}_{|\sqrt{n}a_{i1}|>R}>\varepsilon\right).\end{aligned}$$

If $\|\mathbf{u}\|_2=0$, then μ_n is the dirac measure at 0, and

$$\int_{|x|>R}x^2d\mu_n=0<\varepsilon,$$

with probability 1 for all $R>0$.

Now suppose $\|\mathbf{u}\|_2\neq 0$. Let $f:\mathbb{R}^{n\times k_n}\to\mathbb{R}$ be defined by

$$f_R\left(\mathbf{a}\right):=\frac{1}{n}\sum_{i=1}^n\left(\sqrt{n}a_{i1}\right)^2\varphi_R\left(\sqrt{n}a_{i1}\right).$$

By definition of φ_R,

$$f_R(\mathbf{a})\geq\frac{1}{n}\sum_{i=1}^n\left(\sqrt{n}a_{i1}\right)^2\mathbf{1}_{|\sqrt{n}a_{i1}|>R},$$

therefore,

$$\begin{aligned}\sigma_n\left(\int_{|x|>R}x^2d\mu_n>\varepsilon\right)=&\sigma_n\left(\frac{\|\mathbf{u}\|_2^2}{n}\sum_{i=1}^n\left(\sqrt{n}a_{i1}\right)^2\mathbf{1}_{|\sqrt{n}a_{i1}|>R}>\varepsilon\right)\\\leq&\sigma_n\left(f_R(\mathbf{a})>\frac{\varepsilon}{\|\mathbf{u}\|_2^2}\right).\end{aligned}\tag{6.29}$$

Note that

$$|\partial_{i1}f_R(\mathbf{a})|=|2a_{i1}\varphi_R'\left(\sqrt{n}a_{i1}\right)|\leq 4,$$

where we use $|a_{i1}|\leq 1$ and $\varphi_R'(x)\leq 2$ in the inequality. This implies

$$\|f_R\|_{\mathrm{Lip}}\leq 4.$$

Moreover,

$$\begin{aligned}\mathbb{E}_n\left[f_R(\mathbf{a})\right] &= \mathbb{E}_n\left[\left(\sqrt{n}a_{11}\right)^2 \varphi_R\left(\sqrt{n}a_{11}\right)\right] \\ &\le \mathbb{E}_n\left[\left(\sqrt{n}a_{11}\right)^2 \mathbf{1}_{|\sqrt{n}a_{11}|>R-1}\right] \\ &= \int_0^\infty \sigma_n\left(\left(\sqrt{n}a_{11}\right)^2 \mathbf{1}_{|\sqrt{n}a_{11}|>R-1} > x\right) dx \\ &\le 2R^2 \exp\left(-CR^2\right) + \int_{(R-1)^2}^\infty \sigma_n\left(|\sqrt{n}a_{11}| > \sqrt{x}\right) dx \\ &\le 4R^2 \exp\left(-C'R^2\right),\end{aligned}$$

for some absolute constants C', where the last two inequalities follow from the fact that $\sqrt{n}a_{11}$ is sub-Gaussian with $\|\sqrt{n}a_{11}\|_{\psi_2}$ is bounded by an absolute constant.

Choose $R > 2$ large enough such that

$$4R^2 \exp\left(-C'R^2\right) < \frac{\varepsilon}{2\|\mathbf{u}\|_2^2}.$$

Theorem 6.3 implies that

$$\begin{aligned}\sigma_n\left(f_R(\mathbf{a}) > \frac{\varepsilon}{\|\mathbf{u}\|_2^2}\right) &\le \sigma_n\left(|f_R(\mathbf{a}) - \mathbb{E}_n\left[f_R(\mathbf{a})\right]| > \frac{\varepsilon}{2\|\mathbf{u}\|_2^2}\right) \\ &\le 2\exp\left(-\frac{Cn\varepsilon}{2\|f_R\|_{\mathrm{Lip}}\|\mathbf{u}\|_2^2}\right) \\ &\le 2\exp\left(-C_1 n\right).\end{aligned} \tag{6.30}$$

The proof is complete by combining (6.29) and (6.30).

□

6.5. **Proof of Gaussian Approximation.**

Proof of Lemma 3.6. We recall the definition

$$\Lambda(t_1, t_2) = \log \mathbb{E} \exp\left(t_1 X_1 + t_2 \eta(X_1)\right). \tag{6.31}$$

We set $t_2 = s$ and define

$$\Lambda(t) = \log \mathbb{E} \exp\left(t X_1 + c\eta(X_1)\right) \tag{6.32}$$

Let $R > 1$ be a parameter, and set

$$\Lambda(t) = \varphi_R(t) + \psi_R(t), \tag{6.33}$$

where

$$\varphi_R(t) = \min(\Lambda(t), C(1 + R^2)), \qquad \psi_R(t) = \Lambda(t) - \varphi_R(t). \tag{6.34}$$

We observe that

$$|\psi_R(t)| \le C|x|^2 \mathbb{1}_{|x|>R}. \tag{6.35}$$

Consider the random measures

$$\mu_n^{(j)} = \frac{1}{n}\sum_{i=1}^{n} \delta_{\langle \mathbf{v}^{(n,j)}, \sqrt{n}\mathbf{a}_i\rangle}. \tag{6.36}$$

For all j, n, we have

$$\left|\int \Lambda(t)\, d\mu_n^{(j)} - \int \Lambda(t)\, d\mu\right| \le \left|\int \varphi_R\, d\mu_n^{(j)} - \int \varphi_R\, d\mu\right| \tag{6.37}$$

$$+ C\int_{|x|>R} |x|^2\, d\mu_n^{(j)} \tag{6.38}$$

$$+ C\int_{|x|>R} |x|^2\, d\mu. \tag{6.39}$$

For the last term, the standard Gaussian tail bound shows there exists a constant $C_0 > 0$ such that

$$\int_{|x|>R} |x|^2\, d\mu \le C_0 R^2 \exp(-C_0^{-1} R). \tag{6.40}$$

We note that the left side of this equation is a deterministic quantity.

By Lemma 6.4, we have

$$\mathbb{P}\left(\int_{|x|>R} |x|^2\, d\mu_n^{(j)} \ge \varepsilon/3\right) \le 2\exp\left(-C_1 n\right), \tag{6.41}$$

for $C_1 = C_1(\varepsilon, D)$, by taking R large depending on ε and D.

It remains to control

$$\left|\int \varphi_R\, d\mu_n^{(j)} - \int \varphi_R\, d\mu\right| \tag{6.42}$$

with exponentially high probability. We will first show convergence for the mean

$$\left|\int \varphi_R\, \mathbb{E}[d\mu_n^{(j)}] - \int \varphi_R\, \mathbb{E}[d\mu],\right| \tag{6.43}$$

then show concentration about this mean.

Consider the sequence of (deterministic) measures

$$\mathbb{E}[\mu_1^{(1)}], \mathbb{E}[\mu_1^{(2)}], \ldots, \mathbb{E}[\mu_2^{(1)}], \mathbb{E}[\mu_2^{(2)}], \ldots, \mathbb{E}[\mu_n^{(1)}], \mathbb{E}[\mu_n^{(2)}], \ldots, \tag{6.44}$$

which is the concatenation of the sequences $s_n = \{\mathbb{E}[\mu_n^{(j)}]\}_{j=1}^{d_n}$ for all n. By Lemma 6.2, the corresponding sequence of k-th moments converges to the k-th moment of the Gaussian for all $k \in \mathbb{N}$. Further, the Gaussian measure is characterized by its moments by the Carleman criterion [4, Lemma B.3]. Then a standard argument shows this sequence converges weakly to the Gaussian measure [4, Lemma B.1]. In particular, since φ_R is a continuous and bounded function, we have

$$\lim_{n\to\infty} \sup_j \int \varphi_R\, d\mu_n^{(j)} = \int \varphi_R\, d\mu. \tag{6.45}$$

By our hypothesis on the X_i, $\varphi_R(t)$ is increasing in t. This implies that outside of a compact interval, it is equal to $C(1+R^2)$. Then we see that $\varphi_R(t)$ Lipschitz with constant $L = L(R)$.

We now use Gromov's inequality to complete the proof. For any $\mathbf{u} \in \mathbb{R}^k$,

$$\int \varphi_R \, d\mu_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^{n} \varphi_R(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i \rangle). \tag{6.46}$$

Now we compute the Lipschitz constant of this quantity considered as a function of the matrix $\mathbf{a}$:

$$\frac{1}{n} \sum_{i=1}^{n} \left| \varphi_R(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i \rangle) - \varphi_R(\langle \mathbf{u}, \sqrt{n}\mathbf{b}_i \rangle) \right| \tag{6.47}$$

$$\leq \frac{1}{n^{1/2}} \sum_{i=1}^{n} L \left| \langle \mathbf{u}, \mathbf{a}_i \rangle - \langle \mathbf{u}, \mathbf{b}_i \rangle \right| \tag{6.48}$$

$$\leq \frac{L}{n^{1/2}} \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{k} |a_{ij} - b_{ij}|^2} \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{k} |u_j|^2} \tag{6.49}$$

$$\leq L \|\mathbf{u}\|_2 \|A - B\|_{\mathrm{HS}}. \tag{6.50}$$

Applying Theorem 6.3 shows that

$$\mathbb{P}\left(\left| \int \varphi_R \, d\mu_n^{(j)} - \int \varphi_R \, \mathbb{E}[d\mu_n^{(j)}] \right| \geq \delta D L \right) \leq \exp\left(-c\delta^2 n\right) \tag{6.51}$$

for all $\delta > 0$. We choose δ such that $\delta D L < \varepsilon/3$.

Now the conclusion follows from first combining (6.37), (6.40), (6.41), and (6.51), then taking a union bound over all j, using the assumed bound on d_n and finally applying the Borel–Cantelli lemma. □

7. Preliminary Lemmas: Lower Bound

7.1. **Proof of Gradient Approximation.** Let $m \in \mathbb{N}$ and $(\mathbf{v}, b, c) \in \mathbb{R}^m \times \mathbb{R} \times (-\infty, T)$ be fixed throughout this subsection, where T is the constant from Assumption 2. We use $\mathbb{P}$ and $\mathbb{E}$ to denote the probability and expectation with respect to this sequence, and we use σ_n and $\mathbb{E}_n$ to denote the probability and expectation with respect to the Haar measure on the n-dimensional orthogonal group.

Lemma 7.1. *Let X_1 satisfies Assumption 2 through 4 and let μ denote the law of X_1. Fix $(s_1, s_2) \in \mathbb{R} \times (-\infty, T)$. Define the measure $\tilde{\mu}$ by*

$$\frac{d\tilde{\mu}}{d\mu}(x) = \exp(s_1 x + s_2 \eta(x) - \Lambda(s_1, s_2)). \tag{7.1}$$

Let $\tilde{\mathbb{P}}$ and $\tilde{\mathbb{E}}$ denote the corresponding probability and expectation under $\tilde{\mu}$. Then there exists a constant $C = C(s_2) > 0$ such that

$$\tilde{\mathbb{P}}\left(|X_1 - \tilde{\mathbb{E}}[X_1]| > t\right) \le 2\exp(-t^2/C^2).$$

Proof. By a standard argument using Markov's inequality (see, for example, [27, Proposition 2.5.2]), it suffices to show that

$$\tilde{\mathbb{E}}[\exp(\lambda(X_1 - \tilde{\mathbb{E}}X_1))] \le \exp(C^2\lambda^2),$$

Observe that $\tilde{\mathbb{E}}[X_1] = \partial_1 \Lambda(s_1, s_2)$. Then

$$\begin{aligned}\tilde{\mathbb{E}}[\exp(\lambda(X_1 - \tilde{\mathbb{E}}[X_1]))] &= e^{-\lambda \tilde{\mathbb{E}}X_1 - \Lambda(s_1,s_2)} \int \exp(\lambda x)\exp(s_1 x + s_2\eta(x))\, d\mu(x) \\ &= \exp(\Lambda(s_1+\lambda, s_2) - \Lambda(s_1,s_2) - \lambda\partial_1\Lambda(s_1,s_2)) \\ &\le \exp\left(C^2\lambda^2\right),\end{aligned}$$

where the inequality follows from Assumption 4. □

Let $\{k_n\}_{n=1}^\infty$ be an increasing sequence of positive integers such that $\lim_{n\to\infty} k_n = \infty$. Let $m \in \mathbb{N}$, $\mathbf{v} \in \mathbb{R}^m$ and $b \in \mathbb{R}$ be given. We define an auxiliary sequence of positive integers $\{m_n\}_{n=1}^\infty$ and a sequence of vectors $\{\mathbf{u}^{(n)}\}_{n=1}^\infty$, with $\mathbf{u}^{(n)} \in \mathbb{R}^{k_n}$ in the following way:

$$\begin{aligned} m_n &:= \min\left(k_n - m, n^{1/2}\right), \\ \mathbf{u}^{(n)} &:= (\mathbf{v}, \mathbf{b}^{(n)}, \mathbf{0}^{(n)}), \end{aligned} \tag{7.2}$$

where $\mathbf{b}^{(n)} = b m_n^{-1/2}\mathbf{1}_{m_n} \in \mathbb{R}^{m_n}$ and $\mathbf{0}^{(n)} = \mathbf{0}_{k_n-m-m_n} \in \mathbb{R}^{k_n-m-m_n}$.

For $j \in \{1, \dots, m+m_n\}$, define[3]

$$\nu_n^{(j)} = \frac{1}{n}\sum_{i=1}^n \delta(\sqrt{n}a_{ij}, \langle \mathbf{u}^{(n)}, \sqrt{n}\mathbf{a}_i\rangle), \tag{7.3}$$

and define

$$\nu_n = \frac{1}{n}\sum_{i=1}^n \delta\left(m_n^{-1/2}\sum_{j=m+1}^{m+m_n}\sqrt{n}a_{ij}, \langle \mathbf{u}^{(n)}, \sqrt{n}\mathbf{a}_i\rangle\right).$$

Further, define

$$\nu^{(j)} = \mathrm{Law}\left(g_j, \sum_{l=1}^m v_l g_l + b g_0\right), \quad \nu = \mathrm{Law}\left(g_0, \sum_{l=1}^m v_l g_l + b g_0\right) \tag{7.4}$$

where $g_0, \dots, g_m$ are independent, mean zero, variance one Gaussian random variables. Here $\mathbf{u}^{(n)}$ and $\mathbf{v} = (v_1, \dots, v_m)$ are the vectors chosen above.

[3] We sometimes write $\delta(x)$ instead of δ_x in this subsection for legibility. The difference is only notational.

Lemma 7.2. *Fix $C, R > 0$, and let $h_R : \mathbb{R} \to \mathbb{R}$ be a function satisfying $|h_R(x)| \le CR^{-1}x^2$ for all $x \in \mathbb{R}$. Then*

$$\left| \mathbb{E}_n \left[\int x h_R(y) \, d\nu_n^{(j)}(x,y) \right] \right| \le \frac{C_1 \|\mathbf{u}\|_2^2}{R}, \tag{7.5}$$

for $j = 1, \dots, m$, and

$$\left| \mathbb{E}_n \left[\int x h_R(y) \, d\nu_n^{(j)}(x,y) \right] \right| \le \frac{C_1 m_n^{-1/2} \|\mathbf{u}\|_2^2}{R}, \tag{7.6}$$

for $j = m+1, \dots, m+m_n$, where the constant $C_1 > 0$ depends only on C. We also have the deterministic bounds

$$\int_{|x| \ge R} |x| \, d\nu_n^{(j)}(x,y) \le \frac{1}{R}, \tag{7.7}$$

for $j = 1, \dots, m+m_n$, and

$$\int_{|x| \ge R} |x| \, d\nu_n(x,y) \le \frac{1}{R}. \tag{7.8}$$

Proof. For $j = 1, \dots, m$,

$$\begin{aligned}
\left(\mathbb{E}_n \left[\sum_{i=1}^n a_{ij} h_R\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i \rangle \right) \right] \right)^2 &\le \mathbb{E}_n \left[\left(\sum_{i=1}^n a_{ij} h_R\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i \rangle \right) \right)^2 \right] \\
&\le \mathbb{E}_n \left[\sum_{i=1}^n h_R\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i \rangle \right)^2 \right] \\
&= n\mathbb{E}_n \left[h_R\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_1 \rangle \right)^2 \right] \\
&= n\mathbb{E}_n \left[h_R\left(\|\mathbf{u}\|_2 \sqrt{n} a_{11} \right)^2 \right] \\
&\le Cn\|\mathbf{u}\|_2^4 \mathbb{E}_n \left[\frac{(\sqrt{n}a_{11})^4}{R^2} \right] \\
&\le \frac{C_1 n \|\mathbf{u}\|_2^4}{R^2},
\end{aligned}$$

where the first two inequalities use Cauchy's inequality, the second equality uses the rotational invariance of Haar measure, and the last inequality uses the fact that $\sqrt{n}a_{11}$ is sub-Gaussian with $\|\sqrt{n}a_{11}\|_{\psi_2}$ bounded by a constant independent of n [27, Theorem 3.4.6]. Then equation (7.5) follows by the definition of $\nu_n^{(j)}$.

Now consider $j = m+1, \dots, m+m_n$. Since $u_j = bm_n^{-1/2}$ for all $j = m+1, \dots, m+m_n$, the variables

$$\left\{ \sum_{i=1}^n a_{ij} h_R\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i \rangle \right) \right\}_{j=m+1}^{m+m_n}$$

are identically distributed. Then we have

$$\begin{aligned}
\left(\mathbb{E}_n\left[\sum_{i=1}^n a_{ij} h_R\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\rangle\right)\right]\right)^2 &\le \mathbb{E}_n\left[\left(\sum_{i=1}^n a_{ij} h_R\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\rangle\right)\right)^2\right] \\
&= \frac{1}{m_n}\mathbb{E}_n\left[\sum_{j=m+1}^{m+m_n}\left(\sum_{i=1}^n a_{ij} h_R\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\rangle\right)\right)^2\right] \\
&\le \frac{1}{m_n}\mathbb{E}_n\left[\sum_{i=1}^n h_R\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\rangle\right)^2\right] \\
&= \frac{n}{m_n}\mathbb{E}_n\left[h_R\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_1\rangle\right)^2\right] \\
&= \frac{n}{m_n}\mathbb{E}_n\left[h_R\left(\|\mathbf{u}\|_2\sqrt{n}a_{11}\right)^2\right] \\
&\le \frac{Cn}{m_n}\|\mathbf{u}\|_2^4\mathbb{E}_n\left[\frac{\left(\sqrt{n}a_{11}\right)^4}{R^2}\right] \\
&\le \frac{C_1 n\|\mathbf{u}\|_2^4}{m_n R^2},
\end{aligned}$$

where the second inequality uses the fact that $\mathbf{a}$ is an orthogonal projection. (7.6) follows from the definition of $\nu_n^{(j)}$.

(7.7) can be proved by noticing

$$\int_{|x|\ge R} |x|\, d\nu_n^{(j)}(x,y) \le \frac{1}{R}\int x^2 d\nu_n^{(j)}(x,y) = \frac{1}{R}\sum_{i=1}^n a_{ij}^2 = \frac{1}{R}.$$

And similarly, (7.8) follows from

$$\left|\int_{|x|\ge R} x d\nu_n(x,y)\right| \le \frac{1}{R}\int x^2 d\nu_n(x,y) = \frac{1}{R}m_n^{-1}\sum_{i=1}^n\left(\sum_{j=m+1}^{m+m_n} a_{ij}\right)^2 \le \frac{1}{R},$$

$\square$

Lemma 7.3. *For any fixed $\alpha, \beta \in \mathbb{N}\cup\{0\}$, we have*

$$\mathbb{E}\left[\int x^\alpha y^\beta d\nu_n^j\right] = \left(\int x^\alpha y^\beta d\nu^j\right)\left(1 + O\left(\frac{1}{n}\right)\right),$$

where the implicit constant depends only on α, β.

Proof. Using Lemma B.6 in the fourth equality, we have

$$\begin{aligned}
\mathbb{E}_n\left[\int x^\alpha y^\beta d\nu_n^j\right] &= \frac{1}{n}\mathbb{E}_n\left[\sum_{i=1}^n \left(\sqrt{n}a_{ij}\right)^\alpha \cdot \left\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\right\rangle^\beta\right] \\
&= \mathbb{E}_n\left[\left(\sqrt{n}a_{1j}\right)^\alpha \cdot \left(\sum_{i=1}^{k_n} u_i\sqrt{n}a_{1i}\right)^\beta\right] \\
&= \sum_{i_1,\cdots,i_\beta=1}^{k_n} u_{i_1}\cdots u_{i_\beta} \cdot \mathbb{E}_n\left[\left(\sqrt{n}a_{1j}\right)^\alpha \cdot \sqrt{n}a_{1i_1}\cdots\sqrt{n}a_{1i_\beta}\right] \\
&= \sum_{i_1,\cdots,i_\beta=1}^{k_n} u_{i_1}\cdots u_{i_\beta} \cdot \mathbb{E}_{\mathbf{g}}\left[(g_j)^\alpha \cdot g_{i_1}\cdots g_{i_\beta}\right]\left(1+O\left(\frac{1}{n}\right)\right) \\
&= \mathbb{E}_{\mathbf{g}}\left[(g_j)^\alpha \cdot \left(\sum_{i=1}^{k_n} u_i g_i\right)^\beta\right]\left(1+O\left(\frac{1}{n}\right)\right),
\end{aligned}$$

where the implicit constant depends only on α, β.

By definition $\mathbf{u} = \left(\mathbf{v}, bm_n^{-1/2}\mathbf{1}_{m_n}, \mathbf{0}_{k_n-m_n-m}\right)$, therefore, for $j = 1, \ldots, m$,

$$\left(g_j, \sum_{i=1}^{k_n} u_i g_i\right) \overset{(d)}{=} \left(g_j, \sum_{l=1}^{m} v_l g_l + b g_0\right).$$

Thus by definition of ν^j (7.4),

$$\begin{aligned}
\mathbb{E}_n\left[\int x^\alpha y^\beta d\nu_n^j\right] &= \mathbb{E}_{\mathbf{g}}\left[(g_j)^\alpha \cdot \left(\sum_{i=1}^{k_n} u_i g_i\right)^\beta\right]\left(1+O\left(\frac{1}{n}\right)\right) \\
&= \left(\int x^\alpha y^\beta d\nu^j\right)\left(1+O\left(\frac{1}{n}\right)\right),
\end{aligned}$$

where the implicit constant depends only on α, β. □

Lemma 7.4. *For any fixed $\alpha, \beta \in \mathbb{N} \cup \{0\}$, we have*

$$\mathbb{E}\left[\int x^\alpha y^\beta d\nu_n\right] = \left(\int x^\alpha y^\beta d\nu\right)\left(1+O\left(\frac{1}{n}\right)\right),$$

where the implicit constant depends only on α, β.

Proof. Using Lemma B.6 in the fourth equality, we have

$$\begin{aligned}
\mathbb{E}_n\left[\int x^\alpha y^\beta d\nu_n\right] &= \frac{1}{n}\mathbb{E}_n\left[\sum_{i=1}^n \left(m_n^{-1/2}\sum_{j=m+1}^{m+m_n}\sqrt{n}a_{ij}\right)^\alpha \cdot \langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\rangle^\beta\right] \\
&= \mathbb{E}_n\left[\left(m_n^{-1/2}\sum_{j=m+1}^{m+m_n}\sqrt{n}a_{1j}\right)^\alpha \cdot \left(\sum_{i=1}^{k_n} u_i\sqrt{n}a_{1i}\right)^\beta\right] \\
&= m_n^{-\alpha/2}\sum_{j_1,\dots,j_\alpha=m+1}^{m+m_n}\sum_{i_1,\cdots,i_\beta=1}^{k_n} u_{i_1}\cdots u_{i_\beta}\cdot \mathbb{E}_n\left[\sqrt{n}a_{1j_1}\cdots\sqrt{n}a_{1j_\alpha}\cdot\sqrt{n}a_{1i_1}\cdots\sqrt{n}a_{1i_\beta}\right] \\
&= m_n^{-\alpha/2}\sum_{j_1,\dots,j_\alpha=m+1}^{m+m_n}\sum_{i_1,\cdots,i_\beta=1}^{k_n} u_{i_1}\cdots u_{i_\beta}\cdot \mathbb{E}_{\mathbf{g}}\left[g_{j_1}\cdots g_{j_\alpha}\cdot g_{i_1}\cdots g_{i_\beta}\right]\left(1+O\left(\frac{1}{n}\right)\right) \\
&= \mathbb{E}_{\mathbf{g}}\left[\left(m_n^{-1/2}\sum_{j=m+1}^{m+m_n} g_j\right)^\alpha \cdot \left(\sum_{i=1}^{k_n} u_i g_i\right)^\beta\right]\left(1+O\left(\frac{1}{n}\right)\right),
\end{aligned}$$

where the implicit constant depends only on α,β.

By definition $\mathbf{u} = \left(\mathbf{v}, bm_n^{-1/2}\mathbf{1}_{m_n}, \mathbf{0}_{k_n-m_n-m}\right)$, therefore,

$$\left(m_n^{-1/2}\sum_{j=m+1}^{m+m_n} g_j, \sum_{i=1}^{k_n} u_i g_i\right) \overset{(d)}{=} \left(g_0, \sum_{l=1}^{m} v_l g_l + b g_0\right).$$

Hence,

$$\begin{aligned}
\mathbb{E}_n\left[\int x^\alpha y^\beta d\nu_n\right] &= \mathbb{E}_{\mathbf{g}}\left[\left(m_n^{-1/2}\sum_{j=m+1}^{m+m_n} g_j\right)^\alpha \cdot \left(\sum_{i=1}^{k_n} u_i g_i\right)^\beta\right]\left(1+O\left(\frac{1}{n}\right)\right) \\
&= \left(\int x^\alpha y^\beta d\nu\right)\left(1+O\left(\frac{1}{n}\right)\right),
\end{aligned}$$

where the implicit constant depends only on α,β. ☐

Lemma 7.5. *A sequence of probability measures $\{\mu_n\}$ on $\mathbb{R}^2$ converges weakly to a probability measure μ if the following conditions are satisfied:*

(1) All moments of μ are finite and denote

$$\gamma_{\alpha,\beta} := \int x^\alpha y^\beta d\mu_n(x,y), \quad \forall\, \alpha,\beta \in \mathbb{N}\cup\{0\}.$$

(2) All moments of μ_n are finite and

$$\int x^\alpha y^\beta d\mu_n(x,y) \to \gamma_{\alpha,\beta}, \quad \forall\, \alpha,\beta \in \mathbb{N}\cup\{0\}.$$

(3) μ is uniquely determined by $\{\gamma_{\alpha,\beta}\}_{\alpha,\beta\in\mathbb{N}\cup\{0\}}$.

Proof. By property (1) and (2), for each polynomial P, we have

$$C_p := \sup_{n\geq 1} \int P d\mu_n < \infty.$$

Let $B_R := \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_2 \leq R\}$. Then Markov's inequality implies that

$$\mu_n\left((B_R)^c\right) \leq \frac{C_{x^2+y^2}}{R^2}.$$

Therefore, μ_n is tight. By Prokhorov's Theorem, it suffices to show that, if $\mu_{n_k} \to \nu$ weakly, then $\nu = \mu$.

Pick any polynomial P. Define φ_R to be a nonnegative continuous function such that $\mathbf{1}_{B_R} \leq \varphi_R \leq \mathbf{1}_{B_{R+1}}$. We write

$$\int P d\mu_{n_k} = \int \varphi_R P d\mu_{n_k} + \int (1-\varphi_R) P d\mu_{n_k}. \tag{7.9}$$

For the first term in the right hand side of (7.9), since μ_{n_k} weakly converge to ν,

$$\lim_{k\to\infty} \int \varphi_R P d\mu_{n_k} = \int \varphi_R P d\nu. \tag{7.10}$$

For the second term, use Cauchy's inequality, we have

$$\left|\int (1-\varphi_R) P d\mu_{n_k}\right|^2 \leq \mu_{n_k}\left((B_R)^c\right) \int P^2 d\mu_{n_k} \leq \frac{C_{x^2+y^2} C_{P^2}}{R^2}. \tag{7.11}$$

Moreover, property (2) implies that

$$\lim_{k\to\infty} \int P d\mu_{n_k} = \int P d\mu. \tag{7.12}$$

Combining (7.9)-(7.12), we have

$$\lim_{R\to\infty} \int \varphi_R P d\nu = \int P d\mu. \tag{7.13}$$

Note that P is arbitrary, replacing P by P^2 in the above display and applying monotone convergence theorem implies that

$$\int P^2 d\nu = \lim_{R\to\infty} \varphi_R P^2 d\nu = \int P d\mu.$$

Therefore, $P \in L^2(\nu) \subset L^1(\nu)$. An application of dominant convergence theorem combined with (7.13) now concludes that

$$\int P d\nu = \int P d\mu.$$

Again, since P is arbitrary, setting $P = x^\alpha y^\beta$ $(\alpha, \beta \in \mathbb{N} \cup \{0\})$, and recall property (3), we have $\mu = \nu$. This completes the proof. $\square$

Lemma 7.6. *Fix $C_0 > 0$, let $h : \mathbb{R} \to \mathbb{R}$ satisfy*

$$|h^{(d)}(x)| \le C_0(1 + |x|^{2-d})$$

for $d = 0, 1, 2$. Define

$$f_j^{(n)}\left(\mathbf{a}^{(n)}\right) = \frac{1}{n}\sum_{i=1}^{n} \sqrt{n} a_{ij} h'\left(\left\langle \mathbf{u}^{(n)}, \sqrt{n}\mathbf{a}_i \right\rangle\right) = \int x h'(y) d\nu_n^j(x, y)$$

for $j = 1, \ldots, k_n$. When the context is clear, we omit the superscript (n). Then for σ-a.e. $\mathbf{a}$,

$$\lim_{n\to\infty} f_j^{(n)}\left(\mathbf{a}^{(n)}\right) = \mathbb{E}\left[g_j h'\left(\sum_{l=1}^{m} v_l g_l + b g_0\right)\right]$$

for all $j \in \{1, \ldots, m\}$.

Proof. In this proof, we use C to denote a constant that may depend on C_0, whose value might change from line to line.

Let

$$\hat{f}_j(\mathbf{a}) = \frac{1}{n}\sum_{i=1}^{n} \sqrt{n} a_{ij} \hat{\varphi}_n\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i \rangle\right),$$

where

$$\hat{\varphi}_n(x) = \max\left(\min\left(h'(x), C_0(1 + n^{1/8}\|\mathbf{u}\|_2)\right), -C_0(1 + n^{1/8}\|\mathbf{u}\|_2)\right).$$

Then

$$\begin{aligned}&\sigma_n\left(|f_j(\mathbf{a}) - \mathbb{E}_n\left[f_j(\mathbf{a})\right]| \ge \varepsilon\right)\\ \le&\sigma_n\left(\left|f_j(\mathbf{a}) - \hat{f}_j(\mathbf{a})\right| \ge \varepsilon/3\right) + \sigma_n\left(\left|\hat{f}_j(\mathbf{a}) - \mathbb{E}_n\left[\hat{f}_j(\mathbf{a})\right]\right| \ge \varepsilon/3\right) + \sigma_n\left(\left|\mathbb{E}_n\left[\hat{f}_j(\mathbf{a})\right] - \mathbb{E}_n\left[f_j(\mathbf{a})\right]\right| \ge \varepsilon/3\right)\end{aligned}$$

For the first term, we have

$$\sigma_n\left(\left|f_j(\mathbf{a}) - \hat{f}_j(\mathbf{a})\right| > \varepsilon/3\right) \le \sigma_n\left(\max_i \left|\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\rangle\right| > n^{1/8}\|\mathbf{u}\|_2\right) \le 2n\exp\left(-Cn^{1/4}\right).$$

For the second term, we first notice that the Lipshitz constant of $\hat{f}_j$ is upper bounded by

$$C\left(1 + \|\mathbf{u}\|_2 + \|\mathbf{u}\|_\infty\right).$$

Thus by Theorem 6.3,

$$\sigma_n\left(\left|\hat{f}_j(\mathbf{a}) - \mathbb{E}_n\left[\hat{f}_j(\mathbf{a})\right]\right| \ge \varepsilon/3\right) \le 2\exp\left(-\frac{C\varepsilon^2 n}{\left(1 + \|\mathbf{u}\|_2 + \|\mathbf{u}\|_\infty\right)^2}\right).$$

For the third term, it converges to zero deterministically by (7.5), with $R = n^{1/8}\|\mathbf{u}\|_2$.

By the Borel–Cantelli lemma, we have for σ-a.e. $\mathbf{a}$, there exists $n_0(\varepsilon, \mathbf{a}, \mathbf{v}, b, c) > 0$ such that

$$\max_{1\le j\le m} |f_j(\mathbf{a}) - \mathbb{E}_n\left[f_j(\mathbf{a})\right]| < \varepsilon$$

for $n \ge n_0$.

Therefore, it suffices to show that

$$\lim_{n\to\infty} \mathbb{E}_n\left[f_j(\mathbf{a})\right] = \int x h'(y)\nu^j(x,y),\ \ j=1,\dots,m \tag{7.14}$$

where recall that $\nu^{(j)}$ is the law of the joint distribution of $(g_j, \sum_{l=1}^m v_l g_l + b g_0)$ with $(g_0, g_1, \dots, g_m)$ being independent standard Gaussian random variables.

Let $R, R' > 1$ be two parameters to be chosen later. Define

$$\varphi_R(t) = \max\left(\min\left(h'(t), C_0(1+R)\right), -C_0(1+R)\right),$$

and $\psi_R(t) = h'(t) - \varphi_R(t)$. We observe that

$$|\psi_R(t)| \le C|t|\mathbf{1}_{|t|>R} \le \frac{Ct^2}{R}.$$

Then

$$\begin{aligned}
&\left|\mathbb{E}_n\left[\int x\partial_1\Lambda(y,c)\,d\nu_n^{(j)}(x,y)\right] - \int x\partial_1\Lambda(y,c)\,d\nu^{(j)}(x,y)\right| \\
&\qquad\le \left|\mathbb{E}_n\left[\int x\varphi_R(y)\,d\nu_n^{(j)}(x,y)\right] - \int x\varphi_R(y)\,d\nu^{(j)}(x,y)\right| \\
&\qquad+ \left|\mathbb{E}_n\left[\int x\psi_R(y)\,d\nu_n^{(j)}(x,y)\right]\right| \\
&\qquad+ \int |x|\psi_R(y)\,d\nu^{(j)}(x,y) \\
&\qquad= \mathrm{I} + \mathrm{II} + \mathrm{III},
\end{aligned}$$

where I, II, and III, are the three terms on the left of the equality. Since $\nu^{(j)}$ is the law of joint normal distribution, all moments exist. Therefore, we have the upper bound

$$\mathrm{III} \le \frac{C}{R}\int |x||y|^2 d\nu^{(j)}(x,y) = \frac{C}{R}.$$

Equation (7.5) implies that the second term is bounded by

$$\mathrm{II} \le \frac{C\|\mathbf{u}\|_2^2}{R}.$$

Denote by $\tau_{R'}(x) = \max\left(\min\left(x, R'\right), -R'\right)$. For the first term, we have the bound

$$\begin{aligned}
I &\le \left|\mathbb{E}_n\left[\int \tau_{R'}(x)\varphi_R(y)\,d\nu_n^{(j)}(x,y)\right] - \int \tau_{R'}(x)\varphi_R(y)\,d\nu^{(j)}(x,y)\right| \\
&+ \mathbb{E}_n\left[\int_{|x|>R'} |x\varphi_R(y)|\,d\nu_n^{(j)}(x,y)\right] \\
&+ \int_{|x|>R'} |x\varphi_R(y)|\,d\nu^{(j)}(x,y) \\
&= \mathrm{I}' + \mathrm{II}' + \mathrm{III}'.
\end{aligned}$$

By (7.7),

$$\mathrm{II}' \leq \frac{C(1+R)}{R'}.$$

Since all moments of ν^j exist, the last term can be bounded by

$$\mathrm{III}' \leq \frac{C(1+R)}{R'} \int x^2 d\nu^{(j)}(x,y) = \frac{C(1+R)}{R'}.$$

Choosing R, R' large enough, depending only on ε, we conclude that $\mathrm{II}+\mathrm{III}+\mathrm{II}'+\mathrm{III}' < \varepsilon$. Since jointly normal random vector is determined by its moments [21, Theorem 2], by Lemma 7.3 and Lemma 7.5, we have

$$\mathbb{E}_n\left[\nu_n^{(j)}\right] \to \nu^j \tag{7.15}$$

weakly for all $j = 1, \ldots, m$. Therefore, $\mathrm{I}' + \mathrm{II} + \mathrm{III} + \mathrm{II}' + \mathrm{III}' < 2\varepsilon$ for large enough n for all $j = 1, \ldots, m$. This implies (7.14) and the proof is complete. $\square$

Lemma 7.7. *Retain the notations in Lemma 7.6. Then for σ-a.e. $\mathbf{a}$,*

$$\lim_{n\to\infty} \max_{m+1\leq j\leq m+m_n} \left| m_n^{1/2} f_j^{(n)}\left(\mathbf{a}^{(n)}\right) - \mathbb{E}\left[g_0 h'\left(\sum_{l=1}^m v_l g_l + b g_0\right)\right]\right| = 0.$$

Proof. In this proof, we use C to denote a constant that may depend on C_0, whose value might change from line to line.

Fix $\varepsilon > 0$. Proceeding similarly as in Lemma 7.6, we know there exists $n_0(\varepsilon, \mathbf{a}, \mathbf{v}, b, c)$ such that

$$\max_{m+1\leq j\leq m+m_n} |f_j(\mathbf{a}) - \mathbb{E}_n[f_j(\mathbf{a})]| < m_n^{-1/2}\varepsilon$$

for $n \geq n_0$.

Now it suffices to show that

$$\lim_{n\to\infty} \max_{m+1\leq j\leq m+m_n} \left|\mathbb{E}_n\left[m_n^{1/2} f_j(\mathbf{a})\right] - \int x h'(y)\, d\nu(x,y)\right| = 0, \tag{7.16}$$

where we recall that ν is the law of the joint distribution of $(g_0, \sum_{l=1}^m v_l g_l + b g_0)$, with $(g_0, g_1, \ldots, g_m)$ being independent Gaussian random variables.

Let $R, R' > 0$ be two parameters to be chosen later. Define

$$\varphi_R(t) = \max\left(\min\left(h'(t), C_0(1+R)\right), -C_0(1+R)\right)$$

and $\psi_R(t) = \partial_1 \Lambda(t) - \varphi_R(t)$. We observe that

$$|\psi_R(t)| \leq C|t|\mathbf{1}_{|t|>R} \leq \frac{Ct^2}{R}.$$

Then for $j \in \{m+1, \dots, m+m_n\}$,

$$\begin{aligned}\left|\mathbb{E}_n\left[m_n^{1/2}\int x\partial_1\Lambda(y,c)\,d\nu_n^{(j)}(x,y)\right]-\int x\partial_1\Lambda(y,c)\,d\nu(x,y)\right| &\\ \leq \left|\mathbb{E}_n\left[m_n^{1/2}\int x\varphi_R(y)\,d\nu_n^{(j)}(x,y)\right]-\int x\varphi_R(y)\,d\nu(x,y)\right| &\\ +\left|\mathbb{E}_n\left[m_n^{1/2}\int x\psi_R(y)\,d\nu_n^{(j)}(x,y)\right]\right| &\\ +\int |x|\psi_R(y)\,d\nu(x,y) &\\ =\mathrm{I}+\mathrm{II}+\mathrm{III}. &\end{aligned}$$

where I, II, and III, are the three terms on the left of the equality. Since a jointly normal random variable has all moments finite, the last term can be upper bounded by

$$\mathrm{III} \leq \frac{C}{R}\int |x||y|^2\,d\nu(x,y) = \frac{C}{R}.$$

By (7.6),

$$\mathrm{II} \leq \frac{C\|\mathbf{u}\|_2^2}{R}.$$

Observe that measures $\mathbb{E}_n[\nu_n^{(j)}]$ for $j \in \{m+1, \dots, m+m_n\}$ are equal in distribution. For any such j, we have

$$\begin{aligned}\mathbb{E}_n\left[m_n^{1/2}\int x\varphi_R(y)\,d\nu_n^{(j)}(x,y)\right] &= \mathbb{E}_n\left[m_n^{1/2}\frac{1}{m_n}\sum_{l=m+1}^{m+m_n}\int x\varphi_R(y)\,d\nu_n^l(x,y)\right]\\ &= \mathbb{E}_n\left[\int x\varphi_R(y)\,d\nu_n(x,y)\right],\end{aligned}$$

which is independent of j. Using this and setting $\tau_{R'}(x) = \max\left(\min\left(x, R'\right), -R'\right)$, we have

$$\begin{aligned}I &= \left|\mathbb{E}_n\left[\int x\varphi_R(y)\,d\nu_n(x,y)\right]-\int x\varphi_R(y)\,d\nu(x,y)\right|\\ &\leq \left|\mathbb{E}_n\left[\int \tau_{R'}(x)\varphi_R(y)\,d\nu_n(x,y)\right]-\int \tau_{R'}(x)\varphi_R(y)\,d\nu(x,y)\right|\\ &+\mathbb{E}_n\left[\int_{|x|>R'}|x\varphi_R(y)|\,d\nu_n(x,y)\right]\\ &+\int_{|x|>R'}|x\varphi_R(y)|\,d\nu(x,y)\\ &= \mathrm{I}'+\mathrm{II}'+\mathrm{III}'.\end{aligned}$$

By (7.8),

$$\mathrm{II}' \leq \frac{C(1+R)}{R'}.$$

Since a jointly normal random variable has all moments finite, the last term is bounded by

$$\mathrm{III}' \leq \frac{C(1+R)}{R'} \int x^2 d\nu(x,y) = \frac{C(1+R)}{R'}.$$

By choosing R, R' large enough, depending on ε, we conclude that $\mathrm{II}+\mathrm{III}+\mathrm{II}'+\mathrm{III}' < \varepsilon$. Since jointly normal random vector is determined by its moments [21, Theorem 2], by Lemma 7.4 and Lemma 7.5, we have

$$\mathbb{E}_n\left[\nu_n\right] \to \nu$$

Therefore, $\mathrm{I}'+\mathrm{II}+\mathrm{III}+\mathrm{II}'+\mathrm{III}' < 2\varepsilon$ for large enough n. This implies (7.16) and the proof is complete. $\square$

Lemma 7.8. *Retain the notations in Lemma 7.6. Let $\mathbf{u}^{(n)}$ be defined as in (7.2). Then for σ-a.e. $\mathbf{a}$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i,j=1}^{n} h'\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\rangle\right) h'\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_j\rangle\right) \left(\sum_{l=m+m_n+1}^{k_n} a_{il}a_{jl}\right) = 0. \tag{7.17}$$

Proof. Rewrite (7.17) as

$$\sum_{l=m+m_n+1}^{k_n} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} h'\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\rangle\right) a_{il}\right)^2 = \sum_{l=m+m_n+1}^{k_n} \left(f_l\left(\mathbf{a}\right)\right)^2.$$

Define

$$\begin{aligned}
\varphi_n\left(x\right) &= \max\left(\min\left(h'\left(x\right), C_0(1+n^{1/8}\|\mathbf{u}\|_2)\right), -C(c)\left(1+n^{1/8}\|\mathbf{u}\|_2\right)\right),\\
\tau_n\left(x\right) &= \max\left(\min\left(x, n^{1/4}\right), -n^{1/4}\right).
\end{aligned}$$

Fix $\varepsilon \in (0,1)$. For any $l = m+m_n+1, \ldots, k_n$, note that

$$\begin{aligned}
&\sigma_n\left(|f_l\left(\mathbf{a}\right)| \geq \varepsilon k_n^{-1/2}\right)\\
&\leq \sigma_n\left(\left|\frac{1}{n}\sum_{i=1}^{n} \varphi_n\left(\left\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\right\rangle\right) \tau_n\left(\sqrt{n}a_{il}\right)\right| \geq \varepsilon k_n^{-1/2}\right)\\
&\quad + n\sigma_n\left(\left|\left\langle \mathbf{u}, \sqrt{n}\mathbf{a}_1\right\rangle\right| \geq n^{1/(4q)}\|\mathbf{u}\|_2\right) + n\sigma_n\left(|\sqrt{n}a_{1l}| \geq n^{1/4}\right)\\
&=: \mathrm{I}+\mathrm{II}+\mathrm{III}.
\end{aligned} \tag{7.18}$$

By rotational invariance of Haar measure and the fact that $\sqrt{n}a_{11}$ is sub-Gaussian with $\|\sqrt{n}a_{11}\|_{\psi_2}$ bounded by a constant C independent of n [27, Theorem 3.4.6], we have

$$\begin{aligned}&\mathrm{II} = n\sigma_n\left(|\sqrt{n}a_{11}| \geq n^{1/8}\right) \leq 2n\exp\left(-Cn^{1/4}\right),\\ &\mathrm{III} \leq 2n\exp\left(-Cn^{1/2}\right).\end{aligned} \tag{7.19}$$

For I, first note that $l > m + m_n$ and that $u_j = 0$ for all $j > m + m_n$. Conditioning on the first $m + m_n$ columns of $\mathbf{a}$, and by the invariance of Haar measure under change of sign of single column, we have

$$\mathbb{E}_n\left[\frac{1}{n}\sum_{i=1}^{n}\varphi_n\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\rangle\right)\tau_n\left(\sqrt{n}a_{il}\right) \mid a_{ij},\ i = 1,\ldots,n,\ j = 1,\ldots,m+m_n\right] = 0.$$

Taking expectation over the first $m + m_n$ columns in the above equation gives

$$\mathbb{E}_n\left[\frac{1}{n}\sum_{i=1}^{n}\varphi_n\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\rangle\right)\tau_n\left(\sqrt{n}a_{il}\right)\right] = 0. \tag{7.20}$$

Let

$$\mathfrak{h}(\mathbf{a}) = \frac{1}{n}\sum_{i=1}^{n}\varphi_n\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\rangle\right)\tau_n\left(\sqrt{n}a_{il}\right).$$

For $j > m + m_n$,

$$|\mathfrak{h}(\mathbf{a} + t\mathbf{e}_{ij}) - \mathfrak{h}(\mathbf{a})| \leq \frac{1}{\sqrt{n}}\varphi_n\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\rangle\right)\delta_{jl}|t| \leq C\left(1 + \|\mathbf{u}\|_2\right)n^{-1/4}|t|.$$

For $j \leq m + m_n$,

$$|\mathfrak{h}(\mathbf{a} + t\mathbf{e}_{ij}) - \mathfrak{h}(\mathbf{a})| \leq \frac{1}{\sqrt{n}}|u_j|\tau_n\left(\sqrt{n}a_{il}\right)|t| \leq \|\mathbf{u}\|_\infty n^{-1/4}|t|.$$

The previous two equations imply that the Lipshitz constant of h is upper bounded by $C(1 + \|\mathbf{u}\|_2 + \|\mathbf{u}\|_\infty)n^{-1/4}$. Then Theorem 6.3 and (7.20) together imply that

$$I \leq 2\exp\left(-\frac{C\varepsilon^2 n}{k_n\|h\|_{Lip}^2}\right) \leq 2\exp\left(-\frac{C\varepsilon^2 n^{1/2}}{\left(1 + \|\mathbf{u}\|_2 + \|\mathbf{u}\|_\infty\right)^2}\right). \tag{7.21}$$

Combining (7.18), (7.19) and (7.21), we have

$$\sigma_n\left(|f_l\left(\mathbf{a}\right)| \geq \varepsilon k_n^{-1/2}\right) \leq 2n\exp\left(-\frac{Cn^{1/4}}{(1 + \|\mathbf{u}\|_2 + \|\mathbf{u}\|_\infty)^2}\right).$$

A union bound over all $l > m + m_n$ gives

$$\sigma_n\left(\text{there exists } l > m + m_n \text{ such that } |f_l\left(\mathbf{a}\right)| \geq \varepsilon k_n^{-1/2}\right) \leq 2nk_n\exp\left(-\frac{Cn^{1/4}}{(1 + \|\mathbf{u}\|_2 + \|\mathbf{u}\|_\infty)^2}\right).$$

Combined with the Borel–Cantelli lemma, this completes the proof of (7.17). $\square$

Proof of Lemma 4.2. Recall the definition of $F(\mathbf{u},\mathbf{a},c)$ in (3.4). For $\mathbf{u}^{(n)} \in \mathbb{R}^{k_n}$ as in (7.2), with the given $\mathbf{v}, b, c, m$, define the tilted measure $\tilde{\mathbb{P}}^{(n)}$ by

$$\frac{d\tilde{\mathbb{P}}^{(n)}}{d\mathbb{P}^{(n)}}(\mathbf{x}) = \exp\left(n\left\langle \mathbf{u}^{(n)}, \frac{1}{\sqrt{n}}\left(\mathbf{a}^{(n)}\right)^T \mathbf{x}\right\rangle + c\sum_{i=1}^n \eta(x_i) - nF(\mathbf{u}^{(n)}, \mathbf{a}^{(n)}, c)\right), \ \forall \mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n,$$

where $\mathbb{P}^{(n)}$ denotes the probability with respect to $X_1, \ldots, X_n$. Let $\tilde{\mathbb{E}}^{(n)}$ the corresponding expectation. Note that $\tilde{\mathbb{P}}^{(n)}$ is still a product measure on $\mathbb{R}^n$.

Let $\mathbf{x}^{(n)} = \left(\mathbf{w}, tm_n^{-1/2}\mathbf{1}_{m_n}, \mathbf{0}_{k_n-m_n-m}, s\right)$. Recalling $B = B((\mathbf{w},t,s), r)$ from the statement of Lemma 4.2 and the definition of W, L as in (3.25), we have

$$\begin{aligned}
&\mathbb{P}^{(n)}\left((W_{\leftarrow m}, \|W_{\to m}\|_2, L) \in B, \|W_{\to m}\|_\infty \le 2m^{-1/2}\right)\\
\ge&\mathbb{P}^{(n)}\left((W_{\leftarrow m}, W_{\to m}, L) \in B\left(\mathbf{x}^{(n)}, r\right), \|W_{\to m}\|_\infty \le 2m^{-1/2}\right)\\
\ge&e^{-n(\langle \mathbf{v},\mathbf{w}\rangle + bt + cs - F(\mathbf{u},\mathbf{a},c) + r|(\mathbf{v},b,c)|)}\\
&\times \tilde{\mathbb{P}}^{(n)}\left((W_{\leftarrow m}, W_{\to m}, L) \in B\left(\mathbf{x}^{(n)}, r\right), \|W_{\to m}\|_\infty \le 2m^{-1/2}\right).
\end{aligned}$$

We show below that the following four claims hold σ-a.s. in $\mathbf{a}$ for any $\varepsilon > 0$.

Claim 1.

$$\lim_{n\to\infty} \tilde{\mathbb{P}}^{(n)}\left(|W_j - w_j| \ge \varepsilon\right) = 0, \ j = 1, \ldots, m. \tag{7.22}$$

Claim 2.

$$\lim_{n\to\infty} \tilde{\mathbb{P}}^{(n)}\left(\max_{m+1\le j\le m+m_n}\left|W_j - tm_n^{-1/2}\right| > \varepsilon m_n^{-1/2}\right) = 0. \tag{7.23}$$

Claim 3.

$$\lim_{n\to\infty} \tilde{\mathbb{P}}^{(n)}\left(\sum_{l=m+m_n+1}^{k_n} W_l^2 \ge \varepsilon\right) = 0. \tag{7.24}$$

Claim 4.

$$\lim_{n\to\infty} \tilde{\mathbb{P}}^{(n)}\left(|L - s| \ge \varepsilon\right) = 0. \tag{7.25}$$

Combining (7.22), (7.23), (7.24) and (7.25), we have for σ-a.e. $\mathbf{a}$ that

$$\liminf_{n\to\infty} \tilde{\mathbb{P}}^{(n)}\left((W_{\leftarrow m}, W_{\to m}, L) \in B\left(\mathbf{x}^{(n)}, r\right), \|W_{\to m}\|_\infty \le 2m^{-1/2}\right) = 1.$$

Moreover, $F(\mathbf{u}^{(n)}, \mathbf{a}^{(n)}, c) \to \Lambda_m(\mathbf{v}, b, c)$ for σ-a.e. $\mathbf{a}$ by Lemma 3.6, with $d_n = 1$. We have for σ-a.s. $\mathbf{a}$,

$$\begin{aligned}
&\liminf_{n\to\infty} n^{-1}\mathbb{P}^{(n)}\left((W_{\leftarrow m}, \|W_{\to m}\|_2, L) \in B, \|W_{\to m}\|_\infty \le 2m^{-1/2}\right)\\
&\qquad \ge -\left(\langle \mathbf{v}, \mathbf{w}\rangle + bt + cs - \Lambda_m(\mathbf{v}, b, c) + r|(\mathbf{v}, b, c)|\right) = -\Lambda_m^*(\mathbf{w}, t, s) - r|(\mathbf{v}, b, c)|,
\end{aligned}$$

where the last equality follows from $(\mathbf{w}, t, s) = \nabla\Lambda_m(\mathbf{v}, b, c)$.

To complete the proof, we take $r \to 0$. $\square$

Proof of (7.22). Under $\tilde{\mathbb{P}}^{(n)}$, we have

$$\begin{aligned}
d\tilde{\mathbb{P}}^{(n)}(x_1,\dots,x_n) &= \exp\left(n\langle \mathbf{u}, \frac{1}{\sqrt{n}}\mathbf{a}^T\mathbf{x}\rangle + ncL - nF(\mathbf{u},\mathbf{a},c)\right) d\mathbb{P}^{(n)}(x_1,\dots,x_n) \\
&= \exp\left(-nF(\mathbf{u},\mathbf{a},c)\right)\prod_{i=1}^{n}\exp\left(x_i\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\rangle + c\eta(x_i)\right) d\mathbb{P}^{(n)}\left(x_1,\dots,x_n\right).
\end{aligned}$$

Under the assumption in Proposition 2.4, by Lemma 7.1, we have

$$\tilde{\mathbb{P}}^{(n)}\left(|X_i - \tilde{\mathbb{E}}^{(n)}\left[X_i\right]| > t\right) \le 2\exp(-t^2/C^2),$$

for some constant $C > 0$ depending only on c. Now the Hoeffding's inequality for sub-Gaussian random variables (see [27, Theorem 2.6.2]) implies that

$$\begin{aligned}
\tilde{\mathbb{P}}^{(n)}\left(\max_{1\le j\le m}|W_j - \tilde{\mathbb{E}}^{(n)}[W_j]| \ge \varepsilon/3\right) &\le m\max_{1\le j\le m}\tilde{\mathbb{P}}^{(n)}\left(\left|\sum_{i=1}^{n}\frac{a_{ij}}{\sqrt{n}}(X_i - \tilde{\mathbb{E}}^{(n)}[X_i])\right| \ge \varepsilon/3\right) \\
&\le 2m\exp(-Cn\varepsilon^2) \to 0,
\end{aligned} \tag{7.26}$$

for all $\mathbf{a} \in \mathbb{V}$.

Note that $\tilde{\mathbb{E}}^{(n)}X_i = \partial_1\Lambda\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\rangle, c\right)$, we have

$$\tilde{\mathbb{E}}^{(n)}[W_j] = \frac{1}{n}\sum_{i=1}^{n}\sqrt{n}a_{ij}\partial_1\Lambda\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\rangle, c\right).$$

Now Lemma 7.6 implies that for σ-a.e. $\mathbf{a}$, for each $j = 1,\dots,m$,

$$\lim_{n\to\infty}\tilde{\mathbb{E}}^{(n)}[W_j] = \mathbb{E}_g\left[g_j\partial_1\Lambda\left(\sum_{l=1}^{m}v_lg_l + bg_0, c\right)\right] = \partial_j\Lambda_m(\mathbf{v},b,c) = w_j. \tag{7.27}$$

Combining (7.26) and (7.27), (7.22) follows. □

Proof of (7.23). As in (7.26), Hoeffding's inequality for sub-Gaussian variables implies that

$$\begin{aligned}
&\tilde{\mathbb{P}}^{(n)}\left(\max_{m+1\le j\le m+m_n}|W_j - \tilde{\mathbb{E}}^{(n)}[W_j]| \ge m_n^{-1/2}\varepsilon/3\right) \\
&\qquad \le m_n\max_{m+1\le j\le m+m_n}\tilde{\mathbb{P}}^{(n)}\left(\left|\sum_{i=1}^{n}\frac{a_{ij}}{\sqrt{n}}(X_i - \tilde{\mathbb{E}}^{(n)}[X_i])\right| \ge m_n^{-1/2}\varepsilon/3\right) \\
&\qquad \le 2m_n\exp(-cn\varepsilon^2/m_n) \to 0,
\end{aligned} \tag{7.28}$$

for every $\mathbf{a} \in \mathbb{V}$.

Note that

$$\tilde{\mathbb{E}}^{(n)}[W_j] = \frac{1}{n}\sum_{i=1}^{n}\sqrt{n}a_{ij}\partial_1\Lambda\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\rangle, c\right).$$

The conclusion now follows from (7.28) and Lemma 7.7. □

Proof of (7.24). By Markov's inequality,

$$\begin{aligned}
\tilde{\mathbb{P}}^{(n)}\left(\sum_{l=m+m_n+1}^{k_n} W_l^2 \geq \varepsilon\right) \leq & \varepsilon^{-1}\mathbb{E}^{(n)}\left[\sum_{l=m+m_n+1}^{k_n} W_l^2\right] \\
= & \varepsilon^{-1}n^{-1}\sum_{i,j=1}^{n}\tilde{\mathbb{E}}^{(n)}\left[X_iX_j\right]\sum_{l=m+m_n+1}^{k_n} a_{il}a_{jl} \\
= & \varepsilon^{-1}n^{-1}\left(\sum_{i,j=1}^{n}\tilde{\mathbb{E}}^{(n)}[X_i]\tilde{\mathbb{E}}^{(n)}[X_j]\sum_{l=m+m_n+1}^{k_n} a_{il}a_{jl}+\right. \\
& \left.\sum_{i=1}^{n}\mathrm{Var}_{\tilde{\mathbb{P}}^{(n)}}\left(X_i\right)\sum_{l=m+m_n+1}^{k_n} a_{il}^2\right).
\end{aligned}$$

Note that $\mathrm{Var}_{\tilde{\mathbb{P}}^{(n)}}(X_i) = \partial_1^2\Lambda\left(\left\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\right\rangle, c\right)$.

By Assumption 4 on X_i, $\mathrm{Var}_{\tilde{\mathbb{P}}^{(n)}}(X_i)$ is uniformly bounded for all n and i. Therefore,

$$\tilde{\mathbb{P}}^{(n)}\left(\sum_{l=m+m_n+1}^{k_n} W_l^2 \geq \varepsilon\right) \leq \varepsilon^{-1}n^{-1}\sum_{i,j=1}^{n}\tilde{\mathbb{E}}^{(n)}[X_i]\tilde{\mathbb{E}}^{(n)}(X_j)\sum_{l=m+m_n+1}^{k_n} a_{il}a_{jl} + \varepsilon^{-1}O\left(\frac{k_n}{n}\right).$$

Now it suffices to show

$$\lim_{n\to\infty} n^{-1}\sum_{i,j=1}^{n}\tilde{\mathbb{E}}^{(n)}[X_i]\tilde{\mathbb{E}}^{(n)}[X_j]\sum_{l=m+m_n+1}^{k_n} a_{il}a_{jl} = 0$$

for σ-a.e. $\mathbf{a}$.

Note that $\tilde{\mathbb{E}}^{(n)}[X_i] = \partial_1\Lambda\left(\langle \mathbf{u}, \sqrt{n}\mathbf{a}_i\rangle, c\right)$, the above convergence follows directly from Lemma 7.8. □

Proof of (7.25). Pick small enough $\lambda > 0$. By Chebyshev's inequality and similar calculations as in the proof of Lemma 7.1, we have

$$\begin{aligned}
&\tilde{\mathbb{P}}^{(n)}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\eta\left(X_i\right)-\tilde{\mathbb{E}}^{(n)}\left[\eta\left(X_i\right)\right]\right)>\frac{\varepsilon}{2}\right)\\
\le& e^{-\lambda n\varepsilon/2}\prod_{i=1}^{n}\tilde{\mathbb{E}}^{(n)}\left[\exp\left(\lambda\eta\left(X_i\right)-\lambda\tilde{\mathbb{E}}^{(n)}\left[\eta\left(X_i\right)\right]\right)\right]\\
=& e^{-\lambda n\varepsilon/2}\prod_{i=1}^{n}\exp\left(\Lambda\left(\langle\mathbf{u},\sqrt{n}\mathbf{a}_i\rangle,c+\lambda\right)-\Lambda\left(\langle\mathbf{u},\sqrt{n}\mathbf{a}_i\rangle,c\right)-\lambda\partial_2\Lambda\left(\langle\mathbf{u},\sqrt{n}\mathbf{a}_i\rangle,c\right)\right)\\
=&\exp\left(\frac{\lambda^2}{2}\sum_{i=1}^{n}\partial_2^2\Lambda\left(\langle\mathbf{u},\sqrt{n}\mathbf{a}_i\rangle,\xi_i\right)-\lambda n\frac{\varepsilon}{2}\right),
\end{aligned}$$

where $\xi_i \in (c, c+\lambda)$.

By Assumption 4, we have

$$\tilde{\mathbb{P}}^{(n)}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\eta\left(X_i\right)-\tilde{\mathbb{E}}^{(n)}\left[\eta\left(X_i\right)\right]\right)>\frac{\varepsilon}{2}\right)\le\exp\left(\left(-\frac{\lambda}{2}\varepsilon+C(1+\|\mathbf{u}\|_2^2)\lambda^2\right)n\right)\to 0$$

for small enough $\lambda > 0$ and all $\mathbf{a} \in \mathbb{V}$.

Now it suffices to show that

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbb{E}}^{(n)}\left[\eta\left(X_i\right)\right]\to\partial_3\Lambda_m(\mathbf{v},b,c)=s,$$

for a.s. $\mathbf{a} \in \otimes_n \mathbb{V}_{n,k_n}$.

Fix $\varepsilon > 0$. Note that $\tilde{\mathbb{E}}^{(n)}\eta\left(X_i\right) = \partial_2\Lambda(\langle\mathbf{u}, \sqrt{n}\mathbf{a}_i\rangle, c)$. Consider the random measures

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{\langle\mathbf{u},\sqrt{n}\mathbf{a}_i\rangle}.$$

We can rewrite

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbb{E}}^{(n)}\left[\eta\left(X_i\right)\right]=\int\partial_2\Lambda\left(x,c\right)d\mu_n(x)=:f(\mathbf{a})$$

Upon differentiating and recalling Assumption 4, it is clear that the Lipshitz constant of f is upper bounded by $C\|\mathbf{u}\|_\infty(1 + \|\mathbf{u}\|_2)$.

Using Theorem 6.3 gives

$$\sigma_n\left(\left|f(\mathbf{a})-\mathbb{E}_n\left[f(\mathbf{a})\right]\right|\ge\varepsilon\right)\le 2\exp\left(-\frac{C\varepsilon^2 n}{\|\mathbf{u}\|_\infty^2\left(1+\|\mathbf{u}\|_2\right)^2}\right)$$

for some constant C depending only on c. By Borel–Cantelli lemma, the above inequality implies for σ-a.e. $\mathbf{a}$, we have $|f(\mathbf{a}) - \mathbb{E}_n f(\mathbf{a})| < \varepsilon$ for large enough $n \ge n_0(\varepsilon, \mathbf{a}, \mathbf{u})$.

It remains to show

$$\mathbb{E}_n\left[f(\mathbf{a})\right] \to \int \partial_2\Lambda(x,c)\mu(x), \tag{7.29}$$

where μ is the law of $\|\mathbf{u}\|_2 g$ with g a standard Gaussian.

Let $R > 0$ be a parameter to be chosen later. Define

$$\varphi_R(t) = \max(\min(\partial_2\Lambda(t,c), C(1+R^2)), -C(1+R^2)),$$

where C, depending on c, is the constant appearing in Assumption 4. Let $\psi_R = \partial_2\Lambda(t) - \varphi_R(t)$. Observe that

$$|\psi_R(t)| \leq Ct^2\mathbf{1}_{|t|>R} \leq \frac{Ct^4}{R^2}. \tag{7.30}$$

Then

$$\begin{aligned}\left|\mathbb{E}_n\left[\int \partial_2\Lambda(x,c)d\mu_n\right] - \int \partial_2\Lambda(x,c)d\mu\right| \leq& \left|\mathbb{E}_n\left[\int \varphi_R d\mu_n\right] - \int \varphi_R d\mu\right| \\ &+ \left|\mathbb{E}_n\left[\int \psi_R d\mu_n\right]\right| + \left|\int \psi_R d\mu\right| \\ =& \mathrm{I} + \mathrm{II} + \mathrm{III}.\end{aligned}$$

The second term is bounded by

$$\begin{aligned}\mathrm{II} &\leq \mathbb{E}_n \int \frac{Cx^4}{R^2} d\mu_n \\ &= \frac{C}{R^2}\mathbb{E}_n\left[\left\langle \mathbf{u}, \sqrt{n}\mathbf{a}_1\right\rangle^4\right] \\ &= \frac{C\|\mathbf{u}\|_2^4}{R^2}\mathbb{E}_n\left[\left(\sqrt{n}a_{11}\right)^4\right] \\ &= \frac{C\|\mathbf{u}\|_2^4}{R^2},\end{aligned} \tag{7.31}$$

where the first inequality follows from (7.30), the second equality follows from the orthogonal invariance of Haar measure and the last equality follows from the fact that $\sqrt{n}a_{11}$ is sub-gaussian.

The third term is bounded by

$$\mathrm{III} \leq \frac{C}{R^2}\int x^4 d\mu = \frac{C\|\mathbf{u}\|_2^4}{R^2}. \tag{7.32}$$

Noticing that $\|\mathbf{u}^{(n)}\|_2 = \|(\mathbf{v}, b)\|_2$ is a constant and jointly Gaussian random vector is determined by their moments [21, Theorem 2]. By Lemma 6.2 and a standard moment convergence theorem, we have the weak convergence

$$\mathbb{E}_n\left[\mu_n\right] \to \mu. \tag{7.33}$$

Combining (7.31), (7.32) and (7.33), $\mathrm{I} + \mathrm{II} + \mathrm{III} < \varepsilon$ for large enough n. This proves (7.29) and completes the proof. □

7.2. **Topological Lemmas.** We begin with a preliminary lemma for the proof of Lemma 4.1.

Lemma 7.9. *Fix an open set $O \subset \mathcal{X} \times \mathbb{R}$ and a point $(\mathbf{w}, r, s) \in O$. Then there exists $m_0 \in \mathbb{N}$ and $\rho' > 0$ such that the following holds for $m \geq m_0$. If $(\mathbf{w}', r', s') \in \mathcal{X} \times \mathbb{R}$ and*

$$\left(\mathbf{w}'_{\leftarrow m_0}, \sqrt{r'^2 - \|\mathbf{w}'_{\leftarrow m_0}\|_2^2}, s'\right) \in B\left(\left(\mathbf{w}_{\leftarrow m_0}, \sqrt{r^2 - \|\mathbf{w}\|_2^2}, s\right), \rho'\right),$$

then $(\mathbf{w}', r', s') \in O$.

Proof of Lemma 7.9. Since the topology on $\mathcal{X}$ is the restriction of the product topology on $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}_+ \times \mathbb{R}_+$, there exists $N_0 \in \mathbb{N}$ and $\delta_1, \delta_2, \delta_3 > 0$ such that for any $N > N_0$, the following claim holds:

$$\begin{aligned} &\text{If } (\mathbf{w}', r', s') \in \mathcal{X} \text{ and } (\mathbf{w}'_{\leftarrow N}, r', s') \in B(\mathbf{w}_{\leftarrow N}, \delta_1) \times B(r, \delta_2) \times B(s, \delta_3), \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{then } (\mathbf{w}', r', s') \in O. \end{aligned} \tag{7.34}$$

Fix $m_0 > N_0$ such that $|\mathbf{w}_{m_0\rightarrow}| \leq \delta_2 r/2$. Put $t = \sqrt{r^2 - \|\mathbf{w}\|_2^2}$ and $\rho' = \min(\delta_1, \delta_2/4, \delta_3)$.

By the assumptions on $\mathbf{w}'_{\leftarrow N}$ and s',

$$\mathbf{w}'_{\leftarrow N} \in B\left(\mathbf{w}_{\leftarrow N}, \rho'\right) \subset B\left(\mathbf{w}_{\leftarrow N}, \delta_1\right), \tag{7.35}$$

and

$$s' \in B\left(s, \rho'\right) \subset B\left(s, \delta_3\right). \tag{7.36}$$

By the assumption on ρ',

$$\rho' > \left|\sqrt{r'^2 - \|\mathbf{w}'_{\leftarrow N}\|_2^2} - \sqrt{r^2 - \|\mathbf{w}\|_2^2}\right|.$$

Multiplying both sides by $\sqrt{r'^2 - \|\mathbf{w}'_{\leftarrow N}\|_2^2} + \sqrt{r^2 - \|\mathbf{w}\|_2^2}$ we have

$$\rho'\left(\sqrt{r'^2 - \|\mathbf{w}'_{\leftarrow N}\|_2^2} + \sqrt{r^2 - \|\mathbf{w}\|_2^2}\right) > \left|r^2 - r'^2 - \|\mathbf{w}_{\leftarrow N}\|_2^2 + \|\mathbf{w}'_{\leftarrow N}\|_2^2 - \|\mathbf{w}_{\rightarrow N}\|_2^2\right|.$$

Using the triangle inequality and rearranging terms, we have

$$\begin{aligned} |r^2 - r'^2| &< \left|\|\mathbf{w}_{\leftarrow N}\|_2^2 - \left\|\mathbf{w}'_{\leftarrow N}\right\|_2^2\right| + \|\mathbf{w}_{\rightarrow N}\|^2 + \rho'(r + r') \\ &\leq \left\|\mathbf{w}_{\leftarrow N} - \mathbf{w}'_{\leftarrow N}\right\|_2 (\|\mathbf{w}_{\leftarrow N}\|_2 + \|\mathbf{w}'_{\leftarrow N}\|_2) + \delta_2 r/2 + \rho'(r + r') \\ &\leq 2\rho'(r + r') + \delta_2 r/2. \end{aligned}$$

The previous display implies that

$$|r - r'| < 2\rho' + \delta_2/2 \leq \delta_2. \tag{7.37}$$

Combining (7.34), (7.35), (7.36), and (7.37) proves the lemma. ☐

In the proof of the next lemma, we use the following concept from convex analysis.

Definition 7.10 (Relative interior)**.** For every non-empty convex set C, the relative interior of C, denoted $\mathrm{ri}(C)$, is defined as the set

$$\mathrm{ri}(C) = \{x \in C : \text{for all } y \in C, \text{ there exists some } \mu > 1 \text{ such that } \mu x + (1-\mu)y \in C\}.$$

In other words, every line segment contained in C with one endpoint at x has an extension that remains contained in C.

Proof of Lemma 4.1. Let m_0 be given as in Lemma 7.9 with the fixed open set $\mathcal{O}$ and point $(\mathbf{w}, r, s) \in \mathcal{O}$.

We first show that Λ_{m_0} is essentially smooth, as defined in Definition 2.1. The finiteness of Λ_{m_0} on $\mathbb{R}^{m_0} \times \mathbb{R} \times (-\infty, T)$ follows from the finiteness of Λ on $\mathbb{R} \times (-\infty, T)$. For differentiability of $\Lambda_{m_0}(\mathbf{v}', b', c')$ with $(\mathbf{v}', b', c') \in \mathbb{R}^{m_0} \times \mathbb{R} \times (-\infty, T)$, it follows from the differentiability of Λ, the bounds in Assumption 4, and the dominated convergence theorem. To check the third condition in the definition, first note that since $\partial_2\Lambda\,(t_1, t_2) \geq 0$ when $t_2 \in (-\infty, T)$, we have

$$\|\nabla\Lambda_{m_0}(\mathbf{v}', b', c')\|_2 \geq \mathbb{E}_{\mathbf{g}}\left[\partial_2\Lambda\left(\sum_{l=1}^{m_0} v_l' g_l + b' g_0, c'\right)\right].$$

Now Fatou's lemma implies that

$$\begin{aligned}
\liminf_{\mathbf{v}', b', c' \to \mathbf{v}_0, b_0, T} \|\nabla\Lambda_{m_0}(\mathbf{v}', b', c')\|_2 &\geq \liminf_{\mathbf{v}', b', c' \to \mathbf{v}_0, b_0, T} \mathbb{E}_{\mathbf{g}}\left[\partial_2\Lambda\left(\sum_{l=1}^{m_0} v_l' g_l + b' g_0, c'\right)\right] \\
&\geq \mathbb{E}_{\mathbf{g}}\left[\liminf_{\mathbf{v}', b', c' \to \mathbf{v}_0, b_0, T} \partial_2\Lambda\left(\sum_{l=1}^{m_0} v_l' g_l + b' g_0, c'\right)\right] \\
&= \infty,
\end{aligned}$$

where the last equality follows from the essential smoothness of Λ.

Since Λ_{m_0} is essentially smooth, by [24, Corollary 26.4.1], we have

$$\mathrm{ri}(\mathcal{D}_{m_0^*}) \subset \nabla\Lambda_{m_0}(\mathcal{D}_{m_0}) \subset \mathcal{D}_{m_0^*}.$$

Note that $\Lambda_{m_0}(0,0,0) = 0$ and the cumulant generating function $\Lambda(s_1, s_2) = \log \mathbb{E}[e^{s_1 X_1 + s_2 \eta(X_1)}]$ satisfies $\Lambda(s_1, 0) \geq \mathbb{E}[s_1 X_1] = 0$. Then we see that $(\mathbf{v}', b') \mapsto \Lambda_{m_0}(\mathbf{v}', b', 0)$ achieves minimum of 0 at $(\mathbf{0}, 0)$. Thus $\nabla\Lambda_{m_0}(\mathbf{0}, 0, 0) = (\mathbf{0}, 0, s_0)$ for some $s_0 \in \mathbb{R}_+$. By differentiating Λ_{m_0} two times, we have

$$\nabla^2\Lambda_{m_0}(\mathbf{0}, 0, 0) = \begin{pmatrix} \partial_1^2\Lambda(\mathbf{0}, 0) I_{m_0} & \mathbf{0} \\ \mathbf{0}^T & \partial_2^2\Lambda(\mathbf{0}, 0) \end{pmatrix}.$$

It is clear that $\nabla^2\Lambda_{m_0}(\mathbf{0}, 0, 0)$ is positive definite as long as X_1 is not degenerate and η is not a constant function.. Thus, inverse function theorem implies that $\nabla\Lambda_{m_0}$ is locally a diffeomorphism, which in particular implies that there exists an $\varepsilon_0 > 0$ such that $B_{\varepsilon_0}(0, 0, s_0) \in \nabla\Lambda_{m_0}(\mathcal{D}_{m_0}) \subset \mathcal{D}_{m_0^*}$. Pick $(\tilde{\mathbf{w}}, \tilde{t}, \tilde{s}) \in B_{\varepsilon_0}(0, 0, s_0)$ such that $\tilde{\mathbf{w}}$ is strictly ordered in the sense that $|\tilde{\mathbf{w}}_1| > \ldots > |\tilde{\mathbf{w}}_N|$. Since $(\tilde{\mathbf{w}}, \tilde{t}, \tilde{s}) \in \mathrm{ri}(\mathcal{D}_{m_0^*})$, [24, Theorem 6.1] implies that

$$\lambda(\mathbf{w}_{\leftarrow m_0}, t, s) + (1-\lambda)(\tilde{\mathbf{w}}, \tilde{t}, \tilde{s}) \in \mathrm{ri}(\mathcal{D}_{m_0^*}) \subset \nabla\Lambda_{m_0}(\mathbb{R}^{m_0} \times \mathbb{R} \times \mathbb{R})$$

for $\lambda \in [0,1)$, where $t = \sqrt{r^2 - \|\mathbf{w}\|_2^2}$. Note that, by the definitions of $(\mathbf{w}_{\leftarrow m_0}, t, s)$ and $(\tilde{\mathbf{w}}, \tilde{t}, \tilde{s})$, that the points on the line segment $\lambda(\mathbf{w}_{\leftarrow m_0}, t, s) + (1-\lambda)(\tilde{\mathbf{w}}, \tilde{t}, \tilde{s})$ with $0 \le \lambda < 1$ are strictly ordered in the first component. By continuity of $\Lambda^*_{m_0}$ along the line segment in $\mathcal{D}_{m_0^*}$ and Lemma 7.9, we can find $(\bar{\mathbf{w}}, \bar{t}, \bar{s})$ on this line segment which satisfies all conditions. $\square$

Appendix A. Properties of p-Gaussian Variables

Lemma A.1. *Fix $p \ge 2$, and let X be a random variable with probability density function proportional to $\exp(-p^{-1}|x|^p)$. Define*

$$\widehat{\Lambda}(t) = \log \mathbb{E} \exp(tX).$$

Then there exists a constant $C > 0$ such that:
(1) $0 \le \widehat{\Lambda}(t) \le C\left(1 + t^2\right)$.
(2) $|\widehat{\Lambda}'(t)| \le C(1 + |t|)$.
(3) $\widehat{\Lambda}''(t) \le C$.

Proof. Since the distribution of X is symmetric about 0, $\Lambda(t) = \Lambda(-t)$, so we may assume $t \ge 0$. Let

$$M(t) = \mathbb{E}\exp(tX) = c_p \int_{\mathbb{R}} \exp\left(tx - \frac{|x|^p}{p}\right) dx,$$

where $c_p > 0$ is the appropriate normalizing constant. By the change of variables $x = t^{1/(p-1)}y$, we have

$$M(t) = c_p t^{\frac{1}{p-1}} \int_{\mathbb{R}} \exp\left(t^{\frac{p}{p-1}}\left(y - \frac{|y|^p}{p}\right)\right) dy.$$

Denote $S(y) = y - \frac{|y|^p}{p}$. Note that $y_0 = 1$ is the unique maximizer of $S(y)$, $S(y)$ is smooth in a neighborhood of $y = 1$, and $S''(1) = 1 \neq 0$. An application of Laplace's method [28, Section 19.2.5, Theorem 2] shows that, as $t \to \infty$,

$$M(t) = c_p t^{\frac{1}{p-1}} \sqrt{\frac{2\pi}{p-1}} \exp\left(t^{\frac{p}{p-1}} \frac{p-1}{p}\right) t^{-\frac{p}{2(p-1)}} \left(1 + O\left(t^{-\frac{p}{p-1}}\right)\right), \tag{A.1}$$

where the implicit constant in the big O notation depends only on p. Similarly, as $t \to \infty$,

$$M'(t) = c_p t^{\frac{2}{p-1}} \sqrt{\frac{2\pi}{p-1}} \exp\left(t^{\frac{p}{p-1}} \frac{p-1}{p}\right) t^{-\frac{p}{2(p-1)}} \left(1 + O\left(t^{-\frac{p}{p-1}}\right)\right), \tag{A.2}$$

$$M''(t) = c_p t^{\frac{3}{p-1}} \sqrt{\frac{2\pi}{p-1}} \exp\left(t^{\frac{p}{p-1}} \frac{p-1}{p}\right) t^{-\frac{p}{2(p-1)}} \left(1 + O\left(t^{-\frac{p}{p-1}}\right)\right). \tag{A.3}$$

Further,

$$\widehat{\Lambda}(t) = \log\left(M(t)\right) = O\left(t^{\frac{p}{p-1}}\right) \text{ as } t \to \infty,$$

$$|\widehat{\Lambda}'(t)| = \left|\frac{M'(t)}{M(t)}\right| = t^{\frac{1}{p-1}}\left(1 + O\left(t^{-\frac{p}{p-1}}\right)\right) \text{ as } t \to \infty,$$

and

$$\widehat{\Lambda}''(t) = \frac{M''(t)M(t) - (M'(t))^2}{(M(t))^2} = O\left(t^{\frac{2-p}{p-1}}\right) \text{ as } t \to \infty.$$

The conclusions of the lemma follow from the previous displays and $p \geq 2$. □

Lemma A.2. *Fix $p \geq 2$, and let X be a random variable with probability density function proportional to* $\exp(-p^{-1}|x|^p)$*. Define*

$$\widehat{\Lambda}\left(t_1, t_2\right) = \mathbb{E}\left[\exp\left(t_1 X + t_2 |X|^p\right)\right].$$

Then for all $t_2 < p^{-1}$ and integers $\alpha, \beta \geq 0$ with $\alpha + \beta \leq 2$, there exists a constant $C_{t_2,\alpha,\beta} > 0$ such that

$$\partial_1^\alpha \partial_2^\beta \widehat{\Lambda}\left(t_1, t_2\right) \leq C_{t_2,\alpha,\beta}\left(1 + |t_1|^{2-\alpha}\right)$$

for all $t_1 \in \mathbb{R}$, and the map $t_2 \mapsto C_{t_2,\alpha,\beta}$ is continuous.

Proof. Let $M(t) = \mathbb{E}\exp(tX)$. By [12, Lemma 5.7],

$$\widehat{\Lambda}\left(t_1, t_2\right) = -\frac{1}{p}\log\left(1 - pt_2\right) + \log M\left(\frac{t_1}{\left(1 - pt_2\right)^{1/p}}\right)$$

for $t_1 \in \mathbb{R}$ and $t_2 < p^{-1}$. The conclusion follows from this representation and Lemma A.1. □

Appendix B. Weingarten Calculus

This appendix contains some preliminary remarks on the Weingarten calculus, and then a lemma necessary for the Gaussian approximation results above. We recall the following definitions from [5]. Let $\mathcal{M}(2d)$ be the set of pair partitions $\mathfrak{m}$ of $\{1, 2, \ldots, 2d\}$. These have a canonical form $\{(\mathfrak{m}(1), \mathfrak{m}(2)), \ldots, (\mathfrak{m}(2d-1), \mathfrak{m}(2d))\}$ for $\mathfrak{m}(2i-1) \leq \mathfrak{m}(2i)$ and $\mathfrak{m}(1) < \mathfrak{m}(3) < \cdots \mathfrak{m}(2d-1)$.

Given pair partitions $\mathfrak{m}, \mathfrak{n} \in \mathcal{M}(2d)$, the graph $\Gamma(\mathfrak{m}, \mathfrak{n})$ has vertices $\{1, 2, \ldots, 2d\}$, and edges$\{(\mathfrak{m}(2i-1), \mathfrak{m}(2i)), (\mathfrak{n}(2i-1), \mathfrak{n}(2i))\}_{i=1}^d$. The Gram matrix $G_d^{(n)} = [G^{(n)}(\mathfrak{m}, \mathfrak{n})]_{\mathfrak{m},\mathfrak{n} \in \mathcal{M}(2d)}$ is defined through its entries

$$G^{(n)}(\mathfrak{m}, \mathfrak{n}) = d^{\operatorname{loop}(\mathfrak{m},\mathfrak{n})},$$

where $\operatorname{loop}(\mathfrak{m}, \mathfrak{n})$ is defined as the number of connected components of $\Gamma(\mathfrak{m}, \mathfrak{n})$. The matrix $\mathrm{Wg}^{(d)}$ is defined as the pseudo-inverse of $G^{(n)}(\mathfrak{m}, \mathfrak{n})$. We denote its entries by

$$\mathrm{Wg}^{(n)}(\mathfrak{m}, \mathfrak{n})_d = [\mathrm{Wg}^{(n)}(\mathfrak{m}, \mathfrak{n})]_{\mathfrak{m},\mathfrak{n} \in \mathcal{M}(2d)}.$$

We write $\mathrm{Wg}(\mathfrak{m}, \mathfrak{n}) = \mathrm{Wg}^{(n)}(\mathfrak{m}, \mathfrak{n})$ when the choice of n is clear.

We recall the following theorem, which was originally proved in [6].

Theorem B.1 ([5, Theorem 2.1])**.** *Given $i_1, \dots i_{2d}$ and $j_1, \dots, j_{2d}$ in $\{1, 2, \dots, n\}$,*

$$\int_{g \in O(n)} g_{i_1 j_1} \cdots g_{i_{2d} j_{2d}} \, dg$$
$$= \sum_{\mathfrak{m}, \mathfrak{n} \in \mathcal{M}(2d)} \mathrm{Wg}(\mathfrak{m}, \mathfrak{n}) \prod_{k=1}^{d} \delta(i_{\mathfrak{m}(2k-1)}, i_{\mathfrak{m}(2k)}) \delta(j_{\mathfrak{m}(2k-1)}, j_{\mathfrak{m}(2k)}). \quad \text{(B.1)}$$

In other words, we sum over all choices of pair partitions that pair indices with the same value. In particular, any moment $g_{i_1 j_1} \cdots g_{i_{2d} j_{2d}}$ without an even number of entries in each row and column vanishes.

Example B.2. *In the expectation of $g_{11}^2 g_{22}^2$ computed according to* (B.1)*, the only possibility for a nonzero term in the sum is that both partitions are $\mathfrak{m} = \mathfrak{n} = \{(1, 2), (3, 4)\}$.*

Considering $\mathfrak{m}$ and $\mathfrak{n}$ as members of the symmetric group S_{2d} (products of transpositions), we note that the value of $\mathrm{Wg}^{(n)}(\mathfrak{m}, \mathfrak{n})$ depends only on the value of $\sigma = \mathfrak{m}^{-1}\mathfrak{n}$ [5, Theorem 3.1]. Consider the graph $\Gamma(\sigma)$ whose vertex set is $\{1, 2, \dots, 2d\}$ with edges $\{2i - 1, 2i\}$ and $\{\sigma(2i - 1), \sigma(2i)\}$ for $1 \le i \le d$. Then $\Gamma(\sigma)$ has connected components of even lengths $2\rho_1 \ge 2\rho_2 \ge \cdots$, which determine a partition $\rho = (\rho_1, \rho_2, \dots)$ of d (in the number-theoretic sense). The length $\ell(\rho)$ of ρ is the number of elements ρ_i it contains. We let $\mathrm{Wg}^{(n)}(\rho)$ equal the value of any $\mathrm{Wg}^{(n)}(\mathfrak{m}, \mathfrak{n})$ such that $\sigma = \mathfrak{m}^{-1}\mathfrak{n}$ has the corresponding partition ρ of d.

Theorem B.3 ([6, Corollary 2.7])**.** *Fix $d \in \mathbb{N}$ and a partition ρ of d. As $n \to \infty$,*

$$\mathrm{Wg}^{(n)}(\rho) = \left(\prod_{i \ge 1} c_{\rho_i - 1} \right) n^{-2d + \ell(\rho)} (1 + O(n^{-1})). \quad \text{(B.2)}$$

Here $c_k = \frac{(2k)!}{(k+1)!k!}$ denotes the k-th Catalan number. The implicit constant in the asymptotic notation depends on d and ρ, but not n.

Example B.4. *If $\mathfrak{m} = \mathfrak{n} = \{(1, 2), (3, 4)\}$ as in the previous example, then $\sigma = \mathfrak{m}^{-1}\mathfrak{n}$ is the identity permutation. The graph $\Gamma(\sigma)$ has the edges $\{1, 2\}$ and $\{3, 4\}$, so $\ell(\rho) = 2$, and we conclude the leading order term in the asymptotic is n^{-2}, as in the Gaussian case. The coefficient of this term is 1 because $c_0 = 1$.*

We next recall Wick's theorem on the expectations of products of centered jointly normal random variables.

Theorem B.5 ([13, Theorem 1.28])**.** *Let $g_1, ..., g_n$ be independent normal random variables with mean zero and variance one. Then*

$$\mathbb{E}[g_1 \cdot \dots \cdot g_n] = \begin{cases} 0, & \text{if } n \text{ is odd}, \\ \sum_{\mathfrak{m} \in \mathcal{M}(n)} \prod_{i=1}^{n/2} \mathbb{E}\left[g_{\mathfrak{m}(2i-1)} g_{\mathfrak{m}(2i)}\right] & \text{otherwise}. \end{cases} \quad \text{(B.3)}$$

Lemma B.6. *Fix $k, n \in \mathbb{N}$, and suppose $\mathbf{a} \in \mathbb{V}_{n,k}$ is a random matrix distributed according to the Haar measure on $\mathbb{V}_{n,k}$. Then for all $d \in \mathbb{N}$,*

$$\mathbb{E}\left[\prod_{i=1}^{2d} \sqrt{n} a_{1j_i}\right] = \mathbb{E}\left[\prod_{i=1}^{2d} g_{j_i}\right]\left(1 + O\left(n^{-1}\right)\right). \tag{B.4}$$

where the g_i are independent standard Gaussian, and the implicit constant depends only on d.

Proof. Let $\Xi(\sigma)$ denote the coset type of $\sigma \in S_{2d}$ and set

$$\binom{\mathfrak{m}}{i_1 \dots i_{2d}} = \prod_{l=1}^{d} \delta_{i_{\mathfrak{m}(2l-1),\mathfrak{m}(2l)}}.$$

Let $\mathbf{1}_{2d} \in S_{2d}$ be the identity permutation. With these notations, we have

$$\mathbb{E}\left[\prod_{i=1}^{2d} \sqrt{n} a_{1j_i}\right] \tag{B.5}$$

$$= n^d \sum_{\mathfrak{m},\mathfrak{n} \in \mathcal{M}(2d)} \mathrm{Wg}^{(n)}(\mathfrak{m},\mathfrak{n}) \binom{\mathfrak{n}}{j_1 \dots j_{2d}} \tag{B.6}$$

$$= n^d \sum_{\rho \vdash d} \mathrm{Wg}^{(n)}(\rho) \sum_{\Xi(\mathfrak{m}^{-1}\mathfrak{n}) = \rho} \binom{\mathfrak{n}}{j_1 \dots j_{2d}} \tag{B.7}$$

$$= n^d\, \mathrm{Wg}^{(n)}\left((\mathbf{1}_{2d}), n\right) \mathbb{E}\left[\prod_{i=1}^{2d} g_{j_i}\right] + n^d \sum_{\substack{\rho \vdash d \\ \rho \neq \mathbf{1}_{2d}}} \mathrm{Wg}^{(n)}(\rho) \sum_{\Xi(\mathfrak{m}^{-1}\mathfrak{n}) = \rho} \binom{\mathfrak{n}}{j_1 \dots j_{2d}}, \tag{B.8}$$

where the first equality follows from Theorem B.1, the second equality follows from the fact that $\mathrm{Wg}^{(n)}(\mathfrak{m},\mathfrak{n})$ depends only on the coset type of $\mathfrak{m}^{-1}\mathfrak{n}$ (see [5, Theorem 3.1]), and the third equality follows from Wick's Theorem, Theorem B.5.

By Theorem B.3, the first term in (B.8) is

$$\mathbb{E}\left[\prod_{i=1}^{2d} g_{j_i}\right]\left(1 + O\left(n^{-1}\right)\right). \tag{B.9}$$

By the same theorem, the absolute value of the second term of (B.8) is upper bounded by

$$n^d \mathbb{E}\left[\prod_{i=1}^{2d} g_{j_i}\right] \sum_{\substack{\rho \vdash d \\ \rho \neq \mathbf{1}_{2d}}} |\,\mathrm{Wg}^{(n)}(\rho)| \leq C_d \mathbb{E}\left[\prod_{i=1}^{2d} g_{j_i}\right] n^{-1} \tag{B.10}$$

for n sufficiently large, where $C_d = \max_{\rho \vdash d} \prod_{i \geq 1} c_{\rho_i - 1} + 1$.

Combining (B.8), (B.9) and (B.10), we obtain the desired conclusion. □

References

1. David Alonso-Gutiérrez, Joscha Prochno, and Christoph Thäle, *Large deviations for high-dimensional random projections of ℓ_p^n-balls*, Advances in Applied Mathematics **99** (2018), 1–35.
2. Greg Anderson, Alice Guionnet, and Ofer Zeitouni, *An Introduction to Random Matrices*, no. 118, Cambridge University Press, 2010.
3. Milla Anttila, Keith Ball, and Irini Perissinaki, *The central limit problem for convex bodies*, Transactions of the American Mathematical Society **355** (2003), no. 12, 4723–4735.
4. Zhidong Bai and Jack Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, Springer, 2010.
5. Benoît Collins and Sho Matsumoto, *On some properties of orthogonal Weingarten functions*, Journal of Mathematical Physics **50** (2009), no. 11, 113516.
6. Benoît Collins and Piotr Śniady, *Integration with respect to the Haar measure on unitary, orthogonal and symplectic group*, Communications in Mathematical Physics **264** (2006), no. 3, 773–795.
7. Francis Comets and Amir Dembo, *Ordered overlaps in disordered mean-field models*, Probability Theory and Related Fields **121** (2001), no. 1, 1–29.
8. Amir Dembo and Ofer Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed., Springer, 2009.
9. Persi Diaconis and David Freedman, *Asymptotics of graphical projection pursuit*, The Annals of Statistics (1984), 793–815.
10. Jerome Friedman and John Tukey, *A projection pursuit algorithm for exploratory data analysis*, IEEE Transactions on Computers **100** (1974), no. 9, 881–890.
11. Nina Gantert, Steven Soojin Kim, and Kavita Ramanan, *Cramér's theorem is atypical*, Advances in the mathematical sciences, Springer, 2016, pp. 253–270.
12. ———, *Large deviations for random projections of ℓ^p balls*, The Annals of Probability **45** (2017), no. 6B, 4419–4476.
13. Svante Janson, *Gaussian Hilbert Spaces*, no. 129, Cambridge University Press, 1997.
14. William B Johnson, *Extensions of lipschitz mappings into a Hilbert space*, Contemp. Math. **26** (1984), 189–206.
15. Zakhar Kabluchko, Joscha Prochno, and Christoph Thäle, *High-dimensional limit theorems for random vectors in ℓ_p^n-balls*, Communications in Contemporary Mathematics **23** (2021), no. 03, 1950073.
16. Tom Kaufmann, Holger Sambale, and Christoph Thäle, *Large deviations for uniform projections of p-radial distributions on ℓ_p^n-balls*, arXiv preprint arXiv:2203.00476 (2022), 1–11.
17. Steven Soojin Kim, Yin-Ting Liao, and Kavita Ramanan, *An asymptotic thin shell condition and large deviations for random multidimensional projections*, Advances in Applied Mathematics **134** (2022), 102306.
18. Steven Soojin Kim and Kavita Ramanan, *Large deviation principles induced by the Stiefel manifold, and random multi-dimensional projections*, arXiv preprint arXiv:2105.04685 (2021), 1–18.
19. B Klartag, *Power-law estimates for the central limit theorem for convex sets*, Journal of Functional Analysis **245** (2007), no. 1, 284–310.
20. Bo'az Klartag, *A central limit theorem for convex sets*, Inventiones mathematicae **168** (2007), no. 1, 91–131.
21. Christian Kleiber and Jordan Stoyanov, *Multivariate distributions and the moment problem*, Journal of Multivariate Analysis **113** (2013), 7–18.
22. Yin-Ting Liao and Kavita Ramanan, *Geometric sharp large deviations for random projections of ℓ_p^n spheres and balls*, arXiv preprint arXiv:2001.04053 (2020), 1–38.
23. Odalric Maillard and Rémi Munos, *Compressed least-squares regression*, Advances in neural information processing systems **22** (2009), 1–9.
24. Ralph Tyrrell Rockafellar, *Convex Analysis*, Princeton University Press, 2015.
25. Ralph Tyrrell Rockafellar and Roger Wets, *Variational Analysis*, vol. 317, Springer Science & Business Media, 2009.

26. Gideon Schechtman and Joel Zinn, *On the volume of the intersection of two L_p^n balls*, Proceedings of the American Mathematical Society **110** (1990), no. 1, 217–224.
27. Roman Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge University Press, 2018.
28. Vladimir Antonovich Zorich and Octavio Paniagua, *Mathematical Analysis II*, Springer, 2016.