

# SMALL BALL PROBABILITIES FOR LINEAR IMAGES OF HIGH DIMENSIONAL DISTRIBUTIONS

MARK RUDELSON AND ROMAN VERSHYNIN

ABSTRACT. We study concentration properties of random vectors of the form  $AX$ , where  $X = (X_1, \dots, X_n)$  has independent coordinates and  $A$  is a given matrix. We show that the distribution of  $AX$  is well spread in space whenever the distributions of  $X_i$  are well spread on the line. Specifically, assume that the probability that  $X_i$  falls in any given interval of length  $t$  is at most  $p$ . Then the probability that  $AX$  falls in any given ball of radius  $t\|A\|_{\text{HS}}$  is at most  $(Cp)^{0.9r(A)}$ , where  $r(A)$  denotes the stable rank of  $A$  and  $C$  is an absolute constant.

## 1. INTRODUCTION

Concentration properties of high dimensional distributions have been extensively studied in probability theory. In this paper we are interested in *small ball probabilities*, which describe the spread of a distribution in space. Small ball probabilities have been extensively studied for stochastic processes (see [11]), sums of independent random variables (see [20, 18]) and log-concave measures (see [1, Chapter 5]). Nevertheless, there remain surprisingly basic questions that have not been previously addressed.

The main object of our study is a random vector  $X = (X_1, \dots, X_n)$  in  $\mathbb{R}^n$  with independent coordinates  $X_i$ . Given a fixed  $m \times n$  matrix  $A$ , we study the concentration properties of the random vector  $AX$ . We are interested in results of the following type:

*If the distributions of  $X_i$  are well spread on the line, then the distribution of  $AX$  is well spread in space.*

Special cases of interest are *marginals* of  $X$  which arise when  $A$  is an orthogonal projection, and *sums of independent random variables* which correspond to  $m = 1$ . The problem of describing small ball probabilities even in these two special cases is nontrivial and useful for applications. In particular, a recent interest in this problem was spurred by applications in random matrix theory; see [20, 18] for sums of random variables and [17] for higher dimensional marginals.

This discussion is non-trivial even for continuous distributions, and we shall start from this special case.

**1.1. Continuous distributions.** Our first main result is about distributions with independent continuous coordinates. It states that any  $d$ -dimensional marginal of such distribution has density bounded by  $O(1)^d$ , as long as the densities of the coordinates are bounded by  $O(1)$ .

---

*Date:* October 9, 2014.

Partially supported by NSF grants DMS 1161372, 1001829, 1265782 and USAF Grant FA9550-14-1-0009.

**Theorem 1.1** (Densities of projections). *Let  $X = (X_1, \dots, X_n)$  where  $X_i$  are real-valued independent random variables. Assume that the densities of  $X_i$  are bounded by  $K$  almost everywhere. Let  $P$  be the orthogonal projection in  $\mathbb{R}^n$  onto a  $d$ -dimensional subspace. Then the density of the random vector  $PX$  is bounded by  $(CK)^d$  almost everywhere.*

Here and throughout the paper,  $C, C_1, c, c_1, \dots$  denote positive absolute constants.

Theorem 1.1 is trivial in dimension  $d = n$ , since the product density is bounded by  $K^n$ . A remarkable non-trivial case of Theorem 1.1 is in dimension  $d = 1$ , where it holds with optimal constant  $C = \sqrt{2}$ . This partial case is worth to be stated separately.

**Theorem 1.2** (Densities of sums). *Let  $X_1, \dots, X_n$  be real-valued independent random variables whose densities are bounded by  $K$  almost everywhere. Let  $a_1, \dots, a_n$  be real numbers with  $\sum_{j=1}^n a_j^2 = 1$ . Then the density of  $\sum_{j=1}^n a_j X_j$  is bounded by  $\sqrt{2}K$  almost everywhere.*

*Moreover, this estimate is optimal. The density  $\sqrt{2}K$  is achieved by the sum  $\frac{1}{\sqrt{2}}(X_1 + X_2)$  where  $X_1, X_2$  are uniformly distributed in  $[-\frac{1}{2K}, \frac{1}{2K}]$ .*

Theorem 1.2 follows immediately from a combination of two known results, a theorem of B. Rogozin stated below and a theorem of K. Ball [2].

**Theorem 1.3** (Rogozin [16]). *Let  $Z_1, \dots, Z_n$  be real-valued independent random variables whose densities are bounded by  $d_1, \dots, d_n$  respectively. Then the density of  $\sum_{j=1}^n Z_j$  is uniformly bounded by the value of the density of  $\sum_{j=1}^n U_j$  at the origin, where  $U_j$  are independent random variables uniformly distributed in  $[-\frac{1}{2d_j}, \frac{1}{2d_j}]$ .*

*Proof of Theorem 1.2.* For simplicity, by rescaling we can assume that  $K = 1$ . Rogozin's Theorem 1.3 implies that the density of  $\sum_{j=1}^n a_j X_j$  is uniformly bounded by the density of  $\sum_{j=1}^n a_j Y_j$  at the origin, where  $Y_j$  are independent random variables uniformly distributed in  $[-\frac{1}{2}, \frac{1}{2}]$ .

The random vector  $Y = (Y_1, \dots, Y_n)$  is uniformly distributed in the cube  $[-\frac{1}{2}, \frac{1}{2}]^n$ . Thus the density of  $\sum_{j=1}^n a_j Y_j$  at the origin is the volume of the section of the cube by the hyperplane that contains the origin and is orthogonal to the vector  $a = (a_1, \dots, a_n)$ . Now, a theorem of Ball [2] states that the maximal volume of such section equals  $\sqrt{2}$ . Therefore, the density of  $\sum_{j=1}^n a_j X_j$  is bounded by  $\sqrt{2}$ .  $\square$

**1.2. General distributions.** Using a simple smoothing argument from [4], Theorem 1.1 can be extended for general, not necessarily continuous, distributions. The spread of general distributions is conveniently measured by the *concentration function*. For a random vector  $Z$  taking values in a  $\mathbb{R}^n$ , the concentration function is defined as

$$\mathcal{L}(Z, t) = \max_{u \in \mathbb{R}^n} \mathbb{P} \{ \|Z - u\|_2 \leq t \}, \quad t \geq 0. \quad (1.1)$$

Thus the concentration function controls the *small ball probabilities* of the distribution of  $Z$ . The study of concentration functions of sums of independent random variables originates from the works of Lévy [10], Kolmogorov [9], Rogozin [15], Esseen [6] and Halasz [8]. Recent developments in this area highlighted connections with Littlewood-Offord problem and applications to random matrix theory, see [20, 18].

**Corollary 1.4** (Concentration function of projections). *Consider a random vector  $X = (X_1, \dots, X_n)$  where  $X_i$  are real-valued independent random variables. Let  $t, p \geq 0$  be such that*

$$\mathcal{L}(X_i, t) \leq p \quad \text{for all } i = 1, \dots, n.$$

*Let  $P$  be an orthogonal projection in  $\mathbb{R}^n$  onto a  $d$ -dimensional subspace. Then*

$$\mathcal{L}(PX, t\sqrt{d}) \leq (Cp)^d.$$

This result can be regarded as a tensorization property of the concentration function. It will be deduced from Theorem 1.1 in Section 2.

**1.3. Anisotropic distributions.** Finally, we study concentration of anisotropic high-dimensional distributions, which take the form  $AX$  for a fixed matrix  $A$ . The key exponent that controls the behavior of the concentration function of  $AX$  is the *stable rank* of  $A$ . We define it as

$$r(A) = \left\lfloor \frac{\|A\|_{\text{HS}}^2}{\|A\|^2} \right\rfloor \quad (1.2)$$

where  $\|\cdot\|_{\text{HS}}$  denotes the Hilbert-Schmidt norm.<sup>1</sup> Note that for any non-zero matrix,  $1 \leq r(A) \leq \text{rank}(A)$ .

**Theorem 1.5** (Concentration function of anisotropic distributions). *Consider a random vector  $X = (X_1, \dots, X_n)$  where  $X_i$  are real-valued independent random variables. Let  $t, p \geq 0$  be such that*

$$\mathcal{L}(X_i, t) \leq p \quad \text{for all } i = 1, \dots, n.$$

*Let  $A$  be an  $m \times n$  matrix and  $\varepsilon \in (0, 1)$ . Then*

$$\mathcal{L}(AX, t\|A\|_{\text{HS}}) \leq (C_\varepsilon p)^{(1-\varepsilon)r(A)},$$

*where  $C_\varepsilon$  depends only on  $\varepsilon$ .*

A more precise version of this result is Theorem 8.1 and Corollary 8.5 below. It will be deduced from Corollary 1.4 by replacing  $A$  by a dyadic sum of spectral projections.

*Remark 1.6* (Scaling). To understand Theorem 1.5 better, note that  $\mathbb{E}\|AX\|_2^2 = \|A\|_{\text{HS}}^2$  if all  $X_i$  have zero means and unit variances. This explains the scaling factor  $\|A\|_{\text{HS}}$  in Theorem 1.5. Further, if  $A$  is an orthogonal projection of rank  $d$ , then  $\|A\|_{\text{HS}} = \sqrt{d}$  and  $r(A) = d$ , which recovers Corollary 1.4 in this case up to  $\varepsilon$  in the exponent. Moreover, Theorem 8.1 and Corollary 8.6 below will allow precise recovery, without any loss of  $\varepsilon$ .

*Remark 1.7* (Continuous distributions). In the particular case of continuous distributions, Theorem 1.5 states the following. Suppose the densities of  $X_i$  are bounded by  $K$ . Then obviously  $\mathcal{L}(X_i, t) \leq Kt$  for any  $t \geq 0$ , so Theorem 1.5 yields

$$\mathcal{L}(AX, t\|A\|_{\text{HS}}) \leq (C_\varepsilon t)^{(1-\varepsilon)r(A)}, \quad t \geq 0. \quad (1.3)$$

A similar inequality was proved by Paouris [13] for random vectors  $X$  which satisfy three conditions: (a)  $X$  is isotropic, i.e. all one-dimensional marginals of  $X$  have unit

<sup>1</sup>This definition differs slightly from the traditional definition of stable rank, in which one does not take the floor function, i.e. where  $r(A) = \|A\|_{\text{HS}}^2 / \|A\|^2$ .

variance; (b) the distribution of  $X$  is log-concave; (c) all one-dimensional marginals are uniformly sub-gaussian.<sup>2</sup> The inequality of Paouris states in this case that

$$\mathcal{L}(AX, t\|A\|_{\text{HS}}) \leq (Ct)^{cr(A)}, \quad t \geq 0. \quad (1.4)$$

Here  $C$  is an absolute constant and  $c \in (0, 1)$  depends only on the bound on the sub-gaussian norms. The distributions for which Paouris' inequality (1.4) applies are not required to have independent coordinates. On the other hand, the log-concavity assumption for (1.4) is much stronger than a uniform bound on the coordinate densities in (1.3).

*Remark 1.8* (Large deviations). It is worthwhile to state here a related large deviation bound for  $AX$  from [19]. If  $X_i$  are independent, uniformly sub-gaussian, and have zero means and unit variances, then

$$\mathbb{P}\{|\|AX\|_2 - \|A\|_{\text{HS}}| \geq t\|A\|_{\text{HS}}\} \leq 2e^{-ct^2r(A)}, \quad t \geq 0.$$

Here  $c > 0$  depends only on the bound on the sub-gaussian norms of  $X_i$ .

**1.4. The method.** Let us outline the proof of the key Theorem 1.1, which implies all other results in this paper.

A natural strategy would be to extend to higher dimensions the simple one-dimensional argument leading to Theorem 1.2, which was a combination of Ball's and Rogozin's theorems. A higher-dimensional version of Ball's theorem is indeed available [3]; it states that the maximal volume of a section of the cube by a subspace of codimension  $d$  is  $(\sqrt{2})^d$ . However, we are unaware of any higher-dimensional versions of Rogozin's theorem [16].

An alternative approach to the special case of Theorem 1.1 in dimension  $d = 1$  (and, as a consequence, to Corollary 1.4 in dimension  $d = 1$ ) was developed in an unpublished manuscript of Ball and Nazarov [4]. Although it does not achieve the optimal constant  $\sqrt{2}$  that appears in Theorem 1.2, this approach avoids the delicate combinatorial arguments that appear in the proof of Rogozin's theorem. The method of Ball and Nazarov is Fourier-theoretic; its crucial steps go back to Halasz [7, 8] and Ball [2].

In this paper, we prove Theorem 1.1 by generalizing the method of Ball and Nazarov [4] to higher dimensions using Brascamp-Lieb inequality. For educational purposes, we will start by presenting a version of Ball-Nazarov's argument in dimension  $d = 1$  in Sections 3 and 4. The higher-dimensional argument will be presented in Sections 5–7.

There turned out to be an unexpected difference between dimension  $d = 1$  and higher dimensions, which presents us with an extra challenge. The one-dimensional method works well under assumption that all coefficients  $a_i$  are small, e.g.  $|a_i| \leq 1/2$ . The opposite case where there is a large coefficient  $a_{i_0}$ , is trivial; it can be treated by conditioning on all  $X_i$  except  $X_{i_0}$ .

In higher dimensions, this latter case is no longer trivial. It corresponds to the situation where some  $\|Pe_{i_0}\|_2$  is large (here  $e_i$  denote the coordinate basis vectors). The power of one random variable  $X_{i_0}$  is not enough to yield Theorem 1.1; such argument would lead to a weaker bound  $(CK\sqrt{d})^d$  instead of  $(CK)^d$ .

---

<sup>2</sup>Recall that a random variable  $Z$  is sub-gaussian if  $\mathbb{P}\{|Z| > t\} \leq 2\exp(-t^2/M^2)$  for all  $t \geq 0$ . The smallest  $M \geq 0$  here can be taken as a definition of the sub-gaussian norm of  $Z$ ; see [21].

In Section 6 we develop an alternative way to remove the terms with large  $\|Pe_i\|_2$  from the sum. It is based on a careful tensorization argument for small ball probabilities.

## 2. DEDUCTION OF COROLLARY 1.4 FROM THEOREM 1.1

We begin by recording a couple of elementary properties of concentration functions.

**Proposition 2.1** (Regularity of concentration function). *Let  $Z$  be a random variable taking values in a  $d$ -dimensional subspace of  $\mathbb{R}^n$ . Then for every  $M \geq 1$  and  $t \geq 0$ , we have*

$$\mathcal{L}(Z, t) \leq \overline{\mathcal{L}}(Z, Mt) \leq (3M)^d \cdot \mathcal{L}(Z, t).$$

*Proof.* The lower estimate is trivial. The upper estimate follows once we recall that a ball of radius  $Mt$  in  $\mathbb{R}^d$  can be covered by  $(3M)^d$  balls of radius  $t$ .  $\square$

Throughout this paper, it will be convenient to work with the following equivalent definition of density. For a random vector  $Z$  taking values in a  $d$ -dimensional subspace  $E$  of  $\mathbb{R}^n$ , the density can be defined as

$$f_Z(u) = \limsup_{t \rightarrow 0_+} \frac{1}{|B(t)|} \mathbb{P}\{\|Z - u\|_2 \leq t\}, \quad u \in E, \quad (2.1)$$

where  $|B(t)|$  denotes the volume of a Euclidean ball with radius  $t$  in  $\mathbb{R}^d$ . Lebesgue differentiation theorem states that for random variable  $Z$  with absolutely continuous distribution,  $f_Z(u)$  equals the actual density of  $Z$  almost everywhere.

The following elementary observation connects densities and concentration functions.

**Proposition 2.2** (Concentration function and densities). *Let  $Z$  be a random variable taking values in a  $d$ -dimensional subspace of  $\mathbb{R}^n$ . Then the following assertions are equivalent:*

- (i) *The density of  $Z$  is bounded by  $K^d$  almost everywhere;*
- (ii) *The concentration function of  $Z$  satisfies*

$$\mathcal{L}(Z, t\sqrt{d}) \leq (Mt)^d \quad \text{for all } t \geq 0.$$

Here  $K$  and  $M$  depend only on each other. In the implication (i)  $\Rightarrow$  (ii), we have  $M \leq CK$  where  $C$  is an absolute constant. In the implication (ii)  $\Rightarrow$  (i), we have  $K \leq M$ .

This proposition follows from the known bound  $t^d \leq |B(t\sqrt{d})| \leq (Ct)^d$  (see e.g. formula (1.18) in [14]).  $\square$

Now we are ready to deduce Corollary 1.4 from Theorem 1.1. The proof is a higher-dimensional version of the smoothing argument of Ball and Nazarov [4].

*Proof of Corollary 1.4.* We can assume by approximation that  $t > 0$ ; then by rescaling (replacing  $X$  with  $X/t$ ) we can assume that  $t = 1$ . Furthermore, translating  $X$  if necessary, we reduce the problem to bounding  $\mathbb{P}\{\|PX\|_2 \leq \sqrt{d}\}$ . Consider independent random variables  $Y_i$  uniformly distributed in  $[-1/2, 1/2]$ , which are also

jointly independent of  $X$ . We are seeking to replace  $X$  by  $X' := X + Y$ . By triangle inequality and independence, we have

$$\begin{aligned} \mathbb{P} \left\{ \|PX'\|_2 \leq 2\sqrt{d} \right\} &\geq \mathbb{P} \left\{ \|PX\|_2 \leq \sqrt{d} \text{ and } \|PY\|_2 \leq \sqrt{d} \right\} \\ &= \mathbb{P} \left\{ \|PX\|_2 \leq \sqrt{d} \right\} \mathbb{P} \left\{ \|PY\|_2 \leq \sqrt{d} \right\}. \end{aligned}$$

An easy computation yields  $\mathbb{E} \|PY\|_2^2 = d/12$ , so Markov's inequality implies that  $\mathbb{P}\{\|PY\|_2 \leq \sqrt{d}\} \geq 11/12$ . It follows that

$$\mathbb{P} \left\{ \|PX\|_2 \leq \sqrt{d} \right\} \leq \frac{12}{11} \mathbb{P} \left\{ \|PX'\|_2 \leq 2\sqrt{d} \right\}. \quad (2.2)$$

Note that  $X' = X + Y$  has independent coordinates whose densities can be computed as follows:

$$f_{X'_i}(u) = \mathbb{P} \{|X_i - u| \leq 1/2\} \leq \mathcal{L}(X_i, 1/2).$$

Applying Theorem 1.1, we find that the density of  $PX'$  is bounded by  $(CL)^d$ , where  $L = \max_i \mathcal{L}(X_i, 1/2)$ . Then Proposition 2.2 yields that

$$\mathbb{P} \left\{ \|PX'\|_2 \leq 2\sqrt{d} \right\} \leq \mathcal{L}(PX', 2\sqrt{d}) \leq (C_1 L)^d.$$

Substituting this into (2.2), we complete the proof.  $\square$

*Remark 2.3* (Flexible scaling in Corollary 1.4). Using regularity of concentration function described in Proposition 2.1, one can state the conclusion of Corollary 1.4 in a more flexible way:

$$\mathcal{L}(PX, Mt\sqrt{d}) \leq (CMp)^d, \quad M \geq 1.$$

We will use this observation later.

### 3. DECAY OF CHARACTERISTIC FUNCTIONS

We will now begin preparing our way for the proof of Theorem 1.1. Our argument will use the following tail bound for the characteristic function

$$\phi_X(t) = \mathbb{E} e^{itX}$$

of a random variable  $X$  with bounded density. The estimate and its proof below are essentially due to Ball and Nazarov [4]. The reader can refer to [12, Chapter 3] for the definition and basic properties of the decreasing rearrangements of functions.

**Lemma 3.1** (Decay of characteristic functions). *Let  $X$  be a random variable whose density is bounded by  $K$ . Then the non-increasing rearrangement of the characteristic function of  $X$  satisfies*

$$|\phi_X|^*(t) \leq \begin{cases} 1 - c(t/K)^2, & 0 < t < 2\pi K \\ \sqrt{2\pi K/t}, & t \geq 2\pi K. \end{cases}$$

*Proof.* The estimate for large  $t$  will follow from Plancherel's identity. The estimate for small  $t$  will be based on a regularity argument going back to Halasz [7].

**1. Plancherel.** By replacing  $X$  with  $KX$  we can assume that  $K = 1$ . Let  $f_X(\cdot)$  denote the density of  $X$ . Thus  $\phi_X(t) = \int_{-\infty}^{\infty} f_X(x) e^{itx} dx = \widehat{f_X}(-t/2\pi)$ , according to the standard definition of the Fourier transform

$$\widehat{f}(t) = \int_{-\infty}^{\infty} f(x) e^{-2\pi itx} dx. \quad (3.1)$$

By Plancherel's identity and using that  $\|p_X\|_{L^1} = 1$  and  $\|p_X\|_{L^\infty} \leq K = 1$ , we obtain

$$\|\phi_X\|_{L^2} = \sqrt{2\pi}\|p_X\|_{L^2} \leq \sqrt{2\pi\|p_X\|_{L^1}\|p_X\|_{L^\infty}} \leq \sqrt{2\pi}.$$

Chebyshev's inequality then yields

$$|\phi_X|^*(t) \leq \sqrt{\frac{2\pi}{t}}, \quad t > 0. \quad (3.2)$$

This proves the second part of the claimed estimate. It remains to prove the first part.

**2. Symmetrization.** Let  $X'$  denote an independent copy of  $X$ . Then

$$\begin{aligned} |\phi_X(t)|^2 &= \mathbb{E} e^{itX} \mathbb{E} \overline{e^{itX}} = \mathbb{E} e^{itX} \mathbb{E} e^{-itX'} = \mathbb{E} e^{it(X-X')} \\ &= \phi_{\tilde{X}}(t), \quad \text{where } \tilde{X} := X - X'. \end{aligned}$$

Further, by symmetry of the distribution of  $\tilde{X}$ , we have

$$\phi_{\tilde{X}}(t) = \mathbb{E} \cos(t|X|) = 1 - 2\mathbb{E} \sin^2\left(\frac{1}{2}t|\tilde{X}|\right) =: 1 - \psi(t).$$

We are going to prove a bound of the form

$$\lambda\{\tau : \psi(\tau) \leq s^2\} \leq Cs, \quad 0 < s \leq 1/2 \quad (3.3)$$

where  $\lambda$  denotes the Lebesgue measure on  $\mathbb{R}$ . Combined with the identity  $|\phi_X(t)|^2 = 1 - \psi(t)$ , this bound would imply

$$|\phi_X|^*(Cs) \leq \sqrt{1 - s^2} \leq 1 - s^2/2, \quad 0 < s \leq 1/2.$$

Substituting  $t = Cs$ , we would obtain the desired estimate

$$|\phi_X|^*(t) \leq 1 - s^2/2C^2, \quad 0 < s \leq C/2,$$

which would conclude the proof (provided  $C$  is chosen large enough so that  $C/2 \geq 2\pi$ ).

**3. Regularity.** First we observe that (3.3) holds for some fixed constant value of  $s$ . This follows from the identity  $|\phi_X(\tau)|^2 = 1 - \psi(\tau)$  and inequality (3.2):

$$\lambda\left\{\tau : \psi(\tau) \leq \frac{1}{4}\right\} = \lambda\{\tau : |\phi_X(\tau)| \geq \sqrt{3/4}\} \leq 8\pi/3 \leq 9. \quad (3.4)$$

Next, the definition of  $\psi(\cdot)$  and the inequality  $|\sin(mx)| \leq m|\sin x|$  valid for  $x \in \mathbb{R}$  and  $m \in \mathbb{N}$  imply that

$$\psi(mt) \leq m^2\psi(t), \quad t > 0, \quad m \in \mathbb{N}.$$

Therefore

$$\lambda\left\{\tau : \psi(\tau) \leq \frac{1}{4m^2}\right\} \leq \lambda\left\{\tau : \psi(m\tau) \leq \frac{1}{4}\right\} = \frac{1}{m} \lambda\left\{\tau : \psi(\tau) \leq \frac{1}{4}\right\} \leq \frac{9}{m}, \quad (3.5)$$

where in the last step we used (3.4). This establishes (3.3) for the discrete set of values  $t = \frac{1}{2m}$ ,  $m \in \mathbb{N}$ . We can extend this to arbitrary  $t > 0$  in a standard way, by applying (3.5) for  $m \in \mathbb{N}$  such that  $t \in (\frac{1}{4m}, \frac{1}{2m}]$ . This proves (3.3) and completes the proof of Lemma 3.1.  $\square$

## 4. THEOREM 1.1 IN DIMENSION ONE. DENSITIES OF SUMS.

Now we are going to give a “soft” proof of a version of Theorem 1.2 due to Ball and Nazarov [4]. Their argument establishes Theorem 1.1 in dimension  $d = 1$ . Let us state this result separately.

**Theorem 4.1** (Densities of sums). *Let  $X_1, \dots, X_n$  be real-valued independent random variables whose densities are bounded by  $K$  almost everywhere. Let  $a_1, \dots, a_n$  be real numbers with  $\sum_{j=1}^n a_j^2 = 1$ . Then the density of  $\sum_{j=1}^n a_j X_j$  is bounded by  $CK$  almost everywhere.*

*Proof.* By replacing  $X_j$  with  $KX_j$  we can assume that  $K = 1$ . By replacing  $X_j$  with  $-X_j$  when necessary we can assume that all  $a_j \geq 0$ . We can further assume that  $a_j > 0$  by dropping all zero terms from the sum. If there exists  $j_0$  with  $a_{j_0} > 1/2$ , then the conclusion follows by conditioning on all  $X_j$  except  $X_{j_0}$ . Thus we can assume that

$$0 < a_j < \frac{1}{2} \quad \text{for all } j.$$

Finally, by translating  $X_j$  if necessary we reduce the problem to bounding the density of  $S = \sum_j a_j X_j$  at the origin.

We may assume that  $\phi_{X_j} \in L_1$  by adding to  $X_j$  an independent normal random variable with an arbitrarily small variance. Fourier inversion formula associated with the Fourier transform (3.1) yields that the density of  $S$  at the origin (defined using (2.1)) can be reconstructed from its Fourier transform as

$$f_S(0) = \int_{\mathbb{R}} \widehat{f_S}(x) dx = \int_{\mathbb{R}} \phi_S(2\pi x) dx \leq \int_{\mathbb{R}} |\phi_S(x)| dx =: I. \quad (4.1)$$

By independence, we have  $\phi_S(x) = \prod_j \phi_{X_j}(a_j t)$ , so

$$I = \int_{\mathbb{R}} \prod_j |\phi_{X_j}(a_j x)| dx.$$

We use the generalized Hölder’s inequality with exponents  $1/a_j^2$  whose reciprocals sum to 1 by assumption. It yields

$$I \leq \prod_j \left( \int_{\mathbb{R}} |\phi_{X_j}(a_j x)|^{1/a_j^2} dx \right)^{a_j^2}. \quad (4.2)$$

The value of the integrals will not change if we replace the functions  $|\phi_{X_j}|$  by their non-increasing rearrangements  $|\phi_{X_j}|^*$ . After change of variable, we obtain

$$I \leq \prod_j \left( \frac{1}{a_j} \int_0^\infty |\phi_{X_j}|^*(x)^{1/a_j^2} dx \right)^{a_j^2}.$$

We use Lemma 3.1 to bound the integrals

$$I_j := \int_0^\infty |\phi_{X_j}|^*(x)^{1/a_j^2} dx \leq \int_0^{2\pi} (1 - cx^2)^{1/a_j^2} dx + \int_{2\pi}^\infty (2\pi/x)^{1/(2a_j^2)} dx.$$

Bounding  $1 - cx^2$  by  $e^{-cx^2}$ , we see that the first integral (over  $[0, 2\pi]$ ) is bounded by  $Ca_j$ . The second integral (over  $[2\pi, \infty)$ ) is bounded by

$$\frac{2\pi}{1/(2a_j^2) - 1} \leq 8\pi a_j^2,$$

where we used that  $a_j \leq 1/2$ . Therefore

$$I_j \leq Ca_j + 8\pi a_j^2 \leq 2Ca_j$$

provided that constant  $C$  is chosen large enough. Hence

$$I \leq \prod_j (2C)^{a_j^2} = (2C)^{\sum a_j^2} = 2C.$$

Substituting this into (4.1) completes the proof.  $\square$

## 5. TOWARD THEOREM 1.1 IN HIGHER DIMENSIONS. THE CASE OF SMALL $Pe_j$ .

Our proof of Theorem 1.1 will go differently depending on whether all vectors  $Pe_j$  are small or some  $Pe_j$  are large. In the first case, we proceed with a high-dimensional version of the argument from Section 4, where Hölder's inequality will be replaced by Brascamp-Lieb's inequality. In the second case, we will remove the large vectors  $Pe_j$  one by one, using a new precise tensorization property of concentration functions.

In this section, we treat the case where all vectors  $Pe_j$  are small. Theorem 1.1 can be formulated in this case as follows.

**Proposition 5.1.** *Let  $X$  be a random vector and  $P$  be a projection which satisfy the assumptions of Theorem 1.1. Assume that*

$$\|Pe_j\|_2 \leq 1/2 \quad \text{for all } j = 1, \dots, n.$$

*Then the density of the random vector  $PX$  is bounded by  $(CK)^d$  almost everywhere.*

The proof will be based on Brascamp-Lieb's inequality.

**Theorem 5.2** (Brascamp-Lieb [5], see [3]). *Let  $u_1, \dots, u_n \in \mathbb{R}^d$  be unit vectors and  $c_1, \dots, c_n > 0$  be real numbers satisfying*

$$\sum_{i=1}^n c_i u_i u_i^\top = I_d.$$

*Let  $f_1, \dots, f_n : \mathbb{R} \rightarrow [0, \infty)$  be integrable functions. Then*

$$\int_{\mathbb{R}^n} \prod_{j=1}^n f_j(\langle x, u_j \rangle)^{c_j} dx \leq \prod_{j=1}^n \left( \int_{\mathbb{R}} f_j(t) dt \right)^{c_j}.$$

*Proof of Proposition 5.1.* The singular value decomposition of  $P$  yields the existence of a  $d \times n$  matrix  $R$  satisfying

$$P = R^\top R, \quad RR^\top = I_d.$$

It follows that  $\|Px\|_2 = \|Rx\|_2$  for all  $x \in \mathbb{R}^d$ . This allows us to replace  $P$  by  $R$  in the statement of the proposition. Moreover, by replacing  $X_j$  with  $KX_j$  we can assume that  $K = 1$ . Finally, translating  $X$  if necessary we reduce the problem to bounding the density of  $RX$  at the origin.

As in the proof of Theorem 4.1, Fourier inversion formula associated with the Fourier transform in  $n$  dimensions yields that the density of  $RX$  at the origin (defined using (2.1)) can be reconstructed from its Fourier transform as

$$f_{RX}(0) = \int_{\mathbb{R}^d} \widehat{f_{RX}}(x) dx = \int_{\mathbb{R}^d} \phi_{RX}(2\pi x) dx \leq \int_{\mathbb{R}^d} |\phi_{RX}(x)| dx \quad (5.1)$$

where

$$\phi_{RX}(x) = \mathbb{E} \exp(i\langle x, RX \rangle) \quad (5.2)$$

is the characteristic function of  $RX$ . Therefore, to complete the proof, it suffices to bound the integral in the right hand side of (5.1) by  $C^d$ .

In order to represent  $\phi_{RX}(x)$  more conveniently for application of Brascamp-Lieb inequality, we denote

$$a_j := \|Re_j\|_2, \quad u_j := \frac{Re_j}{\|Re_j\|_2}.$$

Then  $R = \sum_{j=1}^n a_j u_j e_j^\top$ , so the identity  $RR^\top = I_d$  can be written as

$$\sum_{j=1}^n a_j^2 u_j u_j^\top = I_d. \quad (5.3)$$

Moreover, we have  $\langle x, RX \rangle = \sum_{j=1}^n a_j \langle x, u_j \rangle X_j$ . Substituting this into (5.2) and using independence, we obtain

$$\phi_{RX}(x) = \prod_{j=1}^n \mathbb{E} \exp(ia_j \langle x, u_j \rangle X_j).$$

Define the functions  $f_1, \dots, f_n : \mathbb{R} \rightarrow [0, \infty)$  as

$$f_j(t) := |\mathbb{E} \exp(ia_j t X_j)|^{1/a_j^2} = |\phi_{X_j}(a_j t)|^{1/a_j^2}.$$

Recalling (5.3), we apply Brascamp-Lieb inequality for these functions and obtain

$$\begin{aligned} \int_{\mathbb{R}^d} |\phi_{RX}(x)| dx &= \int_{\mathbb{R}^d} \prod_{j=1}^n f_j(\langle x, u_j \rangle)^{a_j^2} dx \\ &\leq \prod_{j=1}^n \left( \int_{\mathbb{R}} f_j(t) dt \right)^{a_j^2} = \prod_{j=1}^n \left( \int_{\mathbb{R}} |\phi_{X_j}(a_j t)|^{1/a_j^2} dt \right)^{a_j^2}. \end{aligned} \quad (5.4)$$

We arrived at the same quantity as we encountered in one-dimensional argument in (4.2). Following that argument, which uses the assumption that all  $a_j \leq 1/2$ , we bound the product the quantity above by

$$(2C)^{\sum_{j=1}^n a_j^2}.$$

Recalling that  $a_j = \|Re_j\|_2$  and, we find that  $\sum_{j=1}^n a_j^2 = \sum_{j=1}^n \|Re_j\|_2^2 = \text{tr}(RR^\top) = \text{tr}(I_d) = d$ . Thus the right hand side of (5.4) is bounded by  $(2C)^d$ . The proof of Proposition 5.1 is complete.  $\square$

## 6. TOWARD THEOREM 1.1 IN HIGHER DIMENSIONS. REMOVAL OF LARGE $Pe_j$ .

Next we turn to the case where not all vectors  $Pe_i$  are small. In this case, we will remove the large vectors  $Pe_i$  one by one. The non-trivial task is how not to lose power at each step. This will be achieved with the help of the following precise tensorization property of small ball probabilities.

**Lemma 6.1** (Tensorization). *Let  $Z_1, Z_2 \geq 0$  be random variables and  $M_1, M_2, p \geq 0$  be real numbers. Assume that*

- (i)  $\mathbb{P}\{Z_1 \leq t \mid Z_2\} \leq M_1 t$  almost surely in  $Z_2$  for all  $t \geq 0$ ;
- (ii)  $\mathbb{P}\{Z_2 \leq t\} \leq M_2 t^p$  for all  $t \geq 0$ .

Then

$$\mathbb{P} \left\{ \sqrt{Z_1^2 + Z_2^2} \leq t \right\} \leq \frac{CM_1M_2}{\sqrt{p+1}} t^{p+1} \quad \text{for all } t \geq 0.$$

*Remark 6.2.* This lemma will be used later in an inductive argument. To make the inductive step, two features will be critical: (a) the term of order  $\sqrt{p}$  in the denominator of the probability estimate; (b) the possibility of choosing different values for the parameters  $M_1$  and  $M_2$ .

*Proof.* Denoting  $s = t^2$ , we compute the probability by iterative integration in  $(Z_1^2, Z_2^2)$  plane:

$$\mathbb{P} \{Z_1^2 + Z_2^2 \leq s\} = \int_0^s \mathbb{P} \left\{ Z_1 \leq (s-x)^{1/2} \mid Z_2^2 = x \right\} dF_2(x) \quad (6.1)$$

where  $F_2(x) = \mathbb{P} \{Z_2^2 \leq x\}$  is the cumulative distribution function of  $Z_2^2$ . Using hypothesis (i) of the lemma, we can bound the right hand side of (6.1) by

$$M_1 \int_0^s (s-x)^{1/2} dF_2(x) = \frac{M_1}{2} \int_0^s F_2(x)(s-x)^{-1/2} dx,$$

where the last equation follows by integration by parts. Hypothesis (ii) of the lemma says that  $F_2(x) \leq M_2 x^{p/2}$ , so the expression above is bounded by

$$\frac{M_1M_2}{2} \int_0^s x^{p/2}(s-x)^{-1/2} dx = \frac{M_1M_2}{2} s^{\frac{p+1}{2}} \int_0^1 u^{p/2}(1-u)^{-1/2} du$$

where the last equation follows by substitution  $x = su$ .

The integral in the right hand side is the value of beta function

$$B\left(\frac{p}{2} + 1, \frac{1}{2}\right).$$

Bounding this value is standard. One can use the fact that

$$B(x, y) \sim \Gamma(y)x^{-y} \quad \text{as } x \rightarrow \infty, \quad y \text{ fixed}$$

which follows from the identity  $B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$  and Stirling's approximation. (Here  $f(x) \sim g(x)$  means that  $f(x)/g(x) \rightarrow 1$ .) It follows that

$$B\left(\frac{p}{2} + 1, \frac{1}{2}\right) \sim \Gamma\left(\frac{1}{2}\right) \left(\frac{p}{2} + 1\right)^{-1/2} \quad \text{as } p \rightarrow \infty.$$

Therefore the ratio of the left and right hand sides is bounded in  $p$ . Hence there exists an absolute constant  $C$  such that

$$B\left(\frac{p}{2} + 1, \frac{1}{2}\right) \leq \frac{C}{\sqrt{p+1}} \quad \text{for all } p \geq 0.$$

We have proved that

$$\mathbb{P} \{Z_1^2 + Z_2^2 \leq s\} \leq \frac{M_1M_2}{2} s^{\frac{p+1}{2}} \frac{C}{\sqrt{p+1}}.$$

Substituting  $s = t^2$  finishes the proof.  $\square$

**Corollary 6.3** (Tensorization, continued). *Let  $Z_1, Z_2 \geq 0$  be random variables and  $K_1, K_2 \geq 0, d > 1$  be real numbers. Assume that*

- (i)  $\mathbb{P} \{Z_1 \leq t \mid Z_2\} \leq K_1 t$  almost surely in  $Z_2$  for all  $t \geq 0$ ;
- (ii)  $\mathbb{P} \{Z_2 \leq t\sqrt{d-1}\} \leq (K_2 t)^{d-1}$  for all  $t \geq 0$ .

Then

$$\mathbb{P} \left\{ \sqrt{Z_1^2 + Z_2^2} \leq t\sqrt{d} \right\} \leq (K_2 t)^d \quad \text{for all } t \geq 0,$$

provided that  $K_1 \leq cK_2$  with a suitably small absolute constant  $c$ .

*Proof.* Random variables  $Z_1, Z_2$  satisfy the assumptions of Lemma 6.1 with

$$M_1 = K_1, \quad M_2 = \left( \frac{K_2}{\sqrt{d-1}} \right)^{d-1}, \quad p = d-1.$$

The conclusion of that lemma is that

$$\mathbb{P} \left\{ \sqrt{Z_1^2 + Z_2^2} \leq t\sqrt{d} \right\} \leq CK_1 \left( \frac{K_2}{\sqrt{d-1}} \right)^{d-1} \frac{1}{\sqrt{d}} (t\sqrt{d})^d.$$

Using the hypothesis that  $K_1 \leq cK_2$ , we bound the right hand side by

$$Cc(K_2 t)^d \left( \frac{d}{d-1} \right)^{\frac{d-1}{2}} \leq 3Cc(K_2 t)^d.$$

If we choose  $c = 1/(3C)$  then the right hand side gets bounded by  $(K_2 t)^d$ , as claimed.  $\square$

**Proposition 6.4** (Removal of large  $Pe_i$ ). *Let  $X$  be a random vector satisfying the assumptions of Theorem 1.1, and let  $P$  be an orthogonal projection in  $\mathbb{R}^n$  onto a  $d$ -dimensional subspace. Let  $\nu > 0$ , and assume that there exists  $i \in \{1, \dots, n\}$  such that*

$$\|Pe_i\|_2 \geq \nu.$$

Define  $Q$  to be the orthogonal projection in  $\mathbb{R}^n$  such that

$$\ker(Q) = \text{span}\{\ker(P), Pe_i\}.$$

Let  $M \geq C_0$  where  $C_0$  is an absolute constant. If

$$\mathbb{P} \left\{ \|QX\|_2 \leq t\sqrt{d-1} \right\} \leq (MKt/\nu)^{d-1} \quad \text{for all } t \geq 0, \quad (6.2)$$

then

$$\mathbb{P} \left\{ \|PX\|_2 \leq t\sqrt{d} \right\} \leq (MKt/\nu)^d \quad \text{for all } t \geq 0.$$

*Proof.* Without loss of generality, we can assume that  $i = 1$ . Let us record a few basic properties of  $Q$ . A straightforward check shows that

$$P - Q \text{ is the orthogonal projection onto } \text{span}(Pe_1). \quad (6.3)$$

It follows that  $(P-Q)e_1 = Pe_1$ , since the orthogonal projection of  $e_1$  onto  $\text{span}(Pe_1)$  equals  $Pe_1$ . Canceling  $Pe_1$  on both sides, we have

$$Qe_1 = 0. \quad (6.4)$$

It follows from (6.3) that  $P$  has the form

$$Px = \left( \sum_{j=1}^n a_j x_j \right) Pe_1 + Qx \quad \text{for } x = (x_1, \dots, x_n) \in \mathbb{R}^n, \quad (6.5)$$

where  $a_j$  are fixed numbers (independent of  $x$ ). Substituting  $x = e_1$ , we obtain using (6.4) that  $Pe_1 = a_1 Pe_1 + Qe_1 = a_1 Pe_1$ . Thus

$$a_1 = 1. \quad (6.6)$$

Furthermore, we note that

$$Qx \text{ does not depend on } x_1 \quad (6.7)$$

since  $Qx = Q(\sum_{i=1}^n x_j e_j) = \sum_{i=1}^n x_j Qe_j$  and  $Qe_1 = 0$  by (6.4). Finally, since  $Pe_1$  is orthogonal to the image of  $Q$ , the two vectors in the right side of (6.5) are orthogonal. Thus

$$\|Px\|_2^2 = \left( \sum_{j=1}^n a_j x_j \right)^2 \|Pe_1\|_2^2 + \|Qx\|_2^2. \quad (6.8)$$

Now let us estimate  $\|PX\|_2$  for a random vector  $X$ . We express  $\|PX\|_2^2$  using (6.8) and (6.6) as

$$\|PX\|_2^2 = \left( X_1 + \sum_{j=2}^n a_j X_j \right)^2 \|Pe_1\|_2^2 + \|QX\|_2^2 =: Z_1^2 + Z_2^2$$

and try to apply Corollary 6.3. Let first us check the hypotheses of that corollary. Since by (6.7)  $Z_2$  is determined by  $X_2, \dots, X_n$  (and is independent of  $X_1$ ), and  $\|Pe_i\|_2 \geq \nu$  by a hypothesis of the lemma, we have

$$\begin{aligned} \mathbb{P} \{ Z_1 \leq t \mid Z_2 \} &\leq \max_{X_2, \dots, X_n} \mathbb{P} \left\{ \left| X_1 + \sum_{j=2}^n a_j X_j \right| \leq t/\nu \mid X_2, \dots, X_n \right\} \\ &\leq \max_{u \in \mathbb{R}} \mathbb{P} \{ |X_1 - u| \leq t/\nu \} \leq Kt/\nu. \end{aligned}$$

The last inequality follows since the density of  $X_1$  is bounded by  $K$ . This verifies hypothesis (i) of Corollary 6.3 with  $K_1 = K/\nu$ . Hypothesis (ii) follows immediately from (6.2), with  $K_2 = MK/\nu$ . If  $M \geq 1/c =: C_0$  then  $K_1 \leq cK_2$  as required in Corollary 6.3. It yields

$$\mathbb{P} \left\{ \sqrt{Z_1^2 + Z_2^2} \leq t\sqrt{d} \right\} \leq (MKt/\nu)^d \quad \text{for all } t \geq 0.$$

This completes the proof.  $\square$

## 7. THEOREM 1.1 IN HIGHER DIMENSIONS: COMPLETION OF THE PROOF

Now we are ready to prove Theorem 1.1. Replacing  $X_j$  with  $KX_j$  we can assume that  $K = 1$ . By Proposition 2.2, it suffices to bound the concentration function as follows:

$$\mathcal{L}(PX, t\sqrt{d}) \leq (C_1 t)^d, \quad t \geq 0,$$

where  $C_1$  is an absolute constant. Translating  $X$  if necessary, we can reduce the problem to showing that

$$\mathbb{P} \left\{ \|PX\|_2 \leq t\sqrt{d} \right\} \leq (C_1 t)^d, \quad t \geq 0. \quad (7.1)$$

We will prove this by induction on  $d$ . Choose  $C_1$  a sufficiently large constant, depending on the value of the constants denoted by  $C$  in Theorem 4.1, Proposition 2.2, Proposition 5.1, and the constant denoted by  $C_0$  in Proposition 6.4.

For the base of the induction, inequality (7.1) for  $d = 1$  follows from Theorem 4.1.

Assume that the statement (7.1) holds in dimension  $d - 1 \in \mathbb{N}$ , so one has

$$\mathbb{P} \left\{ \|QX\|_2 \leq t\sqrt{d-1} \right\} \leq (C_1 t)^{d-1}, \quad t \geq 0 \quad (7.2)$$

for the projection  $Q$  onto any  $(d-1)$ -dimensional subspace of  $\mathbb{R}^n$ . We would like to make an induction step, i.e. prove (7.1) in dimension  $d$ .

If  $\|Pe_i\|_2 < 1/2$  for all  $i \in [n]$ , then (7.1) follows from Proposition 5.1 together with Proposition 2.2. Alternatively, if there exists  $i \in [n]$  such that  $\|Pe_i\|_2 \geq 1/2$ , we can apply Proposition 6.4 with  $M = C_1/2$ . Moreover, since the rank of  $Q$  is  $d-1$ , the assumption (6.2) is also satisfied due to the induction hypothesis (7.2) and the choice of  $M$ . Then an application of Proposition 6.4 yields (7.1). This completes the proof of Theorem 1.1.  $\square$

## 8. THEOREM 1.5. CONCENTRATION OF ANISOTROPIC DISTRIBUTIONS.

In this section we prove a more precise version of Theorem 1.5 for random vectors of the form  $AX$ , where  $A$  is a fixed  $m \times n$  matrix.

The singular values of  $A$  arranged in a non-increasing order are denoted by  $s_j(A)$ ,  $j = 1, \dots, m \wedge n$ . To simplify the notation, we set  $s_j(A) = 0$  for  $j > m \wedge n$ , and we will do the same for singular vectors of  $A$ .

The definition of the *stable rank* of  $A$  from (1.2) reads as

$$r(A) = \left\lfloor \frac{\|A\|_{\text{HS}}^2}{\|A\|^2} \right\rfloor = \left\lfloor \frac{\sum_{j=1}^{\infty} s_j(A)^2}{s_1(A)^2} \right\rfloor.$$

To emphasize the difference between the essential rank and the rank, we set

$$\delta(A) = \frac{\sum_{j=r(A)+1}^{\infty} s_j(A)^2}{\sum_{j=1}^{\infty} s_j(A)^2}. \quad (8.1)$$

Thus  $0 \leq \delta(A) \leq 1$ . Note that the numerator in (8.1) is the square of the distance from  $A$  to the set of matrices of rank at most  $r(A)$  in the Hilbert-Schmidt metric, while the denominator equals  $\|A\|_{\text{HS}}^2$ . In particular,  $\delta(A) = 0$  if and only if  $A$  is an orthogonal projection up to an isometry; in this case  $r(A) = \text{rank}(A)$ .

**Theorem 8.1.** *Consider a random vector  $X = (X_1, \dots, X_n)$  where  $X_i$  are real-valued independent random variables. Let  $t, p \geq 0$  be such that*

$$\mathcal{L}(X_i, t) \leq p \quad \text{for all } i = 1, \dots, n.$$

*Then for every  $m \times n$  matrix  $A$  and for every  $M \geq 1$  we have*

$$\mathcal{L}(AX, Mt\|A\|_{\text{HS}}) \leq \left( CMp/\sqrt{\delta(A)} \right)^{r(A)}. \quad (8.2)$$

*provided  $\delta(A) > 0$ . Moreover, if  $\delta(A) < 0.4$ , then*

$$\mathcal{L}(AX, Mt\|A\|_{\text{HS}}) \leq (CMp)^{r_0(A)}, \quad (8.3)$$

*where  $r_0(A) = \lceil (1 - 2\delta(A))r(A) \rceil$ .*

*Remark 8.2 (Orthogonal projections).* For orthogonal projections we have  $\delta(A) = 0$ ,  $r_0(A) = r(A) = \text{rank}(A)$ , so the second part of Theorem 8.1 recovers Corollary 1.4.

*Remark 8.3 (Stability).* Note the  $\lceil \cdot \rceil$  instead of  $\lfloor \cdot \rfloor$  in the definition of  $r_0(A)$ . This offers some stability of the concentration function. Indeed, a small perturbation of a  $d$ -dimensional orthogonal projection will not change the exponent  $r_0(A) = r(A) = d$  in the small ball probability.

*Remark 8.4* (Flexible scaling). The parameter  $M$  offers some flexible scaling, which may be useful in applications. For example, knowing that  $\mathcal{L}(X_i, t)$  are all small, Theorem 8.1 allows one to bound  $\mathcal{L}(AX, 10t\|A\|_{\text{HS}})$  rather than just  $\mathcal{L}(AX, t\|A\|_{\text{HS}})$ . Note that such result would not trivially follow by applying Remark 2.3.

*Proof of Theorem 8.1.* We will first prove an inequality that is more general than (8.2). Denote

$$S_r(A)^2 = \sum_{j=r+1}^{\infty} s_j(A)^2, \quad r = 0, 1, 2, \dots$$

Then, for every  $r$ , we claim that

$$\mathcal{L}(AX, MtS_r(A)) \leq (CMp)^r. \quad (8.4)$$

This inequality would imply (8.2) by rescaling, since  $S_r(A) = \sqrt{\delta(A)}\|A\|_{\text{HS}}$ .

Before we prove (8.4), let us make some helpful reductions. First, by replacing  $X$  with  $X/t$  we can assume that  $t = 1$ . We can also assume that the vector  $u$  appearing in the definition (1.1) of the concentration function  $\mathcal{L}(AX, MtS_r)$  equals zero; this is obvious by first projecting  $u$  onto the image of  $A$  and then appropriately translating  $X$ . With these reductions, the claim (8.4) becomes

$$\mathbb{P} \{ \|AX\|_2 \leq MS_r(A) \} \leq (CMp)^r. \quad (8.5)$$

Let  $A = \sum_{j=1}^{\infty} s_j(A)u_jv_j^{\top}$  be the singular value decomposition of  $A$ . For  $l = 0, 1, 2, \dots$ , consider the spectral projections  $P_l$  defined as

$$P_0 = \sum_{j=1}^r v_jv_j^{\top} \quad \text{and} \quad P_l = \sum_{j=2^{l-1}r+1}^{2^l r} v_jv_j^{\top}, \quad l = 1, 2, \dots$$

Note that  $\text{rank}(P_0) = r$  and  $\text{rank}(P_l) = 2^{l-1}r$  for  $l = 1, 2, \dots$

We shall bound  $\|AX\|_2$  below and  $S_r(A)$  above and then compare the two estimates. First, using the monotonicity of the singular values, we have

$$\|AX\|_2^2 = \sum_{j=1}^{\infty} s_j(A)^2 \langle X, v_j \rangle^2 \geq \sum_{l=0}^{\infty} s_{2^l r}(A)^2 \|P_l X\|_2^2.$$

Next, again by monotonicity,

$$S_r(A)^2 \leq \sum_{l=0}^{\infty} 2^{lr} \cdot s_{2^l r}(A)^2.$$

Comparing these two estimates term by term, we obtain

$$\mathbb{P} \{ \|AX\|_2 < MS_r(A) \} \leq \sum_{l=0}^{\infty} \mathbb{P} \left\{ \|P_l X\|_2^2 < M^2 2^{lr} \right\}. \quad (8.6)$$

Applying Corollary 1.4 (see Remark 2.3) and noting that  $2^l r \leq 2 \text{rank}(P_l)$ , we find that

$$\mathbb{P} \left\{ \|P_l X\|_2^2 < M^2 2^{lr} \right\} \leq (C_0 M p)^{2^l r}, \quad l = 0, 1, 2, \dots \quad (8.7)$$

where  $C_0$  is an absolute constant. We will shortly conclude that (8.5) holds with  $C = 10C_0$ . Without loss of generality we can assume that  $CMp \leq 1$ , so  $C_0 M p \leq$

1/10. Thus upon substituting (8.7) into (8.6) we obtain a convergent series whose sum is bounded by  $(10C_0Mp)^r$ . This proves (8.5).

We now turn to proving (8.3). By replacing  $X$  with  $X/t$  and  $A$  with  $A/\|A\|$  we can assume that  $t = 1$  and  $\|A\| = 1$ , and reduce our task to showing that

$$\mathbb{P} \{ \|AX\|_2 \leq M \|A\|_{\text{HS}} \} \leq (CMp)^{r_0(A)}. \quad (8.8)$$

For shortness, denote  $r = r(A) = \lfloor \|A\|_{\text{HS}}^2 \rfloor$  and  $\delta = \delta(A)$ . Set  $k = \lfloor (1 - 2\delta)r \rfloor$ . We claim that

$$s_{k+1}(A) \geq \frac{1}{2}. \quad (8.9)$$

Assume the contrary. By definition of  $\delta$  and  $r$ , we have

$$\sum_{j=1}^r s_j(A)^2 = (1 - \delta) \|A\|_{\text{HS}}^2 \geq (1 - \delta)r. \quad (8.10)$$

On the other hand, by our assumption and monotonicity,  $s_j(A)$  are bounded by  $\|A\| = 1$  for all  $j$ , and by  $1/2$  for  $j \geq k + 1$ . Thus

$$\sum_{j=1}^r s_j(A)^2 \leq \sum_{j=1}^k 1^2 + \sum_{j=k+1}^r \left(\frac{1}{2}\right)^2 = k + (r - k)\frac{1}{4}. \quad (8.11)$$

Since the expression in the right hand side increases in  $k$  and  $k \leq (1 - 2\delta)r$ , we can further bound the sum in (8.11) by  $(1 - 2\delta)r + 2\delta r \cdot \frac{1}{4}$ . But this is smaller than  $(1 - \delta)r$ , the lower bound for the same sum in (8.10). This contradiction establishes our claim (8.9).

Similarly to the first part of the proof, we shall bound  $\|AX\|_2$  below and  $S_r(A)$  above and then compare the two estimates. For the lower bound, we consider the spectral projection

$$Q_{k+1} = \sum_{j=1}^{k+1} v_j v_j^\top$$

and using (8.9), we bound

$$\|AX\|_2^2 \geq s_{k+1}(A)^2 \|Q_{k+1}X\|_2^2 \geq \frac{1}{4} \|Q_{k+1}X\|_2^2.$$

For the upper bound, we note that since  $\delta \leq 0.4$  by assumption and  $r = \lfloor \|A\|_{\text{HS}}^2 \rfloor \geq 1$ , we have

$$\|A\|_{\text{HS}}^2 \leq 2r \leq 10(1 - 2\delta)r \leq 10(k + 1)$$

Applying Corollary 1.4 (see Remark 2.3) and recalling that  $\text{rank}(Q_k) = k + 1$ , we conclude that

$$\mathbb{P} \{ \|AX\|_2 \leq M \|A\|_{\text{HS}} \} \leq \mathbb{P} \{ \|Q_{k+1}X\|_2^2 \leq 40M^2(k + 1) \} \leq (CMp)^{k+1}.$$

It remains to note that  $k + 1 \geq r_0(A)$  by definition. This establishes (8.8) whenever  $CMp \leq 1$ . The case when  $CMp > 1$  is trivial. Theorem 8.1 is proved.  $\square$

Theorem 8.1 implies the following more precise version of Theorem 1.5.

**Corollary 8.5.** Consider a random vector  $X = (X_1, \dots, X_n)$  where  $X_i$  are real-valued independent random variables. Let  $t, p \geq 0$  be such that

$$\mathcal{L}(X_i, t) \leq p \quad \text{for all } i = 1, \dots, n.$$

Then for every  $m \times n$  matrix  $A$ , every  $M \geq 1$  and every  $\varepsilon \in (0, 1)$  we have

$$\mathcal{L}(AX, Mt\|A\|_{\text{HS}}) \leq (C_\varepsilon Mp)^{\lceil(1-\varepsilon)r(A)\rceil}.$$

Here  $C_\varepsilon = C/\sqrt{\varepsilon}$  and  $C$  is an absolute constant.

*Proof.* The result follows from Theorem 8.1 where we apply estimate (8.2) whenever  $\delta(A) \geq \varepsilon/2$  and (8.3) otherwise.  $\square$

In the case when the densities of  $X_i$  are uniformly bounded by  $K > 0$ , Corollary 8.5 yields

$$\mathcal{L}(AX, t\|A\|_{\text{HS}}) \leq \left(\frac{CKt}{\sqrt{\varepsilon}}\right)^{\lceil(1-\varepsilon)r(A)\rceil} \quad \text{for all } t, \varepsilon > 0. \quad (8.12)$$

Applying this estimate with  $\varepsilon = \frac{1}{2r(A)}$  and  $\tau = t\sqrt{r(A)}$ , we derive a bound on the Levy concentration function, which is lossless in terms of power of the small ball radius. Such bound may be useful for estimating the negative moments of the norm of  $AX$ .

**Corollary 8.6.** Let  $X = (X_1, \dots, X_n)$  be a vector with independent random coordinates. Assume that the densities on  $X_1, \dots, X_n$  are bounded by  $K$ . Let  $A$  be an  $m \times n$  matrix. Then for any  $\tau > 0$ ,

$$\mathcal{L}(AX, \tau\|A\|) \leq (CK\tau)^{r(A)}.$$

## REFERENCES

- [1] A. Giannopoulos, S. Brazitikos, P. Valettas and B-H. Vritsiou, *Geometry of isotropic convex bodies*. Mathematical Surveys and Monographs, 196. American Mathematical Society, Providence, RI, 2014.
- [2] K. Ball, *Cube slicing in  $\mathbb{R}^n$* , Proc. Amer. Math. Soc. 97 (1986), 465–473.
- [3] K. Ball, *Volumes of sections of cubes and related problems*. In: J. Lindenstrauss and V. D. Milman. Geometric aspects of functional analysis (1987–88). Lecture Notes in Math., Vol. 1376 (1989). Springer, Berlin. pp. 251–260
- [4] K. Ball, F. Nazarov, *Little level theorem and zero-Khinchin inequality*, unpublished, 1996. <http://www.math.msu.edu/~fedja/prepr.html>
- [5] H. J. Brascamp, E. H. Lieb, *Best constants in Young’s inequality, its converse, and its generalization to more than three functions*, Advances in Math. 20 (1976), 151–173
- [6] C. G. Esseen, *On the Kolmogorov-Rogozin inequality for the concentration function*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 5 (1966), 210–216.
- [7] G. Halász, *On the distribution of additive arithmetic functions*, Acta Arith. 27 (1975), 143–152
- [8] G. Halász, *Estimates for the concentration function of combinatorial number theory and probability*, Periodica Mathematica Hungarica 8 (1977), 197–211.
- [9] A. Kolmogorov, *Sur les propriétés des fonctions de concentrations de M. P. Lévy*, Ann. Inst. H. Poincaré 16 (1958), 27–34.
- [10] P. Lévy, *Théorie de l’addition des variables aléatoires*, Gauthier-Villars, 1937.
- [11] W. V. Li, Q.-M. Shao, *Gaussian processes: inequalities, small ball probabilities and applications*. Stochastic processes: theory and methods, 533597, Handbook of Statist., 19, North-Holland, Amsterdam, 2001.
- [12] E. Lieb, M. Loss, *Analysis*. Second edition. Graduate Studies in Mathematics, 14. American Mathematical Society, Providence, RI, 2001.

- [13] G. Paouris, *Small ball probability estimates for log-concave measures*, Trans. Amer. Math. Soc. (2012), 287–308.
- [14] G. Pisier, *The volume of convex bodies and Banach space geometry*. Cambridge Tracts in Mathematics, 94. Cambridge University Press, Cambridge, 1989.
- [15] B. A. Rogozin, *On the increase of dispersion of sums of independent random variables*, Teor. Veroyatnost. i Primenen. 6 (1961), 106–108.
- [16] B. A. Rogozin, *An estimate for the maximum of the convolution of bounded densities*. (Russian) Teor. Veroyatnost. i Primenen. 32 (1987), 53–61.
- [17] M. Rudelson, R. Vershynin, *Smallest singular value of a random rectangular matrix*, Communications on Pure and Applied Mathematics 62 (2009), 1707–1739.
- [18] M. Rudelson, R. Vershynin, *Non-asymptotic theory of random matrices: extreme singular values*, Proceedings of the International Congress of Mathematicians. Volume III, 1576–1602, Hindustan Book Agency, New Delhi, 2010.
- [19] M. Rudelson, R. Vershynin, *Hanson-Wright inequality and sub-gaussian concentration*, Electronic Communications in Probability 18 (2013), 1–9.
- [20] T. Tao, V. Vu, *From the Littlewood-Offord problem to the circular law: universality of the spectral distribution of random matrices*, Bull. Amer. Math. Soc. 46 (2009), 377–396.
- [21] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices*, Compressed sensing, 210–268, Cambridge Univ. Press, Cambridge, 2012.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MICHIGAN, 530 CHURCH ST., ANN ARBOR, MI 48109, U.S.A.

*E-mail address:* {rudelson, romanv}@umich.edu