Electronic Journal of Probability

Electron. J. Probab. **28** (2023), article no. 169, 1–23.
ISSN: 1083-6489 https://doi.org/10.1214/23-EJP1023

# Large deviation principles induced by the Stiefel manifold, and random multidimensional projections*

Steven Soojin Kim† Kavita Ramanan‡

*This article is dedicated to the memory of Elizabeth Meckes*

**Abstract**

For fixed positive integers $k < n$, given an $n$-dimensional random vector $X^{(n)}$, consider its $k$-dimensional projection $\mathbf{a}_{n,k}^{\intercal} X^{(n)}$, where $\mathbf{a}_{n,k}$ is an $n \times k$-dimensional matrix belonging to the Stiefel manifold $\mathbb{V}_{n,k}$ of orthonormal $k$-frames in $\mathbb{R}^n$. For a class of sequences $\{X^{(n)}\}_{n\in\mathbb{N}}$ that includes uniform distributions on suitably scaled $\ell_p^n$ balls, $p \in (1, \infty]$, and product measures with sufficiently light tails, it is shown that the sequence of projected vectors $\{\mathbf{a}_{n,k}^{\intercal} X^{(n)}\}_{n\in\mathbb{N}}$ satisfies a large deviation principle whenever the empirical measures of the rows of $\sqrt{n}\mathbf{a}_{n,k}$ converge, as $n \to \infty$, to a probability measure on $\mathbb{R}^k$. In particular, this implies a (quenched) large deviation principle for the sequence $\{\mathbf{a}_{n,k}^{\intercal} X^{(n)}\}_{n\in\mathbb{N}}$ for almost every realization $\{\mathbf{a}_{n,k}\}_{n\in\mathbb{N}}$ of $\{\mathbf{A}_{n,k}\}_{n\in\mathbb{N}}$, where each $\mathbf{A}_{n,k}$ is a random matrix, independent of $\{X^{(n)}\}_{n\in\mathbb{N}}$, that is distributed according to the normalized Haar measure on $\mathbb{V}_{n,k}$. Moreover, a variational formula is obtained for the rate function of the large deviation principle for the *annealed* projections $\{\mathbf{A}_{n,k}^{\intercal} X^{(n)}\}_{n\in\mathbb{N}}$, in terms of a family of quenched rate functions and a modified entropy term. A key step in this analysis is a large deviation principle for the sequence of empirical measures of rows of the random matrices $\sqrt{n}\mathbf{A}_{n,k}$, $n \geq k$, which may be of independent interest. The study of multidimensional random projections of high-dimensional measures is of interest in asymptotic functional analysis, convex geometry and statistics. Prior results on quenched large deviations for random projections of $\ell_p^n$ balls have been essentially restricted to the one-dimensional setting.

**Keywords:** large deviations; Stiefel manifold; random projections; quenched; annealed; rate function; $\ell_p^n$ balls; asymptotic convex geometry; variational formula.
**MSC2020 subject classifications:** 60F10; 60B20; 52A23.
Submitted to EJP on May 13, 2021, final version accepted on September 14, 2023.

*This work reports results from Steven Soojin Kim's PhD Thesis. Kavita Ramanan was supported in part by NSF-DMS Grants 1713032 and 1954351.

†Brown University. E-mail: steven_kim@alumni.brown.edu
‡Brown University. E-mail: Kavita_Ramanan@brown.edu, https://www.brown.edu/academics/applied-mathematics/faculty/kavita-ramanan/home

# 1 Introduction

## 1.1 Background

The study of high-dimensional measures and their lower-dimensional projections is a central theme in high-dimensional probability, asymptotic functional analysis and convex geometry, where in the latter case the measures of interest are distributions on convex bodies, which are compact, convex sets with non-empty interior (see, e.g., [21, 25]). Multidimensional projections of high-dimensional random vectors are also relevant in statistics, data analysis and computer science [8, 13]. Recent work has shown that large deviation principles (LDPs) that capture the tail behavior of lower-dimensional random projections can provide more interesting information about the original high-dimensonal measures than central limit theorem type results on fluctuations that capture universal phenomena. For example, in the case of $\ell_p^n$ balls, $p \in [1, \infty)$, which are fundamental objects in convex geometry, this was first illustrated by LDPs for one-dimensional projections obtained in [10, 18], and subsequently by LDPs for norms of samples from $\ell_p^n$ balls and their multidimensional projections in [2, 15, 20], as well as corresponding refined large deviation estimates obtained in [22, 23, 17]. LDPs of random projections of high-dimensional measures are broadly of two types, the terminology arising from statistical physics: so-called "quenched" LDPs, where one conditions on the choice of the (sequence of) sub-spaces, bases or directions onto which one projects; or "annealed" LDPs, which average over the randomness arising in the choice of the projection. While most of the work described above on $\ell_p^n$ balls focused on one-dimensional LDPs (either for one-dimensional projections or norms of higher-dimensional projections), in [20] annealed LDPs were also established for multidimensional projections of high-dimensional measures that satisfy a general condition called the asymptotic thin shell condition, and associated refined annealed LDPs (under additional conditions) were obtained in [23]. The asymptoptic thin shell condition and its refinement were shown to be satisfied in [20, 23] by several classes of measures, including product measures with polynomial tail decay, $\ell_p^n$ balls, $p \in [1, \infty]$, and classes of Orlicz balls and Gibbs measures.

In this article, we establish quenched LDPs for multidimensional random projections of a class of $n$-dimensional measures onto a subspace of fixed dimension $k$, as $n$ goes to infinity. Quenched LDPs and their refinements can often provide more geometric information than annealed LDPs because they can be sensitive to projection directions, and thus distinguish how the high-dimensional body looks along different directions. For example, it was shown in [10, Theorem 2.6] that the large deviation decay rate of the tail probabiilty for scaled one-dimensional projections of a vector distributed on a normalized $\ell_1^n$ ball depends on a certain scaled limit of the maximum coordinate of the projection directions. More significantly, the refined estimates of quenched tail probabilities for one-dimensional projections of $\ell_n^p$ balls and spheres obtained in [22] show a dependence, for *all* $2 < p < \infty$, on the sequence of projection directions, in a way that provides insight into their geometry (see [22, Theorem 2.4 and Remark 2.7] for further discussion). However, the analysis of quenched LDPs is typically more difficult than annealed LDPs because one can no longer exploit symmetry properties of the random projection measure that are available in the annealed setting.

To state our results more precisely, for $k \in \mathbb{R}^n$, let $I_k$ denote the $k \times k$ identity matrix, and for $n > k$, let

$$\mathbb{V}_{n,k} := \{A \in \mathbb{R}^{n \times k} : A^{\intercal} A = I_k\} \tag{1.1}$$

denote the Stiefel manifold of orthonormal $k$-frames in $\mathbb{R}^n$. Observe that the set $\mathbb{V}_{n,n}$ can be identified with the set $\mathcal{O}(n)$ of $n \times n$ orthogonal matrices with columns of norm 1. Also, note that for $k, n \in \mathbb{N}$, $k < n$, any element $a_{n,k} \in \mathbb{V}_{n,k}$ defines a linear projection from $n$ to $k$ dimensions. Fixing a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we consider sequences of random vectors $\{X^{(n)}\}_{n\in\mathbb{N}}$ defined on this space that satisfy a certain set of conditions (see Assumption 2.1), which includes, for example, $X^{(n)}$ uniformly distributed on an $\ell_p^n$ ball of radius $n^{1/p}$, $p \geq 2$, or $X^{(n)}$ distributed according to a product measure with sufficiently light tails. For any fixed $k \in \mathbb{N}$, let $\mathbb{N}_k := \{n \in \mathbb{N} : n > k\}$, and consider the sequence of $k$-dimensional projections

$$\{n^{-1/2}\mathbf{a}_{n,k}^{\intercal} X^{(n)}\}_{n\in\mathbb{N}_k}, \tag{1.2}$$

with $\mathbf{a}_{n,k} \in \mathbb{V}_{n,k}$ for each $n \in \mathbb{N}_k$. Also, denoting $\mathbf{a} := \{\mathbf{a}_{n,k}\}_{n\in\mathbb{N}_k}$, let $\mathsf{L}_{n,k}^{\mathbf{a}}$ be the associated sequence of empirical measures of the rows $\sqrt{n}\mathbf{a}_{n,k}(i,\cdot)$, $1 \leq i \leq n$, of $\sqrt{n}\mathbf{a}_{n,k}$:

$$\mathsf{L}_{n,k}^{\mathbf{a}} := \frac{1}{n}\sum_{i=1}^{n} \delta_{\sqrt{n}\mathbf{a}_{n,k}(i,\cdot)}, \quad n \in \mathbb{N}_k, \tag{1.3}$$

where $\delta_y$ represents the Dirac delta measure at $y$. Our first result, Theorem 2.4, shows that whenever $\{\mathsf{L}_{n,k}^{\mathbf{a}}\}_{n\in\mathbb{N}_k}$ converges to a probability measure $\nu$ on $\mathbb{R}^k$ in the $q_\star$-Wasserstein topology for a suitably chosen $q_* > 0$ (see Definition 1.1), then the sequence of random projections $\{n^{-1/2}\mathbf{a}_{n,k}^{\intercal} X^{(n)}\}_{n\in\mathbb{N}_k}$, satisfies an LDP on $\mathbb{R}^k$ with a rate function that we denote by $\mathcal{J}_\nu^{\mathsf{qu}}$. In particular, this implies a quenched LDP for the sequence $\{\mathbf{A}_{n,k}^{\intercal} X^{(n)}\}_{n\in\mathbb{N}_k}$, where the random matrix

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

is independent of $\{X^{(n)}\}_{n\in\mathbb{N}}$ and distributed according to $\sigma_{n,k}$, the normalized Haar measure on $\mathbb{V}_{n,k}$ (i.e., the unique probability measure on $\mathbb{V}_{n,k}$ that is invariant under the group $\mathcal{O}(n)$ of orthogonal transformations). In [20], it was shown that $\{\mathbf{A}_{n,k}^{\intercal} X^{(n)}\}_{n\in\mathbb{N}_k}$ also satisfies an annealed LDP. Our second result, Theorem 2.7, establishes a variational formula for the annealed rate function $\mathcal{J}^{\mathsf{an}}$ in terms of the quenched rate functions $\mathcal{J}_\nu^{\mathsf{qu}}$. Along the way, we establish a result that may be of independent interest (see Theorem 2.8), which states that for any $q \in (0, 2)$, an LDP in the $q$-Wasserestein topology for the random empirical measure sequence $\{\mathsf{L}_{n,k}^{\mathbf{A}}\}_{n\in\mathbb{N}_k}$, where $\mathsf{L}_{n,k}^{\mathbf{A}}$ is defined as in (1.3) but with $\mathbf{A} := (\mathbf{A}_{n,k})_{n\in\mathbb{N}}$ in place of $\mathbf{a}$. Subsequent to the appearance of this work, the article by [14] used projective limits to establish an LDP for the sequence of random matrices themselves $(\mathbf{A}_{n,k})_{n\in\mathbb{N}}$ (rather than just the empirical measures of their rows) and applied that to study LDPs for the laws of the images of product distributions under $\mathbf{A}_{n,k}$. Another related work that appeared subsequent to this work is [24], which establishes a quenched LDP for projections of random vectors $X^{(n)}$ uniformly distributed on scaled $\ell_p^n$ balls onto spaces of possibly growing dimension $k_n$ when $k_n$ is growing sublinearly.

In the next section, we introduce some basic notation and terminology that will be used throughout, and then provide precise statements of our main results in Section 2, with proofs presented in Sections 3–6.

### 1.2 Notation, basic definitions and classical results

For $p \in [1, \infty]$, let $\|\cdot\|_p$ denote the $\ell_p^k$ norm on $\mathbb{R}^k$. When $p = 2$, and when clear from the context, we will omit the subscript and simply write $\|\cdot\|$ for the Euclidean norm. Let $\mathcal{P}(\mathbb{R}^k)$ denote the space of probability measures on the Euclidean space $\mathbb{R}^k$, endowed with its Borel $\sigma$-algebra. By default we will assume $\mathcal{P}(\mathbb{R}^k)$ is equipped with the topology of weak convergence. We also consider the following restricted subsets of probability measures: for $q > 0$, let

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

**Definition 1.1** (Wasserstein topology)**.** *For $q > 0$, we say a sequence of probability measures $\{\nu_n\}_{n\in\mathbb{N}} \subset \mathcal{P}_q(\mathbb{R}^d)$ converges to a limit $\nu$ with respect to the $q$-*Wasserstein topology *if we have both weak convergence, denoted $\nu_n \Rightarrow \nu$, as well as convergence of $q$-th moments $\int_{\mathbb{R}^d} \|x\|^q \nu_n(dx) \to \int_{\mathbb{R}^d} \|x\|^q \nu(dx)$. As noted in [33, Section 6], the $q$-Wasserstein topology is metrizable through the $q$-Wasserstein metric, which we denote by $\mathcal{W}_q$.*

Next, we recall the definition of an LDP; see [7, Section 1.2].

**Definition 1.2** (Large deviation principle)**.** *Let $\mathcal{X}$ be a topological space with Borel sigma-algebra $\mathcal{B}$. A sequence of probability measures $\{P_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(\mathcal{X})$ is said to satisfy* a large deviation principle (LDP) *in $\mathcal{X}$ with rate function $I : \mathcal{X} \to [0,\infty]$ if for all $B \in \mathcal{B}$,*

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

*where $B^\circ$ and $\bar{B}$ denote the interior and closure of $B$, respectively. We say $I$ is a* good rate function (GRF) *if it has compact level sets. Analogously, a sequence of $\mathcal{X}$-valued random variables $\{\mathsf{Y}_n\}_{n\in\mathbb{N}}$ is said to satisfy an LDP with GRF $I$ if the sequence of their laws $\{\mathbb{P} \circ \mathsf{Y}_n\}_{n\in\mathbb{N}}$ does.*

We now recap some standard results from large deviations theory that we will frequently invoke. We start with the contraction principle (see, e.g., [7, Theorem 4.2.1]) that allows one to transfer LDPs for one sequence to another related sequence.

**Theorem 1.3** (Contraction principle)**.** *Let $\mathcal{X}$ and $\mathcal{X}'$ be Hausdorff topological spaces and $f : \mathcal{X} \to \mathcal{X}'$ a continuous map. Suppose $I : \mathcal{X} \to [0,\infty]$ is a GRF on $\mathcal{X}$, and define*

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

*where as usual the infimum over an empty set is taken to be infinity. Then $I'$ is a GRF on $\mathcal{X}'$ and moreover, if $\{P_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(\mathcal{X})$ satisfies an LDP on $\mathcal{X}$ with rate function $I$ then $\{P_n \circ f^{-1}\}_{n\in\mathbb{N}}$ satisfies an LDP on $\mathcal{X}'$ with rate function $I'$.*

Next, we state a result that allows one to strengthen the topology with respect to which an LDP is established. We start with the notion of exponential tightness.

**Definition 1.4** (Exponential tightness)**.** *Let $\mathcal{X}$ be a topological space equipped with a $\sigma$-algbera that contains the Borel $\sigma$-algebra. The sequence of probability measures $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ on $\mathcal{X}$ is said to be exponentially tight if for every $\alpha < \infty$, there exists a compact set $K_\alpha \subset \mathcal{X}$ such that*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_n\left(K_\alpha^c\right) < -\alpha.$$

**Theorem 1.5** (LDP on a finer topology)**.** *Suppose $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ satisfies an LDP on $\mathcal{X}$ with respect to a Hausdorff topology $\tau$ on $\mathcal{X}$, and suppose $\tau'$ is a finer topology on $\mathcal{X}$. If $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ is an exponentially tight family of probability measures on $\mathcal{X}$ with respect to $\tau'$, then $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ also satisfies an LDP on $\mathcal{X}$ equipped with the topology $\tau'$, with the same rate function.*

We now introduce some preliminary definitions. Given a function $f : \mathbb{R}^m \to [-\infty,\infty]$ for some $m \in \mathbb{N}$, we recall that its *Legendre transform* $f^*$ is defined as follows:

$$f^*(t) := \sup_{s\in\mathbb{R}^m} \left[\langle s,t\rangle - f(s)\right], \quad t \in \mathbb{R}^m.$$

Also, given an (extended real-valued) function $f$ defined on a Euclidean space $\mathcal{X}$, its domain, denoted as $D_f$, is defined to be the subset of points in $\mathcal{X}$ for which $f$ is finite. We now state the definition of an essentially smooth function; see, for example, [7, Definition 2.3.5].

**Definition 1.6** (Essentially smooth functions)**.** *An extended real-valued function* $f$ : $\mathbb{R}^m \to [-\infty, \infty]$ *is said to be* essentially smooth *if* $D_f \neq \emptyset$*,* $f$ *is differentiable in the interior* $D_f^\circ$ *of* $D_f$*, and* $f$ *is "steep" (i.e., if* $D_f$ *has a boundary* $\partial D_f$*, then* $\lim_{t \to \partial D_f} ||\nabla f(t)|| = \infty$*).*

We conclude by stating a basic LDP result, namely Cramér's theorem, in the generality that we need (see, e.g., [7, Corollary 6.1.6]).

**Theorem 1.7** (Cramér's theorem)**.** *Let* $\{\mathsf{Y}_n\}$ *be a sequence of independent and identically distributed (i.i.d.)* $\mathbb{R}^m$*-valued random vectors with common log moment generating function (mgf)* $M(t) := \mathbb{E}[e^{tX_1}], t \in \mathbb{R}$*, such that* $0$ *lies in the interior* $D_M^\circ$ *of* $D_M$*. Then the sequence of empirical means* $S_n := \frac{1}{n}\sum_{i=1}^n \mathsf{Y}_i$*,* $n \in \mathbb{N}$*, satisfies an LDP with rate function* $M^*$*.*

## 2 Main results

We now provide a precise statement of our results. For the quenched LDP, we impose the following assumption, noting that only a subset of the enumerated conditions will be used for some of the results.

**Assumption 2.1** (Assumptions on the high-dimensional vectors)**.** *The sequence of random vectors* $\{X^{(n)}\}_{n\in\mathbb{N}}$ *satisfies the following properties:*

*(i)* Representation: *there exists a sequence of i.i.d. real-valued random variables* $\{\xi_j\}_{j\in\mathbb{N}}$*, a Borel measurable function* $\mathbf{r} : \mathbb{R} \to \mathbb{R}_+$*, and a continuous function* $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ *such that*

$$X^{(n)} \overset{(d)}{=} \xi^{(n)} \cdot \rho\left(\frac{1}{n}\sum_{i=1}^n \mathbf{r}(\xi_i)\right), \quad n \in \mathbb{N},$$

*where* $\xi^{(n)} := (\xi_1, \ldots, \xi_n)$*. Let* $\Lambda$ *denote the log mgf of* $(\xi_1, \mathbf{r}(\xi_1))$*:*

$$\Lambda(s_1, s_2) := \log \mathbb{E}\left[\exp\left(s_1\xi_1 + s_2\mathbf{r}(\xi_1)\right)\right], \quad s_1 \in \mathbb{R}, s_2 \in \mathbb{R}. \tag{2.1}$$

*(ii)* Growth of the log mgf: *There exists* $q_\star > 0$ *such that for every* $s_2 \in \{s \in \mathbb{R} : (s_1, s) \in D_\Lambda$ *for some* $s_1 \in \mathbb{R}\}$*, there exists a finite constant* $C_{s_2} > 0$ *(depending only on* $s_2$*) such that*

$$\Lambda(s_1, s_2) \leq C_{s_2}(1 + |s_1|^{q_\star}), \quad \text{for every } (s_1, s_2) \in D_\Lambda. \tag{2.2}$$

*Furthermore, there exists* $T \leq \infty$ *such that* $D_\Lambda = \mathbb{R} \times (-\infty, T)$*.*

*(iii)* Regularity of the integrated log mgf: *For any* $\nu \in \mathcal{P}(\mathbb{R}^k)$*, the function* $\Psi_\nu : \mathbb{R}^{k+1} \to \mathbb{R}$ *obtained as an integrated form of the log mgf,*

$$\Psi_\nu(t_1, t_2) := \int_{\mathbb{R}^k} \Lambda\left(\langle t_1, x\rangle, t_2\right) \nu(dx), \quad t_1 \in \mathbb{R}^k, t_2 \in \mathbb{R}, \tag{2.3}$$

*contains 0 in the interior of its domain, is lower semicontinuous, and is essentially smooth in the sense of Definition 1.6.*

*(iv)* Properties of a related log mgf: *The log mgf* $\bar{\Lambda}$ *of* $(\xi_1^2, \mathbf{r}(\xi_1))$*, given by*

$$\bar{\Lambda}(s_1, s_2) := \log \mathbb{E}[\exp(s_1\xi_1^2 + s_2\mathbf{r}(\xi_1))], \qquad s_1 \in \mathbb{R},\ s_2 \in \mathbb{R}, \tag{2.4}$$

*is finite in a non-empty neighborhood of the origin* $(0, 0)$*.*

*(v)* Tail Bound: *The exponent* $q_\star$ *of part (ii) is bounded above,* $q_\star < 2$*.*

**Remark 2.2.** The inequality (2.2) in Assumption 2.1(ii) implies that for $t_1 \in \mathbb{R}^k$ and $t_2 \in (-\infty, T)$, the map $\nu \mapsto \Psi_\nu(t_1, t_2)$ is continuous with respect to the $q_\star$-Wasserstein topology. Further, for $t_2 \geq T$, we have $\Lambda(y, t_2) = \infty$ for all $y \in \mathbb{R}$, and hence, (2.3) shows that $\Psi_{\bar\nu}(t_1, t_2) = \infty$ for all $t_1 \in \mathbb{R}, t_2 \geq T$ and $\nu \in \mathcal{P}(\mathbb{R}^k)$.

**Remark 2.3.** A wide class of product measures satisfy Assumption 2.1 with $\rho \equiv \mathbf{r} \equiv 1$; namely those that have sufficiently light tails, in the sense of parts (iv) and (v). Examples of sequences of non-product measures satisfying Assumption 2.1 are $\ell_p^n$ spheres. More precisely, fix $p \in [1, \infty)$, and for $n \in \mathbb{N}$, let $\mathbb{D}_{n,p} := \{x \in \mathbb{R}^n : \sum_{i=1}^n |x_i|^p \leq n\}$ be the scaled $\ell_p^n$ ball in $\mathbb{R}^n$, let $\mathbb{S}_p^{n-1} := \partial \mathbb{D}_{n,p}$ be the scaled $\ell_p^n$ sphere in $\mathbb{R}^n$, let $\eta_{n,p}$ be the cone measure on $\mathbb{S}_p^{n-1}$: for Borel subsets $S \subset \mathbb{S}_p^{n-1}$,

$$\eta_{n,p}(S) := \frac{\mathrm{vol}_n(\{cx : x \in S, c \in [0, n^{1/p}]\})}{\mathrm{vol}_n(\mathbb{D}_{n,p})},$$

with $\mathrm{vol}_n$ denoting Lebesgue measure on $\mathbb{R}^n$, and let $X^{(n)} = X^{(n,p)}$ be distributed according to $\eta_{n,p}$. When $p = 2$, we will simply write $\mathbb{S}^{n-1}$ for $\mathbb{S}_2^{n-1}$. Then we have the following observations on properties (i)–(v) of Assumption 2.1:

(i) for $p \in [1, \infty)$, the representation follows from results in [30, 28], with $\{\xi_j\}_{j \in \mathbb{N}}$ being the i.i.d. sequence with common law equal to the generalized $p$-normal distribution (namely, the probability measure on $\mathbb{R}$ with density proportional to $e^{-|y|^p/p}$), $\mathbf{r}(x) = |x|^p$, and $\rho(y) = y^{-1/p}$;

(ii) for $p \in (1, \infty)$, the growth conditions on the log mgf $\Lambda$ are satisfied by [10, Lemma 5.7]; further, $\Lambda$ is symmetric in its first argument due to the symmetry of the generalized $p$-normal distribution;

(iii) for $p \in (1, \infty)$, the conditions on the integrated log mgf were established in [10, Lemma 5.9];

(iv) for $p \in [2, \infty)$, the log mgf condition is easily verified;

(v) for $p \in (2, \infty)$, the precise tail bound exponent was established in [10, Lemma 5.5].

We now our state our first result, whose proof is deferred to Section 5. Recall the $q$-Wasserstein metric $\mathcal{W}_q$ specified in Definition 1.1. Also, for any $\nu \in \mathcal{P}(\mathbb{R}^k)$, we let $\Psi_\nu^*$ denote the Legendre transform of $\Psi_\nu$,

$$\Psi_\nu^*(\tau_1, \tau_2) := \sup_{t_1 \in \mathbb{R}^k, t_2 \in \mathbb{R}} \{\langle t_1, \tau_1 \rangle + t_2 \tau_2 - \Psi_\nu(t_1, t_2)\}, \quad \tau_1 \in \mathbb{R}^k, \tau_2 \in \mathbb{R}. \tag{2.5}$$

Also, let $\gamma$ denote the standard Gaussian distribution on $\mathbb{R}$, and $\gamma^{\otimes k}$ its $k$-fold product.

**Theorem 2.4** (Quenched LDP for multidimensional projections)**.** *Fix $k \in \mathbb{N}$, and suppose $\{X^{(n)}\}_{n \in \mathbb{N}}$ satisfies Assumption 2.1(i, ii, iii) with associated constant $q_\star > 0$ and integrated log mgf functional $\nu \to \Psi_\nu$. Choose any sequence $\mathbf{a} = \{\mathbf{a}_{n,k}\}_{n \in \mathbb{N}_k}$, $\mathbf{a}_{n,k} \in \mathbb{V}_{n,k}$, $n \in \mathbb{N}$, such that the sequence of empirical measures $(\mathsf{L}_{n,k}^{\mathbf{a}})_{n \in \mathbb{N}_k} \subset \mathcal{P}(\mathbb{R}^k)$ defined in* (1.3) *satisfy*

$$\mathcal{W}_{q_\star}(\mathsf{L}_{n,k}^{\mathbf{a}}, \bar\nu) \to 0 \quad \text{as } n \to \infty, \tag{2.6}$$

*for some $\bar\nu \in \mathcal{P}(\mathbb{R}^k)$. Then the following claims hold:*

*(i) The sequence $\{n^{-1/2} \mathbf{a}_{n,k}^\intercal X^{(n)}\}_{n \in \mathbb{N}_k}$ satisfies an LDP in $\mathbb{R}^k$ with GRF $\mathcal{J}_{\bar\nu}^{\mathsf{qu}} : \mathbb{R}^k \to [0, \infty]$ defined by*

$$\mathcal{J}_{\bar\nu}^{\mathsf{qu}}(x) := \inf_{\tau \in \mathbb{R}_+} \Psi_{\bar\nu}^*\left(\tfrac{x}{\rho(\tau)}, \tau\right), \quad x \in \mathbb{R}^k. \tag{2.7}$$

(ii) *If $\sigma$ is any probability measure on $\mathbb{S} := \otimes_{n>k}\mathbb{V}_{n,k}$ whose $n$-th marginal coincides with the Haar measure $\sigma_{n,k}$, then for $\sigma$-a.e. $\mathbf{a} = \{\mathbf{a}_{n,k}\}_{n\in\mathbb{N}_k} \in \mathbb{S}$, the sequence $\{n^{-1/2}\mathbf{a}_{n,k}^{\intercal}X^{(n)}\}_{n\in\mathbb{N}_k}$ satisfies an LDP in $\mathbb{R}^k$ with GRF $\mathcal{J}^{\mathsf{qu}}_{\gamma^{\otimes k}}$.*

(iii) *Let $U$ be a uniformly distributed random variable on $[0,1]$, independent of $\{X^{(n)}\}_{n\in\mathbb{N}}$. If the log mgf $\Lambda$ of (2.1) is symmetric in its first argument, then the claims (i) and (ii) also hold for the sequence $\{n^{-1/2}\mathbf{a}_{n,k}^{\intercal}X^{(n)}U^{1/n}\}_{n\in\mathbb{N}_k}$.*

**Remark 2.5.** Claim (iii) of Theorem 2.4 is motivated by the observation that if $X^{(n,p)}$ is distributed according to the cone measure on the scaled $\ell_p^n$ sphere $\mathbb{S}_p^{n-1}$, then the random variable $U^{1/n}X^{(n,p)}$ is uniformly distributed on the scaled $\ell_p^n$ ball $\mathbb{D}_{n,p}$ [30]. Since, as noted in Remark 2.3, $X^{(n,p)}$ satisfies Assumption 2.1 across a wide range of $p$ (with symmetric $\Lambda$), Theorem 2.4(iii) allows an extension of the LDP results in (i) and (ii) of Theorem 2.4 from $\ell_p^n$ spheres to $\ell_p^n$ balls, which are of greater interest in convex geometry.

Note that the rate function $\mathcal{J}^{\mathsf{qu}}_{\bar{\nu}}$ depends only on the limit $\bar{\nu}$ in (2.6), and is insensitive to further specifics of the projection matrix sequence $\mathbf{a}$. For one-dimensional projections ($k=1$), Theorem 2.4 recovers both [9, Theorem 2], which addresses the case where $X^{(n)}$ has a product distribution, and [10, Theorem 2.5 and Proposition 5.3], which consider the case when $X^{(n)}$ is uniformly distributed on $\mathbb{D}_{n,p}$ or according to the cone measure $\eta_{n,p}$ (as defined in Remark 2.2). One setting of multidimensional projections ($k>1$) considered prior to the above result is the LDP for the projection of $X^{(n)}$ onto the first $k$ canonical directions, which corresponds to $\mathbf{a}_{n,k}$ being equal to the matrix of 1s on the diagonal and 0s elsewhere, (more precisely, $\mathbf{a}_{n,k}(i,j)$ is equal to $1$ if $i=1,\ldots,k$ and $j=i$, and is equal to $0$ otherwise), for which (2.6) does not hold. More recent work [16] establishes asymptotics (law of large numbers and LDPs) for the shape of multidimensional projections of the uniform distribution on a cube or discrete cube. The current article differs from these works by establishing almost everywhere quenched LDP results, first reported in the PhD thesis [18] for multidimensional projections beyond the particular cases of the canonical projection and product measures. Theorem 2.4 provides a potential starting point for obtaining asymptotic results for shapes and instrinsic volumes of projections of non-product measures such as $\ell_p^n$ balls, as well as for obtaining sharp quenched large deviation estimates for multidimensional projections and their norms, which are relevant for understanding volumetric properties of convex bodies and their intersections.

Our second main result concerns a variational representation of the annealed rate function for the sequence of random multidimensional projections $\{\mathbf{A}_{n,k}^{\intercal}X^{(n)}\}_{n\in\mathbb{N}_k}$. We start by stating an annealed LDP counterpart to Theorem 2.4, specialized to the setting considered in this article.

**Theorem 2.6** (Annealed LDP for multidimensional projections)**.** *Consider a sequence $\{X^{(n)}\}_{n\in\mathbb{N}}$ that satisfies Assumption 2.1(i, iv) with associated $\{\xi_j\}_{j\in\mathbb{N}}$, $\rho$, and $\bar{\Lambda}$. Then, for any $k\in\mathbb{N}$, $\{n^{-1/2}\mathbf{A}_{n,k}^{\intercal}X^{(n)}\}_{n\in\mathbb{N}_k}$ satisfies an LDP in $\mathbb{R}^k$ with GRF $\mathcal{J}^{\mathsf{an}} : \mathbb{R}^k \to [0,\infty]$ defined by*

$$\mathcal{J}^{\mathsf{an}}(x) := \inf_{c>0}\left\{J_X\left(\tfrac{\|x\|_2}{c}\right) - \tfrac{1}{2}\log(1-c^2)\right\}, \quad x\in\mathbb{R}^k,$$

*where $J_X$ is given, in terms of the Legendre transform $\bar{\Lambda}^*$ of $\bar{\Lambda}$, by*

$$J_X(x) := \inf_{(t_1,t_2)\in\mathbb{R}_+^2}\left\{\bar{\Lambda}^*(t_1,t_2) : x = t_1^{1/2}\rho(t_2)\right\} = \inf_{t_2>0}\bar{\Lambda}^*\left(\frac{x^2}{\rho^2(t_2)}, t_2\right).$$

*Proof.* It follows from [20, Theorem 2.7] that $\{n^{-1/2}\mathbf{A}_{n,k}^{\intercal}X^{(n)}\}_{n\in\mathbb{N}_k}$ satisfies an LDP in $\mathbb{R}^k$ with GRF $\mathcal{J}^{\mathsf{an}}$ as defined above whenever Assumption A* therein is satisfied with speed $s_n = n$, namely, when the sequence of scaled norms $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP with GRF $J_X$. Since the domain of $\bar{\Lambda}$ contains a neighborhood of the origin due to Assumption 2.1(iv), Cramér's theorem (Theorem 1.7) implies that the sequence $\{\left(\frac{1}{n}\sum_{i=1}^n \xi_i^2, \frac{1}{n}\sum_{i=1}^n \mathbf{r}(\xi_i)\right)\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^2$ with GRF $\bar{\Lambda}^*$. Since Assumption 2.1(i) implies

$$\frac{\|X^{(n)}\|_2}{\sqrt{n}} \overset{(d)}{=} \left(\frac{1}{n}\sum_{i=1}^n \xi_i^2\right)^{1/2} \rho\left(\frac{1}{n}\sum_{i=1}^n \mathbf{r}(\xi_i)\right),$$

with $\rho$ continuous, the contraction principle (Theorem 1.3) shows that the sequence of scaled norms $\{\|X^{(n)}\|_2/\sqrt{n}\}_{n\in\mathbb{N}}$ satisfies an LDP with GRF $J_X$. This completes the proof. □

To state the variational representation for the rate function $\mathcal{J}^{\mathsf{an}}$, we first introduce some notation. For $k \in \mathbb{N}$ and $\nu, \mu \in \mathcal{P}(\mathbb{R}^k)$, define the *relative entropy* of $\nu$ with respect to $\mu$ as

$$H(\nu|\mu) := \int_{\mathbb{R}} \log\left(\frac{d\nu}{d\mu}\right) d\nu \tag{2.8}$$

if $\nu \ll \mu$, and $H(\nu|\mu) := +\infty$ otherwise. Recall that $\gamma$ denotes the standard Gaussian measure on $\mathbb{R}$, and for $\nu \in \mathcal{P}(\mathbb{R}^k)$, let $\mathcal{C} : \mathcal{P}(\mathbb{R}^k) \to \mathbb{R}^{k\times k}$ denote the covariance map,

$$\mathcal{C}(\nu) := \int_{\mathbb{R}^k} [x \otimes x]\, \nu(dx), \quad \nu \in \mathcal{P}(\mathbb{R}^k). \tag{2.9}$$

Recall that $I_k$ denotes the $k \times k$ identity matrix, and write $A \preceq B$ if $B - A$ is positive semidefinite, and define the modified relative entropy functional:

$$\mathbb{H}_k(\nu) := \begin{cases} H(\nu|\gamma^{\otimes k}) + \frac{1}{2}\mathrm{tr}(I_k - \mathcal{C}(\nu)) & \text{if } \mathcal{C}(\nu) \preceq I_k \\ +\infty & \text{else} \end{cases}, \quad \nu \in \mathcal{P}(\mathbb{R}^k). \tag{2.10}$$

**Theorem 2.7** (Variational formula for the GRF of the annealed LDP)**.** *Fix $k \in \mathbb{N}$, suppose that the sequence $\{X^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption 2.1. Let $\mathcal{J}_{\bar{\nu}}^{\mathsf{qu}}$ and $\mathcal{J}^{\mathsf{an}}$ be defined as in Theorems 2.4 and 2.6, respectively. Then we have the following variational formula:*

$$\mathcal{J}^{\mathsf{an}}(x) = \inf_{\bar{\nu}\in\mathcal{P}(\mathbb{R}^k)} \left\{\mathcal{J}_{\bar{\nu}}^{\mathsf{qu}}(x) + \mathbb{H}_k(\bar{\nu})\right\}, \qquad x \in \mathbb{R}^k. \tag{2.11}$$

Note that $\mathbb{H}_k(\bar{\nu}) = 0$ when $\bar{\nu} = \gamma^{\otimes k}$, which implies $\mathcal{J}^{\mathsf{an}} \le \mathcal{J}_{\gamma^{\otimes k}}^{\mathsf{qu}}$, as would be expected from Jensen's inequality given $\mathcal{J}_{\gamma^{\otimes k}}^{\mathsf{qu}}$ is simply the GRF of the quenched LDP for $\{\mathbf{A}_{n,k}^{\intercal} X^{(n)}\}_{n\in\mathbb{N}_k}$. More generally, the optimization problem (2.11) can be interpreted as saying that at the large deviation level, the decay rate of the annealed probability of a rare event is the infimum, over all random "environments" (in this case "sequence of random projection matrices", as captured by the limit $\bar{\nu}$ of the empirical measure of their rows), of the decay rate $\mathcal{J}_{\bar{\nu}}^{\mathsf{qu}}$ of the quenched probability of the rare event conditioned on that environment, plus the cost of the choice of the random environment which in this case is measured by $\mathbb{H}_k(\bar{\nu})$.

While such a relation is intuitive, rigorous proofs of such informal statements are typically non-trivial. For example, such variational representations have been rigorously established only in a few specific cases, such as LDPs for random walks in random environments on $\mathbb{Z}$ in [6] and on supercritical Galton-Watson trees in [1]. The one-dimensional case ($k = 1$) of Theorem 2.7 for $\ell_p^n$ balls recovers [10, Theorem 2.7]. The proof of the the multidimensional case stated in Theorem 2.7), which is given in Section 6, is more involved and relies on an auxiliary LDP for the following sequence of random empirical measures, analogous to those defined in (1.3):

$$\mathsf{L}_{n,k} := \mathsf{L}^{\mathbf{A}}_{n,k} = \frac{1}{n}\sum_{i=1}^{n} \delta_{\sqrt{n}\mathbf{A}_{n,k}(i,\cdot)}, \quad n \in \mathbb{N}. \tag{2.12}$$

**Theorem 2.8** (LDP for empirical measures of rows of Haar-distributed matrices)**.** *Fix $k \in \mathbb{N}$. Then $\mathbb{H}_k$ is a strictly convex GRF and for all $q \in (0,2)$, the sequence $\{\mathsf{L}_{n,k}\}_{n\in\mathbb{N}_k}$ satisfies an LDP in $\mathcal{P}_q(\mathbb{R}^k)$ with GRF $\mathbb{H}_k$.*

This theorem, which is established in Section 4, generalizes [5, Theorem 6.6], which states an LDP for the empirical measure of coordinates drawn uniformly from the sphere $\mathbb{S}^{n-1}$, which corresponds to the case $k = 1$ in our work. In contrast to this case, the $k > 1$ case necessitates more extensive computations which arise due to the non-commutative matrix setting, where the Bartlett decomposition of Proposition 3.2 replaces the usual polar decomposition for a random vector from the sphere. Given that large deviation perspectives have informed the analysis of asymptotics for spherical integrals [11, 26], it is possible that a similar approach could inform asymptotics for integrals over the Stiefel manifold, which arise, for instance, as the normalizing constant of the matrix Bingham distribution [12], or in the study of multi-spiked random covariance matrices [27].

**Remark 2.9.** The first term in the definition (2.10) of $\mathbb{H}_k$ is $H(\cdot|\gamma^{\otimes k})$, the relative entropy functional with respect to the $k$-dimensional standard Gaussian measure, which, by Sanov's theorem (see, e.g., [7, Theorem 6.2.10]), is the large deviation rate function for the sequence of empirical measures of the rows of an $n \times k$ matrix of i.i.d. standard Gaussian elements.

**Remark 2.10.** The second term in the definition (2.10) of $\mathbb{H}_k$ arises from the orthogonality and normalization constraint defining the Stiefel manifold, and offers a way of distinguishing between Haar-distributed matrices on the Stiefel manifold and matrices with i.i.d. standard Gaussian entries at the large deviations scale. Note that because $\mathbb{P}(\mathbf{A}_{n,k}^{\intercal}\mathbf{A}_{n,k} = I_k) = 1$, we have, $\mathbb{P}$-a.s.,

$$\mathrm{tr}(I_k - \mathcal{C}(\mathsf{L}_{n,k})) = \mathrm{tr}\left(I_k - \mathbf{A}_{n,k}^{\intercal}\mathbf{A}_{n,k}\right) = 0, \quad n \in \mathbb{N}_k. \tag{2.13}$$

Nonetheless, the definition of the rate function $\mathbb{H}_k$ includes the trace term $\mathrm{tr}(I_k - \mathcal{C}(\nu))$, and $\mathbb{H}_k(\nu)$ is finite even for $\nu \in \mathcal{P}(\mathbb{R}^k)$ such that $\mathrm{tr}(I_k - \mathcal{C}(\nu)) \neq 0$, due to the fact that the statement of an LDP (Definition 1.2) involves infimization of the rate function $\mathbb{H}_k$ not over a set like $\mathcal{V}_k := \{\nu \in \mathcal{P}(\mathbb{R}^k) : I_k = \mathcal{C}(\nu)\}$, but rather over its interior and closure (in the space of probability measures). In particular, the example set $\mathcal{V}_k$ is neither open nor closed with respect to the weak topology. In fact, it is possible to show from [34] that $\mathcal{V}_k$ is neither open nor closed with respect to *any* topology for which the sequence of empirical measures $\{\mathsf{L}_{n,k}\}_{n\in\mathbb{N}_k}$ satisfies an LDP.

Outside of the large deviations literature, a different comparison between such Stiefel and Gaussian matrices can be found in [32], which analyzes expectations of sublinear convex functions of random matrices.

An immediate consequence of Theorem 2.8 is the following:

**Corollary 2.11** (A LLN for the Empirical Measure Sequence)**.** *Fix $k \in \mathbb{N}$. Then for all $q \in (0,2)$, the sequence $\{\mathsf{L}_{n,k}\}_{n\in\mathbb{N}_k}$ satisfies the strong law of large numbers in $\mathcal{P}_q(\mathbb{R}^k)$. That is, almost surely, as $n \to \infty$, we have $\mathcal{W}_q(\mathsf{L}_{n,k}, \gamma^{\otimes k}) \to 0$.*

*Proof.* By Theorem 2.8, $\mathbb{H}_k$ in (2.10) is a strictly convex rate function. Since $\mathbb{H}_k(\gamma^{\otimes k}) = 0$, $\mathbb{H}_k$ attains its unique minimum over $\mathcal{P}_q(\mathbb{R}^k)$ at $\gamma^{\otimes k}$. For $\epsilon > 0$, due to the LDP for $\{\mathsf{L}_{n,k}\}_{n\in\mathbb{N}_k}$ and the uniqueness of the minimum of $\mathbb{H}_k$, there exists $\delta > 0$ and $N \in \mathbb{N}_k$

such that for $n > N$, $\mathbb{P}(\mathcal{W}_q(\mathsf{L}_{n,k}, \gamma^{\otimes k}) > \epsilon) \leq e^{-n\delta}$, which when combined with the Borel-Cantelli Lemma yields the almost sure convergence of $\mathsf{L}_{n,k}$. □

## 3 The Bartlett decomposition and its consequences

We start by recalling the QR decomposition of a matrix. Fix $k, n \in \mathbb{N}$ with $1 \leq k \leq n$. Let $\mathbb{U}_k$ denote the space of $k \times k$ upper triangular matrices, and recall that $\mathbb{V}_{n,k}$ denotes the Stiefel manifold of orthonormal $k$-frames in $\mathbb{R}^n$.

**Definition 3.1** (QR Decomposition)**.** *The (thin)* QR *decomposition of an $n \times k$ matrix $Z \in \mathbb{R}^{n\times k}$ is the factorization $Z = QR$ of $Z$ as the product of a semi-orthogonal matrix $Q \in \mathbb{V}_{n,k} \subset \mathbb{R}^{n\times k}$ and an upper triangular matrix $R \in \mathbb{U}_k \subset \mathbb{R}^{k\times k}$. Moreover, when $Z$ is of rank $k$, there is a unique such decomposition with the diagonal elements of $R$ being all positive.*

In fact, the well known Gram-Schmidt process for a matrix $Z \in \mathbb{R}^{n\times k}$ provides an explicit QR decomposition. Let the columns of $Z$ be denoted by $z_i := Z(\cdot, i) \in \mathbb{R}^n$, for $1 \leq i \leq k$, and set

$$y_1 := z_1; \qquad q_1 := \frac{y_1}{\|y_1\|_2};$$
$$y_i := z_i - \sum_{m=1}^{i-1} \langle q_m, z_i\rangle q_m; \qquad q_i := \frac{y_i}{\|y_i\|_2}, \quad i = 2, \dots, k.$$

Then we have the decomposition $Z = QR$, where $Q = (q_1, \dots, q_k)$ and

$$R = \begin{pmatrix} \langle q_1, z_1\rangle & \langle q_1, z_2\rangle & \langle q_1, z_3\rangle & \cdots & \langle q_1, z_k\rangle \\ 0 & \langle q_2, z_2\rangle & \langle q_2, z_3\rangle & \cdots & \langle q_2, z_k\rangle \\ 0 & 0 & \langle q_3, z_3\rangle & \cdots & \langle q_3, z_k\rangle \\ \vdots & & & \ddots & \vdots \\ 0 & & & & \langle q_k, z_k\rangle \end{pmatrix}. \tag{3.1}$$

Note that we have the following relation among the elements of $R$:

$$R_{11} = \|z_1\|^{1/2} \tag{3.2}$$

and for $2 \leq j \leq k$ and $1 \leq i \leq j \leq k$,

$$R_{ij} = \langle q_i, z_j\rangle = \frac{\langle z_i, z_j\rangle - \sum_{m=1}^{i-1}\langle q_m, z_i\rangle\langle q_m, z_j\rangle}{\left(\|z_i\|^2 - \sum_{m=1}^{i-1}\langle q_m, z_i\rangle^2\right)^{1/2}} = \frac{\langle z_i, z_j\rangle - \sum_{m=1}^{i-1} R_{mi}R_{mj}}{\left(\|z_i\|^2 - \sum_{m=1}^{i-1} R_{mi}^2\right)^{1/2}}. \tag{3.3}$$

Since an $n \times k$ matrix with i.i.d. standard Gaussian entries has rank $k$ almost surely, this immediately yields the following Bartlett decomposition.

**Proposition 3.2** (Bartlett decomposition [4])**.** *Let $\mathbf{Z}_{n,k}$ be the $n \times k$ random matrix with i.i.d. standard Gaussian entries. Then $\mathbf{Z}_{n,k} = \mathbf{Q}_{n,k}\mathbf{R}_{n,k}$ has an (almost surely unique) QR decomposition with the random matrix $\mathbf{R}_{n,k}$ having positive diagonal entries. The law of $\mathbf{Q}_{n,k}$ is $\sigma_{n,k}$, the Haar measure on $\mathbb{V}_{n,k}$. Moreover, the diagonal entries of $\mathbf{R}_{n,k}$ satisfy $\mathbf{R}_{n,k}(i,i) \sim \chi_{n-i+1}$, the chi distribution with $n-i+1$ degrees of freedom, for $i = 1, \dots, k$.*

**Remark 3.3.** In fact, the random matrices $\mathbf{Q}_{n,k}$ and $\mathbf{R}_{n,k}$ of the Bartlett decomposition are independent, and moreover, the marginal law of the off-diagonal entries of $\mathbf{R}_{n,k}$ are also explicitly known; however, we will not need the latter facts for our analysis. Also, note that when $k = 1$, the Bartlett decomposition corresponds to the classical polar decomposition of the $n$-dimensional Gaussian measure.

Let $\mathsf{L}^{\mathbf{Z}}_{n,k}$ denote the (random) empirical measure of the *rows* of $\mathbf{Z}_{n,k}$,

$$\mathsf{L}^{\mathbf{Z}}_{n,k} := \frac{1}{n}\sum_{i=1}^{n} \delta_{\mathbf{Z}_{n,k}(i,\cdot)} \in \mathcal{P}(\mathbb{R}^k).$$

Then, due to Proposition 3.2, for $\mathbf{A}_{n,k}$ distributed according to the Haar measure $\sigma_{n,k}$ on $\mathbb{V}_{n,k}$, we have

$$\mathbf{A}_{n,k} \overset{(d)}{=} \mathbf{Q}_{n,k} = \mathbf{Z}_{n,k}\mathbf{R}^{-1}_{n,k}. \tag{3.4}$$

In the second equality, we use the fact that $\mathbf{R}_{n,k}$ is (almost surely) invertible, since it is an upper triangular matrix with diagonal entries that are all (almost surely) positive. Recalling the definition of $\mathsf{L}_{n,k}$ from (2.12), and using the representation (3.4), we have for any Borel set $B \subset \mathbb{R}^k$,

$$\mathsf{L}_{n,k}(B) \overset{(d)}{=} \frac{1}{n}\sum_{r=1}^{n} \delta_{\sqrt{n}\mathbf{Z}_{n,k}(r,\cdot)\mathbf{R}^{-1}_{n,k}}(B) = \mathsf{L}^{\mathbf{Z}}_{n,k}\left(B\frac{\mathbf{R}_{n,k}}{\sqrt{n}}\right). \tag{3.5}$$

Fortuitously, the relation (3.3) tells us that each element of the matrix $\mathbf{R}_{n,k}$ can be computed as a function of the *rows* of the matrix $\mathbf{Z}_{n,k}$, and can be written as the image of a linear functional of the measure $\mathsf{L}^{\mathbf{Z}}_{n,k}$. We now define this map precisely.

**Definition 3.4** (Positive Definite Matrices)**.** *Let Sym$_k$ be the space of real symmetric $k\times k$ matrices. For $L, M \in$ Sym$_k$, we write $L \succeq M$ (resp., $L \succ M$) if $L - M$ is positive semi-definite (resp., positive definite).*

We equip $\mathrm{Sym}_k \subset \mathbb{R}^{k\times k}$ with the induced Borel $\sigma$-algebra when viewing it as a measurable space, and the Frobenius norm when viewed as a Banach space.

**Definition 3.5.** *Define the map $\Gamma : \mathbb{R}^{k\times k} \to \mathbb{U}_k$ according to the following iterative procedure: for $M \in \mathbb{R}^{k\times k}$, set $\Gamma(M)_{11} := M_{11}^{1/2}$ and for $j = 2,\dots,k$ and $i = 1,\dots,j$, set*

$$\Gamma(M)_{ij} := \frac{M_{ij} - \sum_{m=1}^{i-1}\Gamma(M)_{mi}\,\Gamma(M)_{mj}}{\left(M_{ii} - \sum_{m=1}^{i-1}\Gamma(M)^2_{mi}\right)^{1/2}}. \tag{3.6}$$

**Remark 3.6.** Note that if $M$ is symmetric and positive semi-definite, then $\Gamma(M)$ computes the *Cholesky decomposition* of $M$, so that $\Gamma(M)^{\intercal}\Gamma(M) = M$.

**Lemma 3.7.** *We have $\mathbb{P}$-a.s., $n^{-1/2}\mathbf{R}_{n,k} = \Gamma(\mathcal{C}(\mathsf{L}^{\mathbf{Z}}_{n,k}))$, where $\mathcal{C}$ is the covariance map of* (2.9).

*Proof.* The result follows from Proposition 3.2, Definition 3.1 and (3.2)–(3.3) upon noticing that for $1 \le i \le j \le k$,

$$\mathcal{C}_{ij}(\mathsf{L}^{\mathbf{Z}}_{n,k}) = \frac{1}{n}\sum_{r=1}^{n}\mathbf{Z}_{n,k}(r,i)\mathbf{Z}_{n,k}(r,j) = \frac{1}{n}\langle\,\mathbf{Z}_{n,k}(\cdot,i),\mathbf{Z}_{n,k}(\cdot,j)\,\rangle. \qquad \square$$

**Example 3.8.** For example, when $k = 1$, we have $\frac{\mathbf{R}_{n,1}}{\sqrt{n}} = \mathcal{C}_{11}(\mathsf{L}_{n,1})^{1/2} = \frac{\|\mathbf{Z}_{n,1}\|_2}{\sqrt{n}}$, and for $k = 2$, we have

$$\frac{\mathbf{R}_{n,2}}{\sqrt{n}} = \begin{pmatrix} \mathcal{C}_{11}(\mathsf{L}_{n,2}) & \frac{\mathcal{C}_{12}(\mathsf{L}_{n,2})}{\mathcal{C}_{11}(\mathsf{L}_{n,2})^{1/2}} \\ 0 & \left(\mathcal{C}_{22}(\mathsf{L}_{n,2}) - \frac{\mathcal{C}_{21}(\mathsf{L}_{n,2})^2}{\mathcal{C}_{11}(\mathsf{L}_{n,2})}\right)^{1/2} \end{pmatrix}.$$

## 4 Proof of the empirical measure large deviations

The representation (3.5) and Lemma 3.7 suggest our plan of attack for the proof of Theorem 2.8: first prove a joint LDP for $\{\mathsf{L}^{\mathbf{Z}}_{n,k}, \mathcal{C}(\mathsf{L}^{\mathbf{Z}}_{n,k})\}_{n\in\mathbb{N}_k}$; then establish an LDP for $\{\mathsf{L}_{n,k}\}_{n\in\mathbb{N}_k}$. Note that we do not attempt to directly establish an LDP for $\{\mathsf{L}_{n,k}\}_{n\in\mathbb{N}_k}$ from that for $\{\mathsf{L}^{\mathbf{Z}}_{n,k}\}_{n\in\mathbb{N}_k}$, because the map $\mathsf{L}^{\mathbf{Z}}_{n,k} \mapsto \mathsf{L}_{n,k}$ is not continuous with respect to the weak topology. Nor do we attempt to directly establish an LDP for $\{\mathsf{L}^{\mathbf{Z}}_{n,k}, \Gamma(\mathcal{C}(\mathsf{L}^{\mathbf{Z}}_{n,k}))\}_{n\in\mathbb{N}_k}$, because the map $\Gamma \circ \mathcal{C}$ is a nonlinear functional on the space of measures on $\mathbb{R}^k$. In contrast, our proposed first step is tractable precisely because $\mathcal{C}$ is a linear functional and the following result, which is stated in [20, Corollary A.2] as a corollary of [5, Proposition 6.4].

**Lemma 4.1** (approximate contraction principle)**.** *Let $\Sigma$ be a Polish space and $\mathcal{X}$ be a separable Banach space with topological dual $\mathcal{X}^*$. Let $\{\mathcal{L}_n\}_{n\in\mathbb{N}}$ be a sequence of $\mathcal{P}(\Sigma)$-valued random variables such that for each $n \in \mathbb{N}$, $\mathcal{L}_n$ is the empirical measure of $n$ i.i.d. $\Sigma$-valued random variables $\mathsf{s}_1, \dots, \mathsf{s}_n$ with common distribution $\mu$ (that does not depend on $n$). For any continuous $W : \Sigma \to \mathbb{R}$, define*

$$\widehat{\Lambda}(W) := \log \mathbb{E}[e^{W(\mathsf{s}_1)}]. \tag{4.1}$$

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

$$\mathcal{D} := \left\{ \alpha \in \mathcal{X}^* : \widehat{\Lambda}(\langle \alpha, \mathbf{c}(\cdot)\rangle) < \infty \right\}, \tag{4.2}$$

*and let $\mathcal{C}_n := \int_\Sigma \mathbf{c}(x)\mathcal{L}_n(dx)$. Lastly, define $F : \mathcal{X} \to \mathbb{R}$ as*

$$F(x) := \sup_{\alpha\in\mathcal{D}_\circ} \langle \alpha, x\rangle, \quad x \in \mathcal{X}. \tag{4.3}$$

*Then, $\{\mathcal{L}_n, \mathcal{C}_n\}_{n\in\mathbb{N}}$ satisfies an LDP with the GRF $\mathbb{I} : \mathcal{P}(\Sigma) \times \mathcal{X} \to [0, \infty]$ defined by*

$$\mathbb{I}(\nu, x) := \begin{cases} H(\nu|\mu) + F\left(x - \int_\Sigma \mathbf{c}\, d\nu\right) & \text{if } H(\nu|\mu) < \infty, \\ +\infty & \text{else,} \end{cases} \qquad \nu \in \mathcal{P}(\Sigma), x \in \mathcal{X}. \tag{4.4}$$

**Lemma 4.2.** *For any $k \in \mathbb{N}$, the sequence $\{\mathsf{L}^{\mathbf{Z}}_{n,k}, \mathcal{C}(\mathsf{L}^{\mathbf{Z}}_{n,k})\}_{n\in\mathbb{N}_k}$ satisfies an LDP in $\mathcal{P}(\mathbb{R}^k) \times \mathrm{Sym}_k$ with GRF $\mathbb{J}_k : \mathcal{P}(\mathbb{R}^k) \times \mathrm{Sym}_k \to [0, \infty]$, defined, for $\nu \in \mathcal{P}(\mathbb{R}^k)$ and $M \in \mathrm{Sym}_k$, to be*

$$\mathbb{J}_k(\nu, M) := \begin{cases} H(\nu|\gamma^{\otimes k}) + \frac{1}{2}\mathrm{tr}\left(M - \int_{\mathbb{R}^k}[z \otimes z]\, \nu(dz)\right) & \text{if } \int_{\mathbb{R}^k}[z \otimes z]\, \nu(dz) \preceq M, \\ +\infty & \text{else.} \end{cases} \tag{4.5}$$

*Proof.* We invoke the approximate contraction principle of Lemma 4.1 with the following parameters: $\Sigma = \mathbb{R}^k$; $\mathcal{X} = \mathrm{Sym}_k$ and $\mathcal{X}^* = \mathrm{Sym}_k$; $\mathbf{c}(z) := [z \otimes z]$ for $z \in \mathbb{R}^k$; $\mathcal{L}_n := \frac{1}{n}\sum_{j=1}^n \delta_{\mathsf{s}_j}$ for $\mathsf{s}_1, \mathsf{s}_2, \dots$ i.i.d. random vectors with common distribution $\gamma^{\otimes k}$; and $\mathcal{C}_n := \int_{\mathbb{R}^k} \mathbf{c}\, d\mathcal{L}_n$. Note that

$$(\mathsf{L}^{\mathbf{Z}}_{n,k}, \mathcal{C}(\mathsf{L}^{\mathbf{Z}}_{n,k})) \stackrel{(d)}{=} (\mathcal{L}_n, \mathcal{C}_n), \qquad n \in \mathbb{N}_k. \tag{4.6}$$

With $\widehat{\Lambda}$ as defined in (4.1), the domain $\mathcal{D}$ specified in (4.2) takes the form

$$\begin{aligned} \mathcal{D} &= \left\{ \zeta \in \mathrm{Sym}_k : \log \int_{\mathbb{R}^k} \exp\left(\langle \zeta, \mathbf{c}(z)\rangle\right) \gamma^{\otimes k}(dz) < \infty \right\} \\ &= \left\{ \zeta \in \mathrm{Sym}_k : \log \int_{\mathbb{R}^k} \frac{1}{(2\pi)^{k/2}} \exp\left(-z^\intercal(\tfrac{1}{2}I_k - \zeta)z\right) dz < \infty \right\} \\ &= \left\{ \zeta \in \mathrm{Sym}_k : \tfrac{1}{2}I_k - \zeta \succ 0 \right\}. \end{aligned}$$

This last expression indicates that $\mathcal{D}$ is a shifted reflection of the positive definite cone, hence open, implying that $\mathcal{D}^\circ = \mathcal{D}$. This expression for the form of $\mathcal{D}$ also makes it clear that $0 \in \mathcal{D}^\circ$. Lastly the value of the supremum in the definition of $F$ takes an explicit form due to the linearity of the trace functional and the fact that the constraint set $\mathcal{D}$ is a cone (due to the positive definiteness constraint): for $\eta \in \mathrm{Sym}_k$,

$$F(\eta) = \sup_{\zeta \in \mathrm{Sym}_k} \left\{ \mathrm{Tr}(\zeta\eta) : \zeta \prec \tfrac{1}{2} I_k \right\} = \begin{cases} \frac{1}{2}\mathrm{Tr}(\eta) & \text{if } \eta \succeq 0, \\ +\infty & \text{else.} \end{cases}$$

Therefore, (4.6) and Lemma 4.1 together imply that $\{(\mathsf{L}^{\mathbf{Z}}_{n,k}, \mathcal{C}(\mathsf{L}^{\mathbf{Z}}_{n,k}))\}_{n\in\mathbb{N}_k}$ satisfies the stated LDP. □

We now establish a relation between the GRF $\mathbb{J}_k$ of Lemma 4.2 and the GRF $\mathbb{H}_k$ of Theorem 2.8.

**Lemma 4.3.** *For any $k \in \mathbb{N}$ and $\nu \in \mathcal{P}(\mathbb{R}^k)$, given $\mathbb{H}_k$ of* (2.10)*, $\mathbb{J}_k$ of* (4.5)*, and $\Gamma$ of* (3.6)*, we have*

$$\mathbb{H}_k(\nu) = \widetilde{\mathbb{H}}_k(\nu) := \inf_{M\in\mathrm{Sym}_k} \mathbb{J}_k(\nu(\,\cdot\times\Gamma(M)^{-1}), M).$$

*Proof.* As noted in Remark 3.6, for $M \in \mathrm{Sym}_k$ and $\Gamma$ as in (3.6), we have $M = \Gamma(M)^\intercal\Gamma(M)$. Given this equality, the constraint $M \succeq \int_{\mathbb{R}^k}[x\otimes x]\,\nu(dx\times\Gamma(M)^{-1})$ in (4.5) can be rewritten, using the notation $\mathcal{C}$ from (2.9), as

$$I_k \succeq \int_{\mathbb{R}^k}[x\otimes x]\nu(dx) = \mathcal{C}(\nu).$$

If the preceding constraint is satisfied, then using the form of $\mathbb{J}_k$ in (4.5) in the first equality below, the chain rule for relative entropy in the third equality, and then the form of the Gaussian distribution $\gamma^{\otimes k}$, we obtain

$$\begin{aligned}
&\mathbb{J}_k(\nu(\,\cdot\times\Gamma(M)^{-1}), M)\\
&\quad= \frac{1}{2}\sum_{i=1}^k\left(M_{ii} - \int_{\mathbb{R}^k} x_i^2\,\nu(dx\times\Gamma(M)^{-1})\right) + H(\nu(\,\cdot\times\Gamma(M)^{-1})|\gamma^{\otimes k})\\
&\quad= \frac{1}{2}\sum_{i=1}^k M_{ii} - \frac{1}{2}\int_{\mathbb{R}^k} y^\intercal\,\Gamma(M)\Gamma(M)^\intercal y\,\nu(dy) + H(\nu|\gamma^{\otimes k}(\,\cdot\times\Gamma(M)))\\
&\quad= \frac{1}{2}\sum_{i=1}^k M_{ii} - \frac{1}{2}\int_{\mathbb{R}^k} y^\intercal\,\Gamma(M)\Gamma(M)^\intercal y\,\nu(dy) - \int_{\mathbb{R}^k}\log\left(\tfrac{d\gamma^{\otimes k}(\,\cdot\times\Gamma(M))}{d\gamma^{\otimes k}}\right)d\nu + H(\nu|\gamma^{\otimes k})\\
&\quad= \frac{1}{2}\sum_{i=1}^k M_{ii} - \frac{1}{2}\int_{\mathbb{R}^k} y^\intercal\,\Gamma(M)\Gamma(M)^\intercal y\,\nu(dy) + \frac{1}{2}\log\det(\Gamma(M)^{-\intercal}\Gamma(M)^{-1})\\
&\qquad\quad + \frac{1}{2}\int_{\mathbb{R}^k} y^\intercal\,(\Gamma(M)\Gamma(M)^\intercal - I_k)y\,\nu(dy) + H(\nu|\gamma^{\otimes k})
\end{aligned}$$

Due to the upper triangular structure of $\Gamma(M)$, we have

$$\frac{1}{2}\log\det(\Gamma(M)^{-\intercal}\Gamma(M)^{-1}) = -\log\det(\Gamma(M)) = -\sum_{i=1}^k \log\Gamma(M)_{ii}.$$

Also, note that $\mathrm{Tr}(I_k - \mathcal{C}(\nu)) = k - \int_{\mathbb{R}^k} y^\intercal y\,\nu(dy)$. Hence, invoking the definition of $\mathbb{H}_k$ in (2.10), we have

$$\mathbb{J}_k(\nu(\,\cdot\times\Gamma(M)^{-1}), M) = \frac{1}{2}\sum_{i=1}^k (M_{ii}-1) - \sum_{i=1}^k \log\Gamma(M)_{ii} + \mathbb{H}_k(\nu).$$

Taking the infimum of the expression above over $M \in \mathrm{Sym}_k$, we see that

$$\widetilde{\mathbb{H}}_k(\nu) = \inf_{M \in \mathrm{Sym}_k} \left\{ \sum_{i=1}^k \left( \frac{M_{ii} - 1}{2} - \log \Gamma(M)_{ii} \right) \right\} + \mathbb{H}_k(\nu). \tag{4.7}$$

Note that by the definition of $\Gamma$ in (3.6),

$$\Gamma(M)_{ii} = \left( M_{ii} - \sum_{h=1}^{i-1} \Gamma(M)_{hi}^2 \right)^{1/2},$$

and by the definition of the Gram-Schmidt process, we have $M_{ii} \geq \sum_{h=1}^{i-1} \Gamma(M)_{hi}^2 \geq 0$. Thus, for all $i = 1, \ldots, k$, for any fixed $M_{ii}$, the maximum value of $\Gamma(M)_{ii}$ is attained when $\Gamma(M)_{hi} = 0$ for $h = 1, \ldots, i-1$. Therefore, once again using $M = \Gamma(M)^\intercal \Gamma(M)$, we obtain

$$\begin{aligned} \inf_{M \in \mathrm{Sym}_k} \left\{ \sum_{i=1}^k \left( \frac{M_{ii} - 1}{2} - \log \Gamma(M)_{ii} \right) \right\} &= \inf_{M_{ii} \geq 0, i=1,\ldots,k} \left\{ \frac{1}{2} \sum_{i=1}^k (M_{ii} - 1 - \log M_{ii}) \right\} \\ &= \frac{1}{2} \sum_{i=1}^k \inf_{M_{ii} \geq 0} \{ M_{ii} - 1 - \log M_{ii} \}, \end{aligned}$$

which is clearly equal to zero. Together with (4.7), this shows that $\mathbb{H}_k = \widetilde{\mathbb{H}}_k$. ☐

**Lemma 4.4.** *Fix $k \in \mathbb{N}$ and consider the following set of probability measures,*

$$\mathcal{K} := \left\{ \nu \in \mathcal{P}(\mathbb{R}^k) : \int_{\mathbb{R}^k} \|x\|^2 \nu(dx) \leq k \right\}. \tag{4.8}$$

*For any $q \in (0,2)$, the set $\mathcal{K} \subset \mathcal{P}_2(\mathbb{R}^d)$ is compact with respect to the $q$-Wasserstein topology. In addition, $\mathcal{K}$ is convex and non-empty. Furthermore, $\mathsf{L}_{n,k}$ defined in (2.12) satisfies $\mathbb{P}(\mathsf{L}_{n,k} \in \mathcal{K}^c) = 0$ for every $n \in \mathbb{N}$.*

*Proof.* The proof of the first statement is an elementary modification of the proof of the $k = 1$ case given in [19, Lemma 3.14]. For the second statement, note that since $\mathbf{A}_{n,k}$ is almost surely supported on $\mathbb{V}_{n,k}$, $\int_{\mathbb{R}^k} |x|^2 \mathsf{L}_{n,k}(dx) = \sum_{i=1}^n \sum_{j=1}^k \mathbf{A}_{n,k}(i,j)^2 = k$ a.s., and so $\mathbb{P}(\mathsf{L}_{n,k} \in \mathcal{K}^c) = 0$. ☐

*Proof of Theorem 2.8.* Let $\Gamma$ be as defined in (3.6). Due to (3.5) and Lemma 3.7, we have

$$\mathsf{L}_{n,k} \overset{(d)}{=} \mathsf{L}_{n,k}^{\mathbf{Z}}(\,\cdot \times \tfrac{\mathbf{R}_{n,k}}{\sqrt{n}}) = \mathsf{L}_{n,k}^{\mathbf{Z}}\left(\,\cdot \times \Gamma\left(\mathcal{C}(\mathsf{L}_{n,k}^{\mathbf{Z}})\right)\right).$$

The image of $\mathcal{C}$ is positive semi-definite matrices, so as noted in Remark 3.6, the map $\Gamma$ maps a matrix to its Cholesky decomposition, hence $M \mapsto \Gamma(M)$ is continuous. By Slutsky's theorem, the map

$$\mathcal{P}(\mathbb{R}^k) \times \mathrm{Sym}_k \ni (\mu, M) \mapsto \mu(\,\cdot \times \Gamma(M)) \in \mathcal{P}(\mathbb{R}^k),$$

is also continuous. Since by Lemma 4.2 the sequence $\{\mathsf{L}_{n,k}^{\mathbf{Z}}, \mathcal{C}(\mathsf{L}_{n,k}^{\mathbf{Z}})\}_{n \in \mathbb{N}_k}$ satisfies an LDP in $\mathcal{P}(\mathbb{R}^k) \times \mathrm{Sym}_k$ with GRF $\mathbb{J}_k$, an application of the contraction principle (Theorem 1.3) to the map above yields an LDP for the sequence $\{\mathsf{L}_{n,k}\}_{n \in \mathbb{N}_k}$ in $\mathcal{P}(\mathbb{R}^k)$ (i.e., with respect to the weak topology), with GRF

$$\inf_{\nu \in \mathcal{P}(\mathbb{R}^k), M \in \mathrm{Sym}_k} \{ \mathbb{J}_k(\mu, M) : \nu = \mu(\,\cdot \times \Gamma(M)) \} = \inf_{M \in \mathrm{Sym}_k} \mathbb{J}_k(\nu(\,\cdot \times \Gamma(M)^{-1}), M) = \mathbb{H}_k(\nu),$$

where the last equality is due to Lemma 4.3.

Fix $q \in (0,2)$. In order to establish the LDP for $\{\mathsf{L}_{n,k}\}_{n\in\mathbb{N}_k}$ in $\mathcal{P}_q(\mathbb{R}^k)$ (i.e., with respect to the stronger $q$-Wasserstein topology), by Theorem 1.5 it suffices to show exponential tightness of $\{\mathsf{L}_{n,k}\}_{n\in\mathbb{N}_k}$ in the $q$-Wasserstein topology. Let $\mathcal{K}$ be the set defined in (4.8). By Lemma 4.4, $\mathcal{K}$ is compact (with respect to the $q$-Wasserstein topology) and $\mathbb{P}(\mathsf{L}_{n,k} \in \mathcal{K}^c) = 0$ for every $n \in \mathbb{N}$, which together trivially imply the exponential tightness of $\{\mathsf{L}_{n,k}\}_{n\in\mathbb{N}_k}$; see Definition 1.4.

Lastly, the strict convexity of $\mathbb{H}_k$ follows from the strict convexity of the relative entropy $H(\cdot|\gamma^{\otimes k})$ and the linearity of the covariance map $\mathcal{C}$. ☐

## 5 Proof of the quenched large deviation principle

In this section, we present the proof of Theorem 2.4. As a precursor, we state two lemmas that will assist with part (iii) of the theorem.

**Lemma 5.1.** *Fix $m \in \mathbb{N}$, and let $\mathcal{F}$ be a set of functions from $\mathbb{R}^m$ to $\mathbb{R}$ such that every $f \in \mathcal{F}$ is symmetric about 0 and convex. Then, defining $g : \mathbb{R}^m \to \mathbb{R}$ as*

$$g(x) := \inf_{f\in\mathcal{F}} f(x), \quad x \in \mathbb{R}^m,$$

*the function $g$ is monotone with respect to scaling in the sense that for all $x \in \mathbb{R}^m$, the mapping*

$$\mathbb{R}_+ \ni c \mapsto g(cx) \in \mathbb{R} \tag{5.1}$$

*is non-decreasing.*

*Proof.* Fix $x \in \mathbb{R}^m$ and $c_1 < c_2 \in \mathbb{R}_+$. For any $f \in \mathcal{F}$, the symmetry about 0 and convexity of $f$ implies that $f$ has a global minimum at 0, hence

$$\begin{aligned} f(c_1 x) &= f(\tfrac{c_1}{c_2} \times c_2 x + \tfrac{c_2 - c_1}{c_2} \times 0) \\ &\le \tfrac{c_1}{c_2} f(c_2 x) + \tfrac{c_2-c_1}{c_2} f(0) \\ &= f(c_2 x) + \tfrac{c_2-c_1}{c_2}(f(0) - f(c_2 x)) \\ &\le f(c_2 x), \end{aligned}$$

where the first inequality follows from convexity, and the second inequality is due to the global minimum at 0. Taking the infimum over all $f \in \mathcal{F}$ on both sides, we find that $g(c_1 x) \le g(c_2 x)$, completing the proof. ☐

**Lemma 5.2.** *Fix $m \in \mathbb{N}$, and let $\mathsf{Y} = \{\mathsf{Y}_n\}_{n\in\mathbb{N}}$ denote a sequence of $\mathbb{R}^m$-valued random variables that satisfies an LDP with GRF $I_{\mathsf{Y}}$. Let $U$ be a uniformly distributed random variable on $[0,1]$ independent of $\{\mathsf{Y}_n\}_{n\in\mathbb{N}}$. If for all $y \in \mathbb{R}^m$, the mapping $\mathbb{R}_+ \ni c \mapsto I_{\mathsf{Y}}(cy) \in [0,\infty]$ is non-decreasing, then the scaled sequence $\{U^{1/n}\mathsf{Y}_n\}_{n\in\mathbb{N}}$ satisfies an LDP with GRF $I_{\mathsf{Y}}$.*

*Proof.* Due to [10, Lemma 3.3], the sequence $\{U^{1/n}\}_{n\in\mathbb{N}}$ satisfies an LDP with the good rate function

$$I_U(u) := \begin{cases} -\log u & u \in (0,1]; \\ +\infty & \text{else.} \end{cases}$$

By independence, the sequence $\{U^{1/n},\ \mathsf{Y}_n\}_{n\in\mathbb{N}}$ satisfies a joint LDP with the GRF $I_{U,\mathsf{Y}} : \mathbb{R}\times\mathbb{R}^m$ defined as $I_{U,\mathsf{Y}}(u,y) := I_U(u) + I_{\mathsf{Y}}(y)$. By the contraction principle (Theorem 1.3), the scaled sequence $\{U^{1/n}\mathsf{Y}_n\}_{n\in\mathbb{N}}$ satisfies an LDP with the rate function $I$, where for $x \in \mathbb{R}^m$,

$$I(x) := \inf_{u\in\mathbb{R},\, y\in\mathbb{R}^m} \{I_U(u) + I_{\mathsf{Y}}(y) : uy = x\} = \inf_{u\in(0,1]} \{-\log u + I_{\mathsf{Y}}(\tfrac{x}{u})\}.$$

The mapping $u \mapsto 1/u$ is monotonically decreasing, which when combined with the assumption on $I_{\mathsf{Y}}$ implies that $u \mapsto I_{\mathsf{Y}}(x/u)$ is monotonically decreasing. Since $u \mapsto -\log u$ is also monotonically decreasing, the infimum above is attained at $u = 1$, hence $I(x) = I_{\mathsf{Y}}(x)$ for all $x \in \mathbb{R}^m$. This concludes the proof of the lemma. □

*Proof of Theorem 2.4.* Suppose Assumption 2.1 holds for some $\{\xi_j\}_{j\in\mathbb{N}}$, $\mathbf{r}$, $\rho$, $q_\star > 0$, and $T \le \infty$, all as defined in the statement of the assumption. Due to the representation of $X^{(n)}$ given by Assumption 2.1(i), we have

$$n^{-1/2}\mathbf{a}_{n,k}^{\intercal} X^{(n)} \overset{(d)}{=} \rho\left((W_{\mathbf{a}}^{(n)})_2\right)(W_{\mathbf{a}}^{(n)})_1, \qquad n \in \mathbb{N}, \tag{5.2}$$

where $W_{\mathbf{a}}^{(n)}$ is the $\mathbb{R}^{k+1}$-valued random variable given by

$$W_{\mathbf{a}}^{(n)} = ((W_{\mathbf{a}}^{(n)})_1, (W_{\mathbf{a}}^{(n)})_2) := \left(n^{-1/2}\,\mathbf{a}_{n,k}^{\intercal}\xi^{(n)},\ \frac{1}{n}\sum_{i=1}^{n}\mathbf{r}(\xi_i)\right), \qquad n \in \mathbb{N}. \tag{5.3}$$

Thus we first prove an LDP for $\{W_{\mathbf{a}}^{(n)}\}_{n\in\mathbb{N}}$. In terms of the log mgf $\Lambda$ of $(\xi_1, \mathbf{r}(\xi_1))$, recalling our notation $\xi^{(n)} = (\xi_1, \dots, \xi_n)$, the scaled log mgf of $W_{\mathbf{a}}^{(n)}$ takes the following form: for $t_1 \in \mathbb{R}^k$ and $t_2 \in \mathbb{R}$,

$$\begin{aligned}
\frac{1}{n}\log\mathbb{E}\left[\exp(n\,\langle t, W_{\mathbf{a}}^{(n)}\rangle)\right] &= \frac{1}{n}\log\mathbb{E}\left[\exp\left(\sum_{i=1}^{n}\left(\sqrt{n}\,\xi_i\langle t_1, \mathbf{a}_{n,k}(i,\cdot)\rangle + t_2\,\mathbf{r}(\xi_i)\right)\right)\right] \\
&= \frac{1}{n}\log\prod_{i=1}^{n}\mathbb{E}\left[\exp\left(\sqrt{n}\,\xi_i\langle t_1, \mathbf{a}_{n,k}(i,\cdot)\rangle + t_2\,\mathbf{r}(\xi_i)\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\Lambda\left(\langle t_1, \sqrt{n}\mathbf{a}_{n,k}(i,\cdot)\rangle, t_2\right) \\
&= \Psi_{\mathsf{L}_{n,k}^{\mathbf{a}}}(t_1, t_2),
\end{aligned}$$

where $\Psi_{\mathsf{L}_{n,k}^{\mathbf{a}}}$ is equal to the integrated log mgf functional defined in (2.3), with $\nu = \mathsf{L}_{n,k}^{\mathbf{a}}$.

Fix $t_1 \in \mathbb{R}^k$. For $t_2 \ge T$, both sides are equal to $+\infty$ due to Remark 2.2. For $t_2 < T$, due to the $q_\star$-Wasserstein continuity of $\nu \to \Psi_\nu$ pointed out in Remark 2.2, together with the $q_\star$-Wasserstein convergence of $\mathsf{L}_{n,k}^{\mathbf{a}}$ to $\bar{\nu}$ in (2.6), we take the limit as $n \to \infty$ of both sides of the last display to find

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\left[\exp(n\,\langle t, W_{\mathbf{a}}^{(n)}\rangle)\right] = \lim_{n\to\infty}\Psi_{\mathsf{L}_{n,k}^{\mathbf{a}}}(t_1, t_2) = \Psi_{\bar{\nu}}(t_1, t_2).$$

Due to the lower semicontinuity and essential smoothness of $\Psi_{\bar{\nu}}$ on $\mathbb{R}^{k+1}$, which follow from Assumption 2.1(iii), the Gärtner-Ellis theorem (see, e.g., [7, Theorem 2.3.6]) yields the LDP for the sequence $\{W_{\mathbf{a}}^{(n)}\}_{n\in\mathbb{N}}$ in $\mathbb{R}^{k+1}$ with the GRF $\Psi_{\bar{\nu}}^*$ from (2.5). Also, note that since $(W_{\mathbf{a}}^{(n)})_2$ is supported on $\mathbb{R}_+$, $\Psi_{\bar{\nu}}^*(\tau_1, \tau_2) = \infty$ whenever $\tau_2 < 0$.

The LDP for $\{W_{\mathbf{a}}^{(n)}\}_{n\in\mathbb{N}}$ and the contraction principle (Theorem 1.3) applied to the continuous mapping $\mathbb{R}^k \times \mathbb{R}_+ \ni (\tau_1, \tau_2) \mapsto \rho(\tau_2)\tau_1 \in \mathbb{R}^k$ yield an LDP for $\{n^{-1/2}\mathbf{a}_{n,k}^{\intercal}X^{(n)}\}_{n\in\mathbb{N}_k}$ in $\mathbb{R}^k$ with GRF $\bar{\mathcal{J}}_{\bar{\nu}}^{\mathsf{qu}}$ defined to be

$$\bar{\mathcal{J}}_{\bar{\nu}}^{\mathsf{qu}}(x) := \inf_{\tau_1\in\mathbb{R}^k,\,\tau_2\in\mathbb{R}_+}\left\{\Psi_{\bar{\nu}}^*(\tau_1, \tau_2) : \tau_1\rho(\tau_2) = x\right\}, \quad x \in \mathbb{R}^k.$$

Substituting the constraint $\tau_1\rho(\tau_2) = x$ and using (2.7), we see that

$$\bar{\mathcal{J}}_{\bar{\nu}}^{\mathsf{qu}}(x) = \inf_{\tau\in\mathbb{R}_+}\Psi_{\bar{\nu}}^*\left(\frac{x}{\rho(\tau)}, \tau\right) = \mathcal{J}_{\bar{\nu}}^{\mathsf{qu}}(x), \quad x \in \mathbb{R}^k. \tag{5.4}$$

This proves part (i) of the theorem.

In turn, the LDP from part (i) implies part (ii) of the theorem since by (2.12) and Corollary 2.11, almost surely, $\mathcal{W}_{q_\star}(\mathsf{L}^{\mathbf{A}}_{n,k},\gamma^{\otimes k}) = \mathcal{W}_{q_\star}(\mathsf{L}_{n,k},\gamma^{\otimes k}) \to 0$ as $n\to\infty$.

We turn to the final claim (iii). Given the assumption on symmetry of $\Lambda$, it is apparent from the definition (2.3) that $\Psi_{\bar\nu}$ is symmetric in its first argument, and then from the definition of the Legendre transform (2.5) that $\Psi^*_{\bar\nu}$ is also symmetric in its first argument. Applying Lemma 5.1 with dimension $m=k$, the set of symmetric convex functions $\mathcal{F} = \{\mathbb{R}^k \ni x \mapsto \Psi^*_{\bar\nu}(\frac{x}{\rho(\tau)},\tau)\in\mathbb{R}\}_{\tau\in\mathbb{R}_+}$, and $g=\mathcal{J}^{\mathsf{qu}}_{\bar\nu}$, we find that the mapping $\mathbb{R}_+\ni c\mapsto \mathcal{J}^{\mathsf{qu}}_{\bar\nu}(cx)\in[0,\infty]$ is non-decreasing. An application of Lemma 5.2 with $\mathsf{Y}_n = n^{-1/2}\mathbf{a}^{\intercal}_{n,k}X^{(n)}$, $n\in\mathbb{N}$, and $I_{\mathsf{Y}}=\mathcal{J}^{\mathsf{qu}}_{\bar\nu}$, completes the proof. ☐

## 6 Proof of the variational formula

In this section, we prove Theorem 2.7, primarily through an application of Theorem 2.8 and Sion's minimax theorem [31]. We start with preliminary results in Lemma 6.1, Lemma 6.2 and Lemma 6.3. Throughout, recall the definition of $\mathbb{H}_k$ from (2.10).

**Lemma 6.1.** *Suppose Assumption 2.1 holds, with associated quantities $T$ and $\Psi_\nu$, $\nu\in\mathcal{P}(\mathbb{R}^k)$, and recall the empirical measure $\mathsf{L}_{n,k}$ from (2.12). For $t_1\in\mathbb{R}^k$, $t_2<T$, and $0<\delta<\infty$, the following condition holds:*

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}\left[e^{\delta n\Psi_{\mathsf{L}_{n,k}}(t_1,t_2)}\right]<\infty\,. \tag{6.1}$$

*Proof.* Let $\Theta^{(n)} := (\Theta^{(n)}_1,\dots,\Theta^{(n)}_n)$ denote a random vector distributed uniformly on $\mathbb{S}^{n-1}$, the Euclidean sphere in $\mathbb{R}^n$ of radius 1. For $t_1\in\mathbb{R}^k$, the random vector $\mathbf{A}_{n,k}t_1$ lies on the Euclidean sphere in $\mathbb{R}^n$ of radius $\|t_1\|_2$ and has a law invariant to orthogonal transformation (due to the law of $\mathbf{A}_{n,k}$ being invariant under orthogonal transformations); hence, $\mathbf{A}_{n,k}t_1 \stackrel{(d)}{=} \|t_1\|_2\Theta^{(n)}$. Fix $t_1\in\mathbb{R}^k$ and $t_2<T$, let $C_{t_2}$ and $q_\star$ be as in Assumption 2.1(ii), and define $g:\mathbb{R}_+\to\mathbb{R}_+$ as

$$g(x) := \exp\left(\delta C_{t_2}[1+(\|t_1\|_2 x)^{q_\star}]\right),\quad x\in\mathbb{R}_+.$$

When combined, the relation $\mathbf{A}_{n,k}t_1 \stackrel{(d)}{=} \|t_1\|_2\Theta^{(n)}$, the bound of Assumption 2.1(ii), the fact that each $g_i$ is an increasing function and the sub-independence of $(|\Theta^{(n)}_1|,\dots,|\Theta^{(n)}_n|)$ established in [3, Theorem 2.11(2)] with $p=2$ therein, yield

$$\begin{aligned}\mathbb{E}\left[e^{\delta n\Psi_{\mathsf{L}_{n,k}}(t_1,t_2)}\right] &= \mathbb{E}\left[\prod_{i=1}^n \exp\left(\delta\Lambda(\sqrt{n}\|t_1\|_2\Theta^{(n)}_i,t_2)\right)\right] \le \mathbb{E}\left[\prod_{i=1}^n g(|\sqrt{n}\Theta^{(n)}_i|)\right]\\ &\le \prod_{i=1}^n \mathbb{E}\left[g(\sqrt{n}|\Theta^{(n)}_i|)\right].\end{aligned} \tag{6.2}$$

Now, let $\{Z_i\}_{i\in\mathbb{N}}$ be i.i.d. standard Gaussian variables, as usual set $Z^{(n)} := (Z_1,\dots,Z_n)$, and note that for each $i=1,\dots,n$, $\sqrt{n}\Theta^{(n)}_i \stackrel{(d)}{=} \sqrt{n}Z_1/\|Z^{(n)}\|_2$ and further, $\sqrt{n}Z_1/\|Z^{(n)}\|_2 \xrightarrow{\text{a.s.}} Z_1$ as $n\to\infty$, and hence almost surely, $\sqrt{n}|Z_1|/\|Z^{(n)}\|_2 \le 2|Z_1|$ for all sufficiently large $n$. Therefore, first dividing both sides of (6.2) by $n$, then taking the limit superior, as $n\to\infty$, and applying the reverse Fatou lemma, which is applicable since $\mathbb{E}[\exp(\delta C_{t_2}\|t_1\|_2^{q_\star}|2Z_1|^{q_\star})]<\infty$ because $q_\star<2$, we obtain

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}\left[e^{\delta n\Psi_{\mathsf{L}_{n,k}}(t_1,t_2)}\right] &\le \limsup_{n\to\infty}\log\mathbb{E}\left[\exp\left(\delta C_{t_2}[1+(\|t_1\|_2\tfrac{\sqrt{n}|Z_1|}{\|Z^{(n)}\|_2})^{q_\star}]\right)\right]\\ &\le \delta C_{t_2} + \log\mathbb{E}\left[\exp(\delta C_{t_2}\|t_1\|_2^{q_\star}|Z_1|^{q_\star})\right].\end{aligned}$$

Since the last term on the right-hand side is finite for all $t_1 \in \mathbb{R}^k$, once again because $q_\star < 2$, (6.1) follows. □

**Lemma 6.2.** *Suppose Assumption 2.1 holds, with associated quantities $\{\xi_j\}_{j\in\mathbb{N}}$, $\mathbf{r}$, $T$, and $\Psi_\nu$, $\nu \in \mathcal{P}(\mathbb{R}^k)$, and let $\mathbf{A}_{n,k}$ be drawn from the Haar measure $\sigma_{n,k}$ on $\mathbb{V}_{n,k}$, independently of $\{\xi_j\}_{j\in\mathbb{N}}$. For $n \in \mathbb{N}$, define*

$$\Phi_n(t_1,t_2) := \frac{1}{n}\log\mathbb{E}\left[\exp\left(\sqrt{n}\,\xi^{(n)}\mathbf{A}_{n,k}t_1 + t_2\sum_{i=1}^n \mathbf{r}(\xi_i)\right)\right], \quad t_1 \in \mathbb{R}^k,\, t_2 \in \mathbb{R}, \tag{6.3}$$

*where $\xi^{(n)} := (\xi_1, \xi_2, \dots, \xi_n)$. Then for $t_1 \in \mathbb{R}^k$ and $t_2 \in \mathbb{R}$,*

$$\Phi_n(t_1,t_2) = \frac{1}{n}\log\mathbb{E}\left[\exp\left(n\Psi_{L_{n,k}(t_1,t_2)}\right)\right], \tag{6.4}$$

*where $\mathsf{L}_{n,k}$ and $\Psi_\nu$ are as defined in (2.12) and (2.3), respectively. Moreover,*

$$\lim_{n\to\infty}\Phi_n(t_1,t_2) = \Phi(t_1,t_2),$$

*where, with $\mathcal{K}$ equal to the set defined in (4.8), we have*

$$\Phi(t_1,t_2) := \sup_{\nu\in\mathcal{P}(\mathbb{R}^k)}\left\{\Psi_\nu(t_1,t_2) - \mathbb{H}_k(\nu)\right\} = \sup_{\nu\in\mathcal{K}}\left\{\Psi_\nu(t_1,t_2) - \mathbb{H}_k(\nu)\right\}. \tag{6.5}$$

*Proof.* Due to the independence of $\xi_1, \xi_2, \dots$, and their independence from $\mathbf{A}_{n,k}$, we can write, for $n \in \mathbb{N}$, $t_1 \in \mathbb{R}^k$ and $t_2 \in \mathbb{R}$,

$$\begin{aligned}\Phi_n(t_1,t_2) &= \frac{1}{n}\log\mathbb{E}\left[\prod_{i=1}^n \mathbb{E}\left[\exp\left(\sqrt{n}\,\xi_i\left(\mathbf{A}_{n,k}t_1\right)_i + t_2\mathbf{r}(\xi_i)\right)\big|\mathbf{A}_{n,k}\right]\right] \\ &= \frac{1}{n}\log\mathbb{E}\left[\exp\left(\sum_{i=1}^n \Lambda(\sqrt{n}\,\langle\mathbf{A}_{n,k}(i,\cdot),t_1\rangle,\, t_2\right)\right],\end{aligned}$$

where $\Lambda$ is as in (2.1). Then (6.4) follows immediately from the definitions of $\mathsf{L}_{n,k}$ and $\Psi_\nu$ in (2.12) and (2.3), respectively. where $\Lambda$ is as in (2.1) Now, let $T \le \infty$ and $q_\star \in (0,2)$ be as specified in (ii) and (v) of Assumption 2.1. For $t_2 \ge T$, by Remark 2.2, both $\Phi_n(\cdot,t_2)$ and $\Phi(\cdot,t_2)$ are identically equal to infinity, and so the limit $\Phi_n(t_1,t_2) \to \Phi(t_1,t_2)$ holds trivially for all $t_1 \in \mathbb{R}^k$. On the other hand, suppose $t_2 < T$. Then, again from Remark 2.2 Assumptions 2.1(ii)–(iii) imply that the map $\mathcal{P}_{q^*}(\mathbb{R}^k) \ni \nu \mapsto \Psi_\nu(t_1,t_2) \in \mathbb{R}$ is continuous (with respect to the $q_\star$-Wasserstein topology). Moreover, from Theorem 2.8 that the sequence $\{\mathsf{L}_{n,k}\}_{n\in\mathbb{N}_k}$ satisfies an LDP in $\mathcal{P}_q(\mathbb{R}^k)$ for all $q \in (0,2)$, with the GRF $\mathbb{H}_k$. Since $q_* \in (0,2)$ by Assumption 2.1(v), by Varadhan's lemma [7, Theorem 4.3.1], which is applicable due to the integrability estimate (6.1) of Lemma 6.1, it follows that the limit of $\Phi_n(t_1,t_2)$ is given by $\Phi(t_1,t_2)$ defined in (6.5).

To complete the proof of the lemma, it only remains to establish the last equality in (6.5), but this is an immediate consequence of the definitions of $\mathcal{K}$ and $\mathbb{H}_k$ in (4.8) and (2.10), respectively, which directly imply $\mathbb{H}_k(\nu) = \infty$ for $\nu \notin \mathcal{K}$. □

**Lemma 6.3.** *Suppose Assumption 2.1 holds, and for each $\nu \in \mathcal{P}(\mathbb{R}^k)$, let $\Psi_\nu$ be as defined in (2.3), let $\Psi^*_\nu$ denote its Legendre transform, as specified in (2.5), and let $\mathcal{K} \subset \mathcal{P}(\mathbb{R}^k)$ be the set defined in (4.8). Then the Legendre transform $\Phi^*$ of the function $\Phi$ defined in (6.5) satisfies, for $\tau_1 \in \mathbb{R}^k$ and $\tau_2 \in \mathbb{R}$,*

$$\Phi^*(\tau_1,\tau_2) = \inf_{\nu\in\mathcal{K}}\left\{\Psi^*_\nu(\tau_1,\tau_2) + \mathbb{H}_k(\nu)\right\} = \inf_{\nu\in\mathcal{P}(\mathbb{R}^k)}\left\{\Psi^*_\nu(\tau_1,\tau_2) + \mathbb{H}_k(\nu)\right\}. \tag{6.6}$$

*Proof.* First, note that the second equality in (6.6) holds because $\mathbb{H}_k(\nu) = \infty$ for $\nu \notin \mathcal{K}$ due to (4.8) and (2.10). Next, fix the following:

- let $\Lambda$, $T$ be as in Assumption 2.1(ii), and define $\mathcal{D}_T := \mathbb{R}^k \times (-\infty, T)$;
- let $q_\star \in (0, 2)$ be as in Assumption 2.1(ii), and let $\mathcal{M}_{q_\star}(\mathbb{R}^k)$ denote the space of finite signed measures (not necessarily probability measures) on $\mathbb{R}^k$, equipped with the $q_\star$-Wasserstein topology.

Fix $\tau = (\tau_1, \tau_2) \in \mathbb{R}^k \times \mathbb{R}$. Then by the definition (2.5) of $\Psi_\nu^*$,

$$\begin{aligned}\Psi_\nu^*(\tau_1, \tau_2) &= \sup_{(t_1,t_2)\in\mathbb{R}^k\times\mathbb{R}} \{\langle \tau_1, t_1\rangle + \tau_2 t_2 - \Psi_\nu(t_1, t_2)\} \\ &= \sup_{(t_1,t_2)\in\mathcal{D}_T} \{\langle \tau_1, t_1\rangle + \tau_2 t_2 - \Psi_\nu(t_1, t_2)\}, \end{aligned} \tag{6.7}$$

where the second equality holds because, by Remark 2.2, $\Psi_\nu(t_1, t_2) = \infty$ if $t_2 \geq T$. Thus, the right-hand side of (6.6) is equal to $\inf_{\nu\in\mathcal{K}} \sup_{t=(t_1,t_2)\in\mathcal{D}_T} F_\tau(\nu, t)$, where

$$F_\tau(\nu, t) := \langle \tau_1, t_1\rangle + \tau_2 t_2 - \Psi_\nu(t_1, t_2) + \mathbb{H}_k(\nu), \quad \nu \in \mathcal{P}(\mathbb{R}^k),\, t = (t_1, t_2) \in \mathbb{R}^k \times \mathbb{R}.$$

On the other hand, by the definition of $\Phi^*$ and the representation (6.5) for $\Phi$,

$$\begin{aligned}\Phi^*(\tau_1, \tau_2) &= \sup_{(t_1,t_2)\in\mathbb{R}^{k+1}} \{\langle \tau_1, t_1\rangle + \tau_2 t_2 - \Phi(t_1, t_2)\} \\ &= \sup_{t=(t_1,t_2)\in\mathbb{R}^{k+1}} \inf_{\nu\in\mathcal{K}} F_\tau(\nu, t), \\ &= \sup_{t=(t_1,t_2)\in\mathcal{D}_T} \inf_{\nu\in\mathcal{K}} F_\tau(\nu, t),\end{aligned}$$

where the last equality uses the fact that for $t_2 > T$, $\Psi_\nu(t_1, t_2) = \infty$ and hence, $F_\tau(\nu, t) = -\infty$ (see Remark 2.2). Thus, to prove the first equality in (6.6), it suffices to show that for all $(\tau_1, \tau_2) \in \mathbb{R}^k \times \mathbb{R}$,

$$\inf_{\nu\in\mathcal{K}} \sup_{(t_1,t_2)\in\mathcal{D}_T} F_\tau(\nu, (t_1, t_2)) = \sup_{(t_1,t_2)\in\mathcal{D}_T} \inf_{\nu\in\mathcal{K}} F_\tau(\nu, (t_1, t_2)). \tag{6.8}$$

To justify the exchange of infimum and supremum in (6.8), we verify the conditions of Sion's minimax theorem [31, Corollary 3.3]. That is, for $(\tau_1, \tau_2) \in \mathbb{R}^k \times \mathbb{R}$, we note that

- the set $\mathcal{D}_T = \mathbb{R}^k \times (-\infty, T)$ is a convex subset of the topological vector space $\mathbb{R}^{k+1}$;
- due to Lemma 4.4 and the fact that $q_* \in (0, 2)$, $\mathcal{K}$ is a convex compact subset of the topological vector space $\mathcal{M}_{q_\star}(\mathbb{R}^k)$;
- for $t = (t_1, t_2) \in \mathcal{D}_T$: the lower semicontinuity of $F_\tau(\cdot, t)$ follows from the lower semicontinuity of $\nu \to \Psi_\nu(t)$ due to Assumption 2.1(iii) and of $\mathbb{H}_k$ (as it is a GRF); the convexity of $F_\tau(\cdot, t)$ follows from the linearity of $\nu \mapsto \Psi_\nu(t)$ and the convexity of $\mathbb{H}_k$, which was established in Theorem 2.8;
- for $\nu \in \mathcal{K}$: the lower semicontinuity of $t \to \Psi_\nu(t)$ on $\mathcal{D}_T$ follows from Assumption 2.1(iii); the convexity of $\Psi_\nu$ on $\mathcal{D}_T$ follows from linearity of expectation, the definition (2.3), and the fact that $\Lambda$ is convex since it is a log mgf by (2.1);
- the convexity of $\Lambda$ also implies that for each $\nu \in \mathcal{P}(\mathbb{R}^k)$ $(t_1, t_2) \mapsto Psi_\nu(t_1, t_2)$ is convex and it is also lower semicontinuous by Assumption 2.1(iii); together with the fact that for each $\tau = (\tau_1, \tau_2) \in \mathbb{R}^k \times \mathbb{R}$, $(t_1, t_2) \mapsto \langle \tau_1, t_1\rangle + \tau_2 t_2$ is continuous and linear, it follows that $F_\tau(\nu, \cdot)$ is upper semicontinuous and concave on $\mathcal{D}_T$.

Due to the conditions verified above, the minimax theorem can be applied to establish (6.8) and hence, the first equality in (6.6). This completes the proof of the lemma. □

*Proof of Theorem 2.7.* Let $\{\xi_j\}_{j\in\mathbb{N}}$, $\rho$ and $\mathbf{r}$ be as in Assumption 2.1, let $\xi^{(n)} := (\xi_1, \dots, \xi_n)$ and consider the sequence $\{W^{(n)}\}_{n\in\mathbb{N}}$ in $\mathbb{R}^{k+1}$, where $W^{(n)} = W^{(n)}_{\mathbf{A}}$ is defined by

$$W^{(n)} = (W^{(n)}_1, W^{(n)}_2) := \left(n^{-1/2}\, \mathbf{A}^{\intercal}_{n,k}\xi^{(n)},\ \frac{1}{n}\sum_{i=1}^{n}\mathbf{r}(\xi_i)\right), \qquad n \in \mathbb{N}. \tag{6.9}$$

Then by Assumption 2.1(i) we can write

$$\sqrt{n}\mathbf{A}^{\intercal}_{n,k}X^{(n)} = W^{(n)}_1\rho(W^{(n)}_2), \tag{6.10}$$

where $\rho$ is continuous. Analogous to the proof of Theorem 2.4 we will start by first establishing an LDP for $\{W^{(n)}\}_{n\in\mathbb{N}}$.

Note that the functional $\Phi_n$ defined in (6.3) is the scaled mgf of $W^{(n)}$:

$$\Phi_n(t_1, t_2) = \frac{1}{n}\log\mathbb{E}\left[\exp\left(n\langle(t_1, t_2), W^{(n)}\rangle\right)\right], \quad (t_1, t_2) \in \mathbb{R}^{k+1}, \tag{6.11}$$

and hence, by Lemma 6.2 it follows that for every $(t_1, t_2) \in \mathbb{R}^k \times \mathbb{R}$,

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\left[e^{n\langle(t_1,t_2),W^{(n)}\rangle}\right] = \Phi(t_1, t_2), \tag{6.12}$$

where $\Phi$ is as in (6.5). Also, note that the law of $\mathbf{A}_{n,k}$ is invariant to orthogonal transformation and independent of $\xi^{(n)}$, hence $\mathbf{A}^{\intercal}_{n,k}\frac{\xi^{(n)}}{\|\xi^{(n)}\|_2} \overset{(d)}{=} \mathbf{A}_{n,k}(1,\cdot)$ and $\mathbf{A}^{\intercal}_{n,k}\frac{\xi^{(n)}}{\|\xi^{(n)}\|_2}$ is independent of $\xi^{(n)}$; we refer to [10, Lemma 6.3] for the proof of the simpler case when $k = 1$. As a consequence,

$$W^{(n)} \overset{(d)}{=} \left(n^{-1/2}\mathbf{A}_{n,k}(1,\cdot)\|\xi^{(n)}\|_2, \frac{1}{n}\sum_{i=1}^{n}\mathbf{r}(\xi_i)\right).$$

Define the $\mathbb{R}^{k+2}$-valued sequence of random variables,

$$S^{(n)} := \left(\mathbf{A}_{n,k}(1,\cdot), \frac{1}{n}\|\xi^{(n)}\|_2^2, \frac{1}{n}\sum_{i=1}^{n}\mathbf{r}(\xi_i)\right), \quad n \in \mathbb{N}.$$

Since by part (iv) of Assumption 2.1, the domain of $\bar{\Lambda}$, the log mgf of $(\xi_1^2, \mathbf{r}(\xi_1))$, contains a neighborhood of the origin, by Cramér's theorem (Theorem 1.7) $\{(\|\xi^{(n)}\|_2^2, \frac{1}{n}\sum_{i=1}^{n}\mathbf{r}(\xi_i))\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}^2$ with the convex GRF $\bar{\Lambda}^*$, equal to the Legendre transform of $\bar{\Lambda}$. The independence of $\mathbf{A}_{n,k}$ from $\{\xi_j\}_{j\in\mathbb{N}}$, along with [3, Theorem 3.4] (applied to the case of $p = 2$ therein, with their canonically projected $X^{(k)}$ equivalent to our $\mathbf{A}_{n,k}(1,\cdot)$) then implies that the sequence $\{S^{(n)}\}_{n\in\mathbb{N}}$ satisfies an LDP with the convex GRF $J : \mathbb{R}^{k+2} \to [0, \infty]$ defined by

$$J(a, b, c) := -\tfrac{1}{2}\log(1 - \|a\|_2^2) + \widehat{J}(b, c),$$

for $a \in \mathbb{R}^k$ such that $\|a\|_2^2 < 1$ and $b, c \in \mathbb{R}$, for some function $\widehat{J} : \mathbb{R}^2 \to [0, \infty]$. Then, by the contraction principle (Theorem 1.3), $\{W^{(n)}\}_{n\in\mathbb{N}}$ satisfies an LDP with the GRF $J_W : \mathbb{R}^{k+1} \to \mathbb{R}$ defined as follows:

$$J_W(x, z) := \inf_{y\in\mathbb{R}:y>\|x\|^2} J(xy^{-1/2}, y, z), \quad x \in \mathbb{R}^k, z \geq 0.$$

Note that $J_W$ is convex due to [10, Lemma 6.2] and [29, Theorem 5.3].

We now claim that

$$\Phi(t_1,t_2) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}[e^{n\langle t, W^{(n)}\rangle}] = \sup_{\tau_1\in\mathbb{R}^k, \tau_2\in\mathbb{R}} \{\langle t_1,\tau_1\rangle + t_2\tau_2 - J_W(\tau_1,\tau_2)\}. \tag{6.13}$$

The first equality is just (6.12). To justify the second equality in (6.13), let $T$ be as in Assumption 2.1(ii) and first fix $t_1 \in \mathbb{R}^k$, $t_2 < T$, and let $t = (t_1,t_2)$. For $0 < \delta < \infty$, (6.11) and (6.4) together imply

$$\mathbb{E}[e^{\delta n\langle t, W^{(n)}\rangle}] = e^{n\Phi_n(\delta t_1,\delta t_2)} = \mathbb{E}\left[e^{n\Psi_{\mathsf{L}_{n,k}}(\delta t_1,\delta t_2)}\right].$$

Since $t_2 < T$, there exists $\delta > 1$ such that $\delta t_2 < T$ and so the last relation and Lemma 6.1 imply

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}[e^{\delta n\langle t, W^{(n)}\rangle}] < \infty \quad \text{for some } \delta > 1. \tag{6.14}$$

Hence, (6.12), the fact that $\{W^{(n)}\}_{n\in\mathbb{N}}$ satisfies an LDP with rate function $J_W$ and Varadhan's lemma [7, Theorem 4.3.1], whose application is justified by (6.14), imply (6.13) holds for all $t_1 \in \mathbb{R}^k$ and $t_2 < T$.

Now fix $t_1 \in \mathbb{R}^k$ and $t_2 \geq T$. We claim that (6.13) continues to hold, but now with both sides equal to infinity. The fact that $\Phi(t_1,t_2) = \infty$ follows from the definition (6.5) of $\Phi$ and the observation that $\Phi_\nu(t_1,t_2) = \infty$ for every $\nu \in \mathcal{P}(\mathbb{R}^k)$ when $t_2 \geq T$ (by Remark 2.2). To show that the term on the right-hand side of (6.13) is also equal to infinity, for $s_2 \in \mathbb{R}$, define $\widetilde{\Lambda}(s_2) := \Lambda(0,s_2)$. Note that $\widetilde{\Lambda}$ is the log mgf of $\mathbf{r}(\xi_1)$. Due to Assumption 2.1(iv), the domain of $\widetilde{\Lambda}$ contains a non-empty neighborhood around 0, hence by Cramér's theorem (Theorem 1.7) the sequence $\{\frac{1}{n}\sum_{i=1}^n \mathbf{r}(\xi_i)\}_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ with GRF $\widetilde{\Lambda}^*$. However, due to the contraction principle (Theorem 1.3) and the continuity of the coordinate projection map, we also know that $\widetilde{\Lambda}^*(\tau_2) = \inf_{\tau_1\in\mathbb{R}^k} J_W(\tau_1,\tau_2)$ for all $\tau_2 \in \mathbb{R}$. Note that this infimum is attained at some $\tau_1^* \in \mathbb{R}^k$ because, as a GRF, $J_W$ is lower semicontinuous with compact level sets. Therefore, on the right-hand side of (6.13), if $t_2 \geq T$, then

$$\begin{aligned}
\sup_{\tau_1\in\mathbb{R}^k,\tau_2\in\mathbb{R}} \{\langle t_1,\tau_1\rangle + t_2\tau_2 - J_W(\tau_1,\tau_2)\} &\geq \sup_{\tau_2\in\mathbb{R}} \{\langle t_1,\tau_1^*\rangle + t_2\tau_2 - J_W(\tau_1^*,\tau_2)\} \\
&= \langle t_1,\tau_1^*\rangle + \sup_{\tau_2\in\mathbb{R}} \{t_2\tau_2 - \widetilde{\Lambda}^*(\tau_2)\} \\
&= \langle t_1,\tau_1^*\rangle + \widetilde{\Lambda}(t_2) \\
&= \langle t_1,\tau_1^*\rangle + \Lambda(0,t_2) \\
&= \infty,
\end{aligned}$$

where the first equality used the definition of $\tau_1^*$, the second equality used the identity $(\widetilde{\Lambda}^*)^* = \widetilde{\Lambda}$, which holds since $\widetilde{\Lambda}$ is convex, the third equality uses the definition of $\widetilde{\Lambda}$, and the last equality follows from Remark 2.2. Hence, (6.13) holds for *all* $t_1 \in \mathbb{R}^k$ and $t_2 \in \mathbb{R}$.

Note that (6.13) shows that $\Phi = J_W^*$. Due to the convexity of $J_W$ and Legendre duality (see, e.g., [7, Lemma 4.5.8]), we also have $J_W = \Phi^*$. Moreover, since $W_2^{(n)}$ is supported on $\mathbb{R}_+$, clearly $\Phi(\tau_1,\tau_2) = \infty$ whenever $\tau_2 < 0$. By the representation for $\sqrt{n}\mathbf{A}_{n,k}^{\mathsf{T}} X^{(n)}$ in terms of $W^{(n)}$ in (6.10), invoking the LDP for $\{W^{(n)}\}_{n\in\mathbb{N}}$ and applying the contraction principle (Theorem 1.3) to the continuous map $\mathbb{R}^k \times \mathbb{R}_+ \ni (w_1,w_2) \mapsto w_1\rho(w_2) \in \mathbb{R}^k$ (recall that $\rho$ is continuous), we find that the annealed rate function $\mathcal{J}^{\mathsf{an}}$ for $\sqrt{n}\mathbf{A}_{n,k}^{\mathsf{T}} X^{(n)}$

in Theorem 2.6 can be written as

$$\begin{aligned}\mathcal{J}^{\mathsf{an}}(x) &= \inf_{\tau_1\in\mathbb{R}^k,\tau_2\in\mathbb{R}_+:\tau_1\rho(\tau_2)=x} \Phi^*(\tau_1,\tau_2)\\ &= \inf_{\tau_1\in\mathbb{R}^k,\tau_2\in\mathbb{R}_+:\tau_1\rho(\tau_2)=x}\ \inf_{\nu\in\mathcal{P}(\mathbb{R}^k)} \{\Psi^*_\nu(\tau_1,\tau_2)+\mathbb{H}_k(\nu)\}\\ &= \inf_{\nu\in\mathcal{P}(\mathbb{R}^k)}\ \inf_{\tau_1\in\mathbb{R}^k,\tau_2\in\mathbb{R}_+:\tau_1\rho(\tau_2)=x} \{\Psi^*_\nu(\tau_1,\tau_2)+\mathbb{H}_k(\nu)\}\\ &= \inf_{\nu\in\mathcal{P}(\mathbb{R}^k)} \left\{\inf_{\tau\in\mathbb{R}_+} \Psi^*_\nu\left(\frac{x}{\rho(\tau)},\tau\right)+\mathbb{H}_k(\nu)\right\}\\ &= \inf_{\nu\in\mathcal{P}(\mathbb{R}^k)} \{\mathcal{J}^{\mathsf{qu}}_\nu(x)+\mathbb{H}_k(\nu)\},\end{aligned}$$

where the second equality invokes (6.6) of Lemma 6.3, and the last equality relies on (2.7). This completes the proof of Theorem 2.7. □

## References


[1] E. Aidékon, *Large deviations for transient random walks in random environment on a Galton–Watson tree*, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques **46** (2010), no. 1, 159–189. MR2641775

[2] D. Alonso-Gutiérrez, J. Prochno, and C. Thäle, *Large deviations for high-dimensional random projections of $\ell_p^n$-balls*, Advances in Applied Mathematics **99** (2018), no. 12, 1–35. MR3806754

[3] F. Barthe, F. Gamboa, L. Lozada-Chang, and A. Rouault, *Generalized Dirichlet distributions on the ball and moments*, ALEA Latin American Journal of Probability and Mathematical Statistics **7** (2010), 319–340. MR2737561

[4] M. S. Bartlett, *On the theory of statistical regression*, Proceedings of the Royal Society of Edinburgh **53** (1933), 260–283.

[5] G. Ben Arous, A. Dembo, and A. Guionnet, *Aging of spherical spin glasses*, Probability Theory and Related Fields **120** (2001), no. 1, 1–67. MR1856194

[6] F. Comets, N. Gantert, and O. Zeitouni, *Quenched, annealed and functional large deviations for one-dimensional random walk in random environment*, Probability Theory and Related Fields **118** (2000), no. 1, 65–114. MR1785454

[7] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, 2 ed., vol. 38, Springer Science & Business Media, 2009. MR2571413

[8] P. Diaconis and D. Freedman, *Asymptotics of graphical projection pursuit*, The Annals of Statistics **12** (1984), no. 3, 793–815. MR0751274

[9] N. Gantert, S. S. Kim, and K. Ramanan, *Cramér's theorem is atypical*, Advances in the Mathematical Sciences: Research from the 2015 Association for Women in Mathematics Symposium (Gail Letzter, Kristin Lauter, Erin Chambers, Nancy Flournoy, Julia Elisenda Grigsby, Carla Martin, Kathleen Ryan, and Konstantina Trivisa, eds.), Springer International Publishing, Cham, 2016, pp. 253–270. MR3654489

[10] N. Gantert, S. S. Kim, and K. Ramanan, *Large deviations for random projections of $\ell^p$ balls*, The Annals of Probability **45** (2017), no. 6B, 4419–4476. MR3737915

[11] A. Guionnet and M. Maïda, *A Fourier view on the R-transform and related asymptotics of spherical integrals*, Journal of Functional Analysis **222** (2005), no. 2, 435–490. MR2132396

[12] P. D. Hoff, *Simulation of the matrix Bingham-von Mises-Fisher distribution, with applications to multivariate and relational data*, Journal of Computational and Graphical Statistics **18** (2009), no. 2, 438–456. MR2749840

[13] W. B. Johnson and J. Lindenstrauss, *Extensions of Lipschitz mappings into a Hilbert space*, Conference in modern analysis and probability (New Haven, Conn., 1982), Contemp. Math., vol. 26, Amer. Math. Soc., Providence, RI, 1984, pp. 189–206. MR737400

[14] Z. Kabluchko and J. Prochno, *Large deviations for random matrices in the orthogonal group and Stiefel manifold with applications to random projections of product distributions*, arXiv e-print, arXiv:2110.12977, 2021. MR4310599

[15] Z. Kabluchko, J. Prochno, and C. Thaele, *High-dimensional limit theorems for random vectors in $\ell_p^n$-balls*, Communications in Contemporary Mathematics **21** (2019), no. 1, 1–30. MR3904638

[16] Z. Kabluchko, J. Prochno, and C. Thaele, *A new look at random projections of the cube and general product measures*, Bernoulli **27** (2021), no. 3, 2117-2138. MR4260512

[17] T. Kaufmann, *Sharp Asymptotics for $q$-Norms of Random Vectors in High-Dimensional $\ell_p^n$-Balls*, Modern Stochastics: Theory and Applications, **8** (2021), no. 2, 239–274. MR4279877

[18] S. S. Kim, *Problems at the interface of probability and convex geometry: Random projections and constrained processes*, Ph.D. thesis, Brown University, 2017.

[19] S. S. Kim and K. Ramanan, *A conditional limit theorem for high-dimensional $\ell^p$-spheres*, Journal of Applied Probability **55** (2018), no. 4, 1060–1077. MR3899928

[20] S. S. Kim, Y.-T. Liao, and K. Ramanan, *An asymptotic thin shell condition and large deviations for random multidimensional projections*, Advances in Applied Mathematics, **134** (2022), 102306. MR4352552

[21] B. Klartag, *A central limit theorem for convex sets*, Inventiones mathematicae **168** (2007), 91–131. MR2285748

[22] Y.-T. Liao and K. Ramanan, *Geometric sharp large deviations for random projections of $\ell_p^n$ balls and spheres*, *Electronic Journal of Probability*, to appear, 2023; arXiv e-print, arXiv:2001.04053, 2020.

[23] Y.-T. Liao and K. Ramanan, *A refined asymptotic thin shell condition and its applications*, Preprint, 2023.

[24] P. Lopatto, K. Ramanan and X. Xie, *Quenched large deviation principles for random projections of $\ell_p^n$ balls*, arXiv e-print, arXiv:2308.00649, 2023.

[25] E. Meckes, *Projections of probability distributions: A measure-theoretic Dvoretzky theorem*, Geometric Aspects of Functional Analysis, Lecture Notes in Mathematics, vol. 2050, Springer, 2012, pp. 317–326. MR2985300

[26] A. Onatski, M. J. Moreira, and M. Hallin, *Asymptotic power of sphericity tests for high-dimensional data*, The Annals of Statistics **41** (2013), no. 3, 1204–1231. MR3113808

[27] A. Onatski, M. J. Moreira, and M. Hallin, *Signal detection in high dimension: The multispiked case*, The Annals of Statistics **42** (2014), no. 1, 225–254. MR3189485

[28] S. T. Rachev and L. Rüschendorf, *Approximate independence of distributions on spheres and their stability properties*, The Annals of Probability **19** (1991), no. 3, 1311–1337. MR1112418

[29] R. T. Rockafellar, *Convex analysis*, no. 28, Princeton University Press, 1970. MR0274683

[30] G. Schechtman and J. 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. MR1015684

[31] M. Sion, *On general minimax theorems*, Pacific Journal of Mathematics **8** (1958), no. 1, 171–176. MR0097026

[32] J. A. Tropp, *A comparison principle for functions of a uniformly random subspace*, Probability Theory and Related Fields **153** (2012), no. 3-4, 759–769. MR2948692

[33] C. Villani, *Optimal transport: Old and new*, vol. 338, Springer Science & Business Media, 2008. MR2459454

[34] R. Wang, X. Wang, and L. Wu, *Sanov's theorem in the Wasserstein distance: a necessary and sufficient condition*, Statistics & Probability Letters **80** (2010), no. 5, 505–512. MR2593592