ISRAEL JOURNAL OF MATHEMATICS **212** (2016), 289–314
DOI: 10.1007/s11856-016-1287-8

THE SMALLEST SINGULAR VALUE OF RANDOM RECTANGULAR MATRICES WITH NO MOMENT ASSUMPTIONS ON ENTRIES

BY

Konstantin E. Tikhomirov

Department of Mathematics, University of Alberta
Edmonton, Alberta T6G 2G1, Canada
e-mail: ktikhomi@ualberta.ca

ABSTRACT

Let $\delta > 1$ and $\beta > 0$ be some real numbers. We prove that there are positive u, v, N_0 depending only on β and δ with the following property: for any N, n such that $N \geq \max(N_0, \delta n)$, any $N \times n$ random matrix $A = (a_{ij})$ with i.i.d. entries satisfying $\sup_{\lambda\in\mathbb{R}} \mathbb{P}\{|a_{11} - \lambda| \leq 1\} \leq 1 - \beta$ and any non-random $N \times n$ matrix B, the smallest singular value s_n of $A + B$ satisfies $\mathbb{P}\{s_n(A + B) \leq u\sqrt{N}\} \leq \exp(-vN)$. The result holds without any moment assumptions on the distribution of the entries of A.

1. Introduction

In the last years, spectral properties of random matrices with fixed dimensions (the corresponding theory is often called **non-asymptotic**) have attracted considerable attention of researchers, whose efforts have been mostly concentrated on studying distributions of the largest and the smallest singular values. For detailed information on the development of the subject, we refer the reader to surveys [12], [22].

Let $N \geq n$. Given an $N \times n$ random matrix A, we employ a usual notation $s_1(A) := \max_{y\in S^{n-1}} \|Ay\|$; $s_n(A) := \inf_{y\in S^{n-1}} \|Ay\|$, where $\|\cdot\|$ is the standard Euclidean norm in $\mathbb{R}^n$. A limiting result of Z. D. Bai and Y. Q. Yin [3] suggests that for an $N \times n$ matrix with i.i.d. mean zero entries with unit variance and

Received May 25, 2014 and in revised form November 19, 2014

a finite fourth moment, its largest and smallest singular values should "concentrate" near $\sqrt{N}+\sqrt{n}$ and $\sqrt{N}-\sqrt{n}$, respectively. In the non-asymptotic setting one is interested, in particular, in finding the weakest possible conditions on random matrices that would imply $s_1 \lesssim \sqrt{N}+\sqrt{n}$ and $s_n \gtrsim \sqrt{N}-\sqrt{n}$ with a large probability.

For a random $N\times n$ matrix A with i.i.d. mean zero subgaussian entries, an elementary application of the standard ε-net argument yields $s_1(A)\le C(\sqrt{N}+\sqrt{n})$ with an overwhelming probability. Distribution of the smallest singular value when $N\approx n$ requires a more delicate analysis. A. Litvak, A. Pajor, M. Rudelson and N. Tomczak-Jaegermann showed in [7] that if N and n satisfy $N/n\ge 1+c_1(\ln N)^{-1}$ then $\mathbb{P}\{s_n(A)\le c_2\sqrt{N}\}\le \exp(-c_3N)$, where c_1, c_3 depend only on the variance and the subgaussian moment, and c_2 on the moments and the aspect ratio N/n. The approach initiated in [7] was further developed by M. Rudelson and R. Vershynin who combined it with certain Littlewood–Offord-type theorems. In [15], Rudelson and Vershynin treated square matrices and later, in [14], rectangular matrices with an arbitrary aspect ratio and i.i.d. mean zero subgaussian entries, thereby sharpening and generalizing the result of [7]. We note that the Littlewood–Offord theory has gained an important role in the study of random matrices primarily due to T. Tao and V. Vu (see, in particular, [19]).

Various estimates for the extremal singular values were obtained when studying the problem of approximating the covariance matrix of a random vector by the empirical covariance matrix. Answering a question of R. Kannan, L. Lovász and M. Simonovits, the authors of [1] treated log-concave random vectors. Later, the log-concavity was replaced by weaker assumptions (see, in particular, [2], [18], [9], [4]).

Recently, it has become apparent that different conditions are required to bound the largest and the smallest singular value, and these two questions should be handled separately. One of the results proved by N. Srivastava and R. Vershynin in [18] provides a lower estimate for the second moment of $s_n(A)$, where A is an $N\times n$ matrix with independent isotropic rows satisfying a $(2+\varepsilon)$-moment condition and certain assumptions on the aspect ratio N/n. It is important to note that the conditions imposed on A are too weak to imply the "usual" upper bound $s_1(A)\lesssim\sqrt{N}$ with a large probability [8]. This result of [18] was strengthened by V. Koltchinskii and S. Mendelson in [5] under similar assumptions on the matrix. Another theorem of [5] states the following: given an n-dimensional isotropic random vector X satifying $\inf_{y\in S^{n-1}} \mathbb{P}\{|\langle X, y\rangle| \geq \alpha\} \geq \beta$ for some $\alpha, \beta > 0$, there are $C_1, c_2, c_3 > 0$ depending only on α, β such that for $N \geq C_1 n$ and the $N \times n$ random matrix A with i.i.d. rows distributed like X, one has

$$\mathbb{P}\{s_n(A) \geq c_2\sqrt{N}\} \geq 1 - \exp(-c_3 N).$$

We note that a closely related question of bounding random quadratic forms from below was considered by R. I. Oliveira in [10].

The isotropy of a random vector or, more generally, boundedness of variances of its coordinates is quite a natural assumption which appears as part of requirements on rows of a matrix in all of the aforementioned papers. However, for a deeper understanding of non-asymptotic characteristics of random matrices, an important question is whether any moment assumptions on entries are really necessary in order to get satisfactory lower estimates for the smallest singular value.

Unlike in [18] and [5] where the matrix entries within a given row are not necessarily independent, in our paper we consider the classical setting when a rectangular matrix has i.i.d. entries. However, in contrast with all the mentioned results, the lower estimate for the smallest singular value that we prove does not use any moment assumptions; the only requirement is that the distribution of the entries satisfies a "spreading" condition given in terms of the Lévy concentration function. Moreover, compared to [18] and [5], we significantly relax the assumptions on the aspect ratio of the matrix.

Given a real random variable ξ, **the concentration function** of ξ is defined as

$$\mathcal{Q}(\xi, \alpha) = \sup_{\lambda\in\mathbb{R}} \mathbb{P}\{|\xi - \lambda| \leq \alpha\}, \quad \alpha \geq 0.$$

The notion of the concentration function was introduced by P. Lévy [6] in the context of studying distributions of sums of random variables. Note that for a random variable ξ with zero median satisfying $\mathbb{E}|\xi|^p \geq m$ and $\mathbb{E}|\xi|^q \leq M$ for some $0 < p < q$ and $m, M > 0$, we have $\mathcal{Q}(\xi, \alpha) \leq 1 - \beta$ for some $\alpha, \beta > 0$ depending only on p, q, m, M. At the same time, the condition $\mathcal{Q}(\xi, \alpha) \leq 1 - \beta$ for some $\alpha, \beta > 0$ does not imply any upper bounds on positive moments of ξ.

The main result of our paper is the following theorem:

Theorem 1: *For any real $\beta > 0$ and $\delta > 1$ there are $u, v > 0$ and $N_0 \in \mathbb{N}$ depending only on β and δ with the following property: Let $N, n \in \mathbb{N}$ satisfy*

$N \geq \max(N_0, \delta n)$; $A = (a_{ij})$ be an $N \times n$ random matrix with i.i.d. entries, such that for some $\alpha > 0$ the concentration function of the entries satisfies

$$\mathcal{Q}(a_{11}, \alpha) \leq 1 - \beta. \tag{1}$$

Then for any non-random $N \times n$ matrix B we have

$$\mathbb{P}\{s_n(A+B) \leq \alpha u\sqrt{N}\} \leq \exp(-vN). \tag{2}$$

Adding the non-random component B in the theorem does not increase complexity of the proof; on the other hand, it demonstrates "shift-invariance" of the lower estimate. Note that the problem of estimating the smallest singular value of non-random shifts of square matrices is important in the analysis of algorithms [16], [17], [20], [21].

It is easy to see that a restriction of type (1) is necessary for (2) to hold. Indeed, suppose that for some $N \times n$ matrix A with i.i.d. entries and some numbers $u, v, \alpha > 0$, (2) is true whenever $B = \lambda I$, $\lambda \in \mathbb{R}$. Then, obviously,

$$\mathbb{P}\Big\{\sum_{i=1}^{N}(a_{i1} - \lambda)^2 \leq \alpha^2 u^2 N\Big\} \leq \exp(-vN), \quad \lambda \in \mathbb{R},$$

implying $\mathcal{Q}(a_{11}, \alpha u) = \sup_{\lambda \in \mathbb{R}} \mathbb{P}\{|a_{11} - \lambda| \leq \alpha u\} \leq \exp(-v)$.

Our proof of Theorem 1 is based on two key elements: on a modification of a standard ε-net argument for matrices (Proposition 3) and on estimates of the distance between a random vector and a fixed linear subspace that follow from a result of [13] (Theorem 4 and Corollary 6 of our paper). Our method is similar in many aspects to the approach developed in [7] and later in [14], [15]. In particular, as in the mentioned papers, we decompose the unit sphere S^{n-1} into several subsets which are studied separately from one another. On the other hand, our modification of the ε-net argument and its technical realization in regard to splitting a random matrix into "regular" and "non-regular" parts are apparently new.

We will discuss the main idea of the proof more concretely and in more detail at the end of the next section, after we define notation and state the modified ε-net argument.

2. Preliminaries

Throughout the text, $(\Omega, \Sigma, \mathbb{P})$ denotes a probability space. Given a vector $x \in \mathbb{R}^N$, $\|x\|$ is the standard Euclidean norm and $\|x\|_\infty$ is the ℓ_∞^N-norm of x.

By S^{N-1} (respectively, B_2^N) we denote the Euclidean unit sphere (respectively, the closed unit ball) in $\mathbb{R}^N$. Given a set $K \subset \mathbb{R}^N$ and a vector x, the Euclidean distance between K and x is $\mathrm{d}(x, K) = \inf\{\|x - y\| : y \in K\}$. We use the same notation for the distance between two subsets of $\mathbb{R}^N$, i.e.,

$$\mathrm{d}(K_1, K_2) = \inf\{\|x - y\| : x \in K_1, y \in K_2\}.$$

We will sometimes use the standard identification of $N \times n$ matrices and linear operators from $\mathbb{R}^n$ to $\mathbb{R}^N$. In particular, for an $N \times n$ matrix D, by $\|D\|$ we mean the operator norm of D treated as the linear operator $D : \ell_2^n \to \ell_2^N$. For a set $K \subset \mathbb{R}^n$, $D(K)$ is the image of K in $\mathbb{R}^N$ under the action of D. For an $N \times n$ matrix D, $\mathrm{col}_j(D)$ is the j-th column of D and $\mathrm{span} D$ is the linear span of columns of D in $\mathbb{R}^N$. The $N \times n$ matrix of ones is denoted by $\mathbf{1_{N\times n}}$. For a linear subspace $E \subset \mathbb{R}^n$, $E^\perp$ is the orthogonal complement of E in $\mathbb{R}^n$ and $\mathrm{Proj}_E : \mathbb{R}^n \to \mathbb{R}^n$ is the orgothogonal projection onto E. In the special case when E is the linear span of a subset $\{e_j\}_{j \in J}$ ($J \subset \{1, 2, \ldots, n\}$) of the standard unit basis in $\mathbb{R}^n$, we will often write $x\chi_J$ in place of $\mathrm{Proj}_E(x)$.

In the paper, we define many universal constants and functions that are frequently referred to later in the text. For convenience, we add to the name of every such constant or function a subscript indicating the statement where it was defined. For example, C_{12} is the universal constant from Lemma 12, etc.

Let K be a subset of $\mathbb{R}^n$ and let $\varepsilon \in (0, 1]$. A subset $\mathcal{N} \subset K$ is called an **ε-net** for K if for any $y \in K$ there is $y' \in \mathcal{N}$ with $\|y - y'\| \le \varepsilon$. We will use a well-known fact that any subset $K \subset B_2^n$ admits an ε-net $\mathcal{N}$ for K with cardinality $|\mathcal{N}| \le (3/\varepsilon)^n$.

Given an ε-net $\mathcal{N}$ for S^{n-1}, the matrix $A+B$ from Theorem 1 trivially satisfies $s_n(A+B) \ge \min_{y' \in \mathcal{N}} \|Ay' + By'\| - \varepsilon\|A+B\|$. This standard ε-net argument is not applicable in our setting as $A+B$ may have a very large norm with a large probability. A modification of the method in such a way that $\|A + B\|$ does not participate in the estimate for $s_n(A + B)$ is an important element of our proof. In this section we provide a "non-probabilistic" form of the argument. Given a non-random $N \times n$ matrix D, we shall represent it as a sum of two matrices D_1 and D_2; then we are able to estimate $s_n(D)$ from below in terms of the norm $\|D_1\|$ of the "regular part" of the matrix D and distances between certain vectors and subspaces in $\mathbb{R}^N$ (determined by matrices D_1 and D_2). We start with a simpler version of the argument:

Lemma 2: *Let $N, n \in \mathbb{N}$, $h, \varepsilon > 0$ and let D_1, D_2, D be $N \times n$ (non-random) matrices with $D = D_1 + D_2$. Further, let $\mathcal{N}$ be an ε-net on S^{n-1} such that for any $y' \in \mathcal{N}$ we have*

$$\mathrm{d}(D_1 y', \mathrm{span} D_2) \geq h.$$

Then

$$s_n(D) \geq \inf_{y \in S^{n-1}} \mathrm{d}(D_1 y, \mathrm{span} D_2) \geq h - \varepsilon \|D_1\|.$$

Proof. Choose any $y \in S^{n-1}$ and $y' \in \mathcal{N}$ such that $\|y - y'\| \leq \varepsilon$. Then

$$\|Dy\| = \|D_1 y + D_2 y\| \geq \mathrm{d}(D_1 y, \mathrm{span} D_2) \geq \mathrm{d}(D_1 y', \mathrm{span} D_2) - \varepsilon \|D_1\| \geq h - \varepsilon \|D_1\|.$$

By taking the infimum over all $y \in S^{n-1}$, we obtain the result. ∎

Note that Lemma 2 cannot be used to handle matrices with the aspect ratio less than 2. Indeed, the lower estimate $s_n(D) \geq \inf_{y \in S^{n-1}} \mathrm{d}(D_1 y, \mathrm{span} D_2)$ is non-trivial only if $\mathrm{span} D_1 \cap \mathrm{span} D_2 = 0$, which is not true when $N < 2n$ and both D_1 and D_2 have full rank. The following strengthening of Lemma 2 resolves the problem:

Proposition 3: *Let $N, n \in \mathbb{N}$, $S \subset S^{n-1}$ and let D_1, D_2, D be $N \times n$ (non-random) matrices with $D = D_1 + D_2$. Further, suppose that numbers $h, \varepsilon > 0$, a subset $\mathcal{N} \subset \mathbb{R}^n$ and a collection of linear subspaces $\{E_{y'} \subset \mathbb{R}^n : y' \in \mathcal{N}\}$ satisfy the following three conditions:*

(a) *$y' \in E_{y'}$ for all $y' \in \mathcal{N}$;*

(b) *for any $y' \in \mathcal{N}$ we have*

$$\mathrm{d}(D_1 y', D(E_{y'}^{\perp}) + D_2(E_{y'})) \geq h; \tag{3}$$

(c) *for any $y \in S$ there is $y' \in \mathcal{N}$ such that $\|\mathrm{Proj}_{E_{y'}}(y) - y'\| \leq \varepsilon$.*

Then

$$\inf_{y \in S} \|Dy\| \geq h - \varepsilon \|D_1\|.$$

Proof. Take any $y \in S$ and let $y' \in \mathcal{N}$ be such that $\|\mathrm{Proj}_{E_{y'}}(y) - y'\| \leq \varepsilon$. Then

$$\begin{aligned}
\|Dy\| &= \|D_1(\mathrm{Proj}_{E_{y'}}(y)) + (D(\mathrm{Proj}_{E_{y'}^{\perp}}(y)) + D_2(\mathrm{Proj}_{E_{y'}}(y)))\| \\
&\geq \mathrm{d}(D_1(\mathrm{Proj}_{E_{y'}}(y)), D(E_{y'}^{\perp}) + D_2(E_{y'})) \\
&\geq \mathrm{d}(D_1 y', D(E_{y'}^{\perp}) + D_2(E_{y'})) - \varepsilon \|D_1\| \\
&\geq h - \varepsilon \|D_1\|.
\end{aligned}$$

Taking the infimum over S, we get the result. ∎

To apply Proposition 3 we need an estimate for the distance between a random vector in $\mathbb{R}^N$ with independent coordinates and a fixed linear subspace. For any random vector X in $\mathbb{R}^N$ define the concentration function of X by

$$\mathcal{Q}(X,h) = \sup_{\lambda\in\mathbb{R}^N} \mathbb{P}\{\|X-\lambda\| \le h\}, \quad h \ge 0.$$

Note that for $N = 1$ the above definition is consistent with that given in the introduction. The following result is proved by M. Rudelson and R. Vershynin in [13]:

Theorem 4 ([13]): *Let $X = (X_1, X_2, \ldots, X_m)$ be a random vector in $\mathbb{R}^m$ with independent coordinates such that*

$$\mathcal{Q}(X_i, h) \le \eta, \quad i = 1, 2, \ldots, m$$

for some $h > 0, \eta \in (0, 1)$. Then for any $d \in \{1, 2, \ldots, m\}$ and any d-dimensional non-random subspace $E \subset \mathbb{R}^m$ we have

$$\mathcal{Q}(\mathrm{Proj}_E X, h\sqrt{d}) \le (C_4\eta)^d,$$

where $C_4 > 0$ is a (sufficiently large) universal constant.

This theorem gives a non-trivial estimate for the concentration only for η sufficiently close to zero. Below, we provide an elementary extension of this result covering the case of "more concentrated" coordinates. First, let us recall a theorem of B. Rogozin:

Theorem 5 ([11]): *Let $k\in\mathbb{N}$, $\xi_1, \xi_2, \ldots, \xi_k$ be independent random variables and let $h_1, h_2, \ldots, h_k > 0$ be some real numbers. Then for any $h \ge \max_{j=1,2,\ldots,k} h_j$,*

$$\mathcal{Q}\biggl(\sum_{j=1}^k \xi_j, h\biggr) \le C_5 h \biggl(\sum_{j=1}^k (1-\mathcal{Q}(\xi_j, h_j))h_j^2\biggr)^{-1/2},$$

where $C_5 > 0$ is a universal constant.

Now, an easy application of Theorems 4 and 5 gives

Corollary 6: *Let $X = (X_1, X_2, \ldots, X_m)$ be a random vector with independent coordinates such that*

$$\mathcal{Q}(X_i, h) \le 1-\tau, \quad i = 1, 2, \ldots, m$$

for some $h > 0, \tau \in (0, 1)$. Then for any $d \in \{1, 2, \ldots, m\}$, $\ell \in \mathbb{N}$ and any d-dimensional non-random subspace $E \subset \mathbb{R}^m$ the concentration function of

$\mathrm{Proj}_E X$ *satisfies*

$$\mathcal{Q}(\mathrm{Proj}_E X, h\sqrt{d}/\ell) \le (C_4 C_5/\sqrt{\ell\tau})^{d/\ell}.$$

Proof. Let $X^1, X^2, \ldots, X^\ell$ be independent copies of X and

$$S = (S_1, S_2, \ldots, S_m) = \sum_{j=1}^{\ell} X^j.$$

In view of the condition on the coordinates of X and Theorem 5, we obtain

$$\mathcal{Q}(S_i, h) \le C_5(\ell(1 - \mathcal{Q}(X_i, h)))^{-1/2} \le \frac{C_5}{\sqrt{\ell\tau}}, \quad i = 1, 2, \ldots, m.$$

Then Theorem 4 gives

$$\mathcal{Q}(\mathrm{Proj}_E S, h\sqrt{d}) \le (C_4 C_5/\sqrt{\ell\tau})^{d}.$$

On the other hand, the definition of S together with the triangle inequality implies that

$$\mathcal{Q}(\mathrm{Proj}_E X, h\sqrt{d}/\ell)^\ell \le \mathcal{Q}(\mathrm{Proj}_E S, h\sqrt{d}),$$

and the proof is complete. ∎

Remark 1: Note that for any non-zero τ we can choose $\ell \in \mathbb{N}$ such that the upper estimate for the concentration function provided by Corollary 6 is non-trivial (strictly less than 1). In fact, a slightly weaker version of Corollary 6 still sufficient for our purposes could be proved using the original result of P. Lévy from [6] instead of Theorem 5.

As an immediate application of Corollary 6, we prove a statement about peaky vectors. We call a vector $y \in S^{n-1}$ θ-**peaky** for some $\theta > 0$ if $\|y\|_\infty \ge \theta$. The set of all θ-peaky unit vectors in $\mathbb{R}^n$ shall be denoted by $S_p^{n-1}(\theta)$.

Proposition 7 (Peaky vectors): *Let $\delta > 1$ and let $n, N \in \mathbb{N}$ satisfy $N \ge \delta n$. Further, assume we are given $\theta, \gamma > 0$ and let $U = (u_{ij})$ be an $N \times n$ random matrix with independent entries (not necessarily identically distributed), each entry u_{ij} satisfying*

$$\mathcal{Q}(u_{ij}, 1) \le 1 - \gamma.$$

Then

$$\mathbb{P}\{\inf_{y \in S_p^{n-1}(\theta)} \|Uy\| \le h_7 \theta \sqrt{N}\} \le n \exp(-w_7 N),$$

where the $h_7, w_7 > 0$ depend only on γ and δ.

Proof. By Corollary 6, for $d = N - n + 1$, any $\ell \in \mathbb{N}$ and any fixed $(n-1)$-dimensional subspace $F \subset \mathbb{R}^N$ we have

$$\begin{aligned}\mathbb{P}\{\mathrm{d}(\mathrm{col}_j(U), F) \le \sqrt{d}/\ell\} &\le \mathcal{Q}(\mathrm{Proj}_{F^\perp}(\mathrm{col}_j(U)), \sqrt{d}/\ell)\\ &\le (C_4 C_5/\sqrt{\ell\gamma})^{d/\ell}, \quad j = 1, 2, \dots, n.\end{aligned}$$

Take $\ell := \lceil 4C_4^2 C_5^2/\gamma \rceil$. Since for each $j = 1, 2, \dots, n$, $\mathrm{col}_j(U)$ is independent of the span of the other columns of U, from the above estimate we obtain

$$\mathbb{P}\{\mathrm{d}(\mathrm{col}_j(U), \mathrm{span}\{\mathrm{col}_k(U)\}_{k\neq j}) \le h\sqrt{d}\} \le \exp(-wd), \quad j = 1, 2, \dots, n$$

for some $h, w > 0$ depending only on γ. Let

$$\mathcal{E} = \{\omega \in \Omega : \mathrm{d}(\mathrm{col}_j(U(\omega)), \mathrm{span}\{\mathrm{col}_k(U(\omega))\}_{k\neq j}) > h\sqrt{d} \text{ for all } j = 1, 2, \dots, n\}.$$

Then $\mathbb{P}(\mathcal{E}) \ge 1 - n\exp(-wd)$. Take arbitrary $\omega \in \mathcal{E}$. For any $y = (y_1, y_2, \dots, y_n)$ in $S_p^{n-1}(\theta)$ there is $j = j(y)$ such that $|y_j| \ge \theta$, hence

$$\begin{aligned}\|U(\omega)y\| &= \|U(\omega)(y_j e_j) + U(\omega)(y - y_j e_j)\|\\ &\ge \theta \mathrm{d}(\mathrm{col}_j(U(\omega)), \mathrm{span}\{\mathrm{col}_k(U(\omega))\}_{k\neq j})\\ &> h\theta\sqrt{d}.\end{aligned}$$

Thus,

$$\mathbb{P}\{\inf_{y\in S_p^{n-1}(\theta)} \|Uy\| \le h\theta\sqrt{d}\} \le n\exp(-wd),$$

and the statement follows. ∎

Next, we introduce two notions important for us that will be used throughout the rest of the text. For any number $s \in \mathbb{R}$ and any Borel subset $H \subset \mathbb{R}$, define the ***H*-part of *s*** as

$$s_H = \begin{cases} s, & \text{if } s \in H,\\ 0, & \text{otherwise.}\end{cases}$$

The "complementary" $\mathbb{R}\setminus H$-part of s will be denoted by $s_{\overline{H}}$. Obviously, $s = s_H + s_{\overline{H}}$. The name and the notation resemble the positive and negative part of a real number; in fact $s_+ = s_H$ for $H = [0, \infty)$. For a real-valued random variable ξ we define the H-part of ξ pointwise: $\xi_H(\omega) = \xi(\omega)_H$ for all $\omega \in \Omega$. When a variable has a subscript, we will use parentheses to separate the subscript from the H-part notation, for example $(\xi_1)_H$ is the H-part of a random variable ξ_1. Given a matrix $A = (a_{ij})$, its H-part A_H is defined entry-wise, i.e., $(A_H)_{ij} = (a_{ij})_H$ for all admissible i, j.

For any $N \times n$ matrices M, M' (whether random or not), a Borel set $H \subset \mathbb{R}$ and a linear subspace $E \subset \mathbb{R}^n$ let

$$V_{M,M'}(H,E) := (M + M')(E^\perp) + (M_{\overline{H}} + M')(E).$$

Note that $V_{M,M'}(H,E)$ is a linear subspace of $\mathbb{R}^N$ of dimension at most n. When the matrices M, M' are clear from the context, we shall write $V(H,E)$ in place of $V_{M,M'}(H,E)$. When one or both matrices M, M' are random, $V_{M,M'}(H,E)$ is a **random subspace in** $\mathbb{R}^N$ of dimension at most n.

Let us conclude the section by describing the main idea of the proof of Theorem 1. Let S be a subset of S^{n-1}. As we already noted before, the main obstacle in using the standard ε-net argument to get a lower estimate for $\inf_{y\in S}\|Ay + By\|$ is the need to control the norm of the matrix $A+B$ which is not possible unless we impose strong restrictions on its entries. Proposition 3 provides a workaround: we represent $A + B$ as a sum of two random matrices, "regular" and "irregular", satisfying certain conditions, so that the lower bound for $\inf_{y\in S}\|Ay+By\|$ involves the norm of only the "regular" matrix. The splitting shall be defined with help of the above concept of H-part. Namely, for some specially chosen $\lambda \in \mathbb{R}$ and $H \subset \mathbb{R}$ we define the "regular" part as $(A - \lambda\mathbf{1_{N\times n}})_H$ and the "irregular" as $A+B-(A - \lambda\mathbf{1_{N\times n}})_H$ (which is identical to $(A - \lambda\mathbf{1_{N\times n}})_{\overline{H}} + B + \lambda\mathbf{1_{N\times n}}$). The set H shall be bounded which implies boundedness of the entries of $(A - \lambda\mathbf{1_{N\times n}})_H$. This, together with the appropriately chosen "shift" λ, allows us to easily control $\|(A - \lambda\mathbf{1_{N\times n}})_H\|$ from above. We will define H as the union of two specially constructed closed intervals on $\mathbb{R}$. The choice of H depends on the set S and may depend on the characteristics of the distribution of the entries of A (we leave this problem for the last section).

The crucial property that our set H shall satisfy is: letting $\tilde{A} = A - \lambda\mathbf{1_{N\times n}}$ and $\tilde{B} = B + \lambda\mathbf{1_{N\times n}}$, for a certain finite subset of vectors $\mathcal{N} \subset \mathbb{R}^n$ and a collection of linear subspaces $\{E_{y'} \subset \mathbb{R}^n\}_{y'\in\mathcal{N}}$ (see Proposition 3) we have

$$\inf_{y'\in\mathcal{N}} \mathrm{d}(\tilde{A}_H y', V_{\tilde{A},\tilde{B}}(H, E_{y'})) \gtrsim \sqrt{N}$$

with a large probability. This restriction on H naturally corresponds to the condition (3) in Proposition 3. In practice, we shall verify this property of H by proving that for every vector $y \in B_2^n$ satisfying certain upper bounds on $\|y\|_\infty$ and lower bounds on $\|y\|$ and for $E = \mathrm{span}\{e_j\}_{j\in\mathrm{supp}\,y}$, the distance $\mathrm{d}(\tilde{A}_H y, V_{\tilde{A},\tilde{B}}(H,E))$ is large with an overwhelming probability. This condition demands a "rich" structure from $\tilde{A}_H$; consequently, the set H cannot be very small in diameter. On the other hand, the "upper" restrictions on H are dictated by the necessity to control the norm of $\tilde{A}_H$. Thus, we have to find a balance between the two requirements.

In order to estimate the distance between the random vector $\tilde{A}_H y$ and the random subspace $V_{\tilde{A},\tilde{B}}(H,E)$, we will use Corollary 6. However, since in general $V_{\tilde{A},\tilde{B}}(H,E)$ is dependent (in a probabilistic sense) on $\tilde{A}_H y$, an immediate application of the corollary is not possible; instead, we will combine it with a conditioning argument, which is presented in the next section.

3. The distribution of $\mathrm{d}(A_H y, V_{A,B}(H,E))$

Assume that we are given $\delta > 1$, $N, n \in \mathbb{N}$ with $N \geq \delta n$, a random $N \times n$ matrix A with i.i.d. entries, a non-random $N \times n$ matrix B and a Borel subset $H \subset \mathbb{R}$ with $\mathbb{P}\{a_{11} \in H\} > 0$. The purpose of this section is to study the distribution of the distance between a random vector $A_H y$ and the random subspace

$$V_{A,B}(H,E) = (A+B)(E^\perp) + (A_{\overline{H}} + B)(E),$$

where $E = \operatorname{span}\{e_j\}_{j \in \operatorname{supp} y}$. We give sufficient conditions on A, H and y which guarantee that $\mathrm{d}(A_H y, V_{A,B}(H,E))$ is large with a large probability (Proposition 11). Note that generally $A_H y$ and $V_{A,B}(H,E)$ are dependent. In order to overcome this problem, we apply a decoupling argument.

We adopt the following notation: For any subset

$$W \subset \{1,2,\ldots,N\} \times \{1,2,\ldots,n\}$$

let

$$\Omega_W = \{\omega \in \Omega : a_{ij}(\omega) \in H \text{ for all } (i,j) \in W \text{ and } a_{ij}(\omega) \in \overline{H} \text{ for all } (i,j) \notin W\}.$$

Given an event $\mathcal{E} \subset \Omega$ with $\mathbb{P}(\mathcal{E}) > 0$, we denote by $(\mathcal{E}, \Sigma_{\mathcal{E}}, \mathbb{P}_{\mathcal{E}})$ the probability space where the σ-algebra $\Sigma_{\mathcal{E}}$ of subsets of $\mathcal{E}$ is naturally induced by the σ-algebra Σ on Ω, and $\mathbb{P}_{\mathcal{E}}$ is defined by

$$\mathbb{P}_{\mathcal{E}}(K) = \mathbb{P}(\mathcal{E})^{-1}\mathbb{P}(K) \quad (K \in \Sigma_{\mathcal{E}}).$$

Lemma 8 (Conditional independence): *Let A, B and H be as above, $y \in \mathbb{R}^n$, $E = \operatorname{span}\{e_j\}_{j \in \operatorname{supp} y}$ and let $W \subset \{1,2,\ldots,N\} \times \{1,2,\ldots,n\}$ be such that $\mathbb{P}(\Omega_W) > 0$. Then the random vector $A_H y$ in $\mathbb{R}^N$ and the random subspace $V_{A,B}(H,E) \subset \mathbb{R}^N$ are conditionally independent given event Ω_W. Moreover, the coordinates of $A_H y$ are conditionally independent given Ω_W.*

The proof of the lemma is quite straightforward, so we omit it. Lemma 8 shows that Corollary 6 can be applied to $A_H y$ and the subspace $V_{A,B}(H,E)$ "inside" each Ω_W. Hence, to give a satisfactory lower estimate for

$$\mathrm{d}(A_H y, V_{A,B}(H,E))$$

on the entire Ω, it is enough to verify that there is a subset

$$M \subset 2^{\{1,2,\dots,N\}\times\{1,2,\dots,n\}}$$

such that the $\mathbb{P}$-measure of the union of Ω_W's ($W \in M$) is close to 1 and for each $W \in M$, the restriction of the vector $A_H y$ to Ω_W has sufficiently "spread" coordinates. Of course, such a set M may exist only under certain assumptions on A, H and y. In Lemma 9, we formulate those assumptions using random variables that agree on a part of the probability space and are independent when restricted to the other part of Ω. Let us remark that, whereas the use of such variables has some advantages (in our opinion), it should not be regarded as a necessary ingredient of the proof.

Let ξ, ξ' be two random variables such that $\mathbb{P}\{\xi \in H\} > 0$. We say that ξ, ξ' are **conditionally i.i.d. given event** $\{\omega \in \Omega : \xi(\omega) \in H\}$ **and identical on** $\{\omega \in \Omega : \xi(\omega) \in \overline{H}\}$ if the following is true: setting $\mathcal{E} = \{\omega \in \Omega : \xi(\omega) \in H\}$, the restrictions of ξ, ξ' to the probability space $(\mathcal{E}, \Sigma_{\mathcal{E}}, \mathbb{P}_{\mathcal{E}})$ are i.i.d. and $\xi(\omega) = \xi'(\omega)$ for $\omega \in \Omega\setminus\mathcal{E}$. The definition implies that ξ' has the same individual distribution (on Ω) as ξ and for any Borel subsets $K, K' \subset \mathbb{R}$

$$\mathbb{P}\{(\xi,\xi') \in K \times K'\} = \frac{\mathbb{P}\{\xi \in H \cap K\}\mathbb{P}\{\xi \in H \cap K'\}}{\mathbb{P}\{\xi \in H\}} + \mathbb{P}\{\xi \in \overline{H} \cap K \cap K'\}; \tag{4}$$

in particular, $\mathbb{P}\{(\xi,\xi') \in H \times \overline{H}\} = \mathbb{P}\{(\xi,\xi') \in \overline{H} \times H\} = 0$. Note that $\xi_{\overline{H}}$ and ${\xi'}_{\overline{H}}$ are equal a.s. on Ω. It is a trivial observation that $\xi_H - {\xi'}_H$ is symmetrically distributed.

For any event $\mathcal{E} \subset \Omega$ with $\mathbb{P}(\mathcal{E}) > 0$ and any random variable ξ on Ω, let $\mathcal{Q}_{\mathcal{E}}(\xi, \cdot)$ be the concentration function of the restriction of ξ to the probability space $(\mathcal{E}, \Sigma_{\mathcal{E}}, \mathbb{P}_{\mathcal{E}})$.

Lemma 9: *Let H be a Borel subset of $\mathbb{R}$; $N \ge \delta n$ for some $\delta > 1$ and let $A = (a_{ij})$ be an $N \times n$ random matrix with i.i.d. entries and $\mathbb{P}\{a_{11} \in H\} > 0$. Further, let $A' = (a'_{ij})$ be an $N \times n$ random matrix having the same distribution as A such that 2-dimensional vectors (a_{ij}, a'_{ij}) ($1 \le i \le N$, $1 \le j \le n$) are i.i.d. and for any admissible i and j the variables a_{ij} and a'_{ij} are conditionally* *i.i.d. given event* $\{\omega \in \Omega : a_{ij}(\omega) \in H\}$ *and identical on* $\{\omega \in \Omega : a_{ij}(\omega) \in \overline{H}\}$. *Let* $y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$ *and* $s > 0$ *be such that*

$$\mathbb{P}\bigg\{\bigg|\sum_{j=1}^{n}((a_{ij})_H - (a'_{ij})_H)y_j\bigg| > s\bigg\} \geq \delta^{-1/4}, \quad i = 1, 2, \ldots, N. \tag{5}$$

Define M *as the collection of all subsets* $W \subset \{1, 2, \ldots, N\} \times \{1, 2, \ldots, n\}$ *satisfying*

$$\mathbb{P}(\Omega_W) > 0$$

and

$$\bigg|\bigg\{i \in \{1, 2, \ldots, N\} : \mathcal{Q}_{\Omega_W}\bigg(\sum_{j=1}^{n}(a_{ij})_H y_j, \frac{s}{2}\bigg) \leq 1 - \tau\bigg\}\bigg| \geq N\delta^{-1/2}$$

with $\tau = \frac{1}{2}(\delta^{-1/4} - \delta^{-1/3})$. *Then*

$$\mathbb{P}\bigg(\bigcup_{W \in M} \Omega_W\bigg) \geq 1 - \exp(-w_9 N),$$

where $w_9 > 0$ *depends only on* δ.

Proof. For each $i = 1, 2, \ldots, N$ and $J \subset \{1, 2, \ldots, n\}$ let

$$\Omega_J^i = \{\omega \in \Omega : a_{ij}(\omega) \in H \text{ for all } j \in J \text{ and } a_{ij}(\omega) \in \overline{H} \text{ for all } j \notin J\},$$

and for $i = 1, 2, \ldots, N$ define

$$L_i = \bigg\{J \subset \{1, 2, \ldots, n\} : \mathbb{P}(\Omega_J^i) > 0 \text{ and } \mathcal{Q}_{\Omega_J^i}\bigg(\sum_{j=1}^{n}(a_{ij})_H y_j, \frac{s}{2}\bigg) \leq 1 - \tau\bigg\};$$

$$\mathcal{E}_i = \bigcup_{J \in L_i} \Omega_J^i.$$

It is not difficult to see that the events $\mathcal{E}_i \subset \Omega$ $(i = 1, 2, \ldots, N)$ are independent in view of the independence of the entries of A.

Fix for a moment any $i \in \{1, 2, \ldots, N\}$. One can verify that for any $j \in \{1, 2, \ldots, n\}$ and $J \subset \{1, 2, \ldots, n\}$ the variables $(a_{ij})_H$ and $(a'_{ij})_H$ are i.i.d. given event Ω_J^i. It follows that

$$\sum_{j=1}^{n}(a_{ij})_H y_j \text{ and } \sum_{j=1}^{n}(a'_{ij})_H y_j \text{ are i.i.d. given } \Omega_J^i, \text{ for all } J \subset \{1, 2, \ldots, n\}. \tag{6}$$

Take any subset $J \subset \{1, 2, \ldots, n\}$ satisfying

$$\text{(7)} \qquad \mathbb{P}(\Omega^i_J) > 0 \quad \text{and} \quad \mathbb{P}_{\Omega^i_J}\Big\{\Big|\sum_{j=1}^n ((a_{ij})_H - (a'_{ij})_H) y_j\Big| > s\Big\} \geq 2\tau.$$

For all $\lambda \in \mathbb{R}$ we have, in view of (6),

$$\begin{aligned}
&\mathbb{P}_{\Omega^i_J}\Big\{\lambda - \frac{s}{2} \leq \sum_{j=1}^n (a_{ij})_H y_j \leq \lambda + \frac{s}{2}\Big\}^2 \\
&\quad= \mathbb{P}_{\Omega^i_J}\Big\{\lambda - \frac{s}{2} \leq \sum_{j=1}^n (a_{ij})_H y_j \leq \lambda + \frac{s}{2} \text{ and } \lambda - \frac{s}{2} \leq \sum_{j=1}^n (a'_{ij})_H y_j \leq \lambda + \frac{s}{2}\Big\} \\
&\quad\leq \mathbb{P}_{\Omega^i_J}\Big\{\Big|\sum_{j=1}^n ((a_{ij})_H - (a'_{ij})_H) y_j\Big| \leq s\Big\} \\
&\quad\leq 1 - 2\tau,
\end{aligned}$$

implying

$$\mathcal{Q}_{\Omega^i_J}\Big(\sum_{j=1}^n (a_{ij})_H y_j, \frac{s}{2}\Big) \leq \sqrt{1 - 2\tau} \leq 1 - \tau.$$

Thus, any J satisfying (7) belongs to L_i. Clearly,

$$\mathbb{P}\Big\{\Big|\sum_{j=1}^n ((a_{ij})_H - (a'_{ij})_H) y_j\Big| > s\Big\} = \sum_J \mathbb{P}_{\Omega^i_J}\Big\{\Big|\sum_{j=1}^n ((a_{ij})_H - (a'_{ij})_H) y_j\Big| > s\Big\} \mathbb{P}(\Omega^i_J),$$

where the summation is taken over $J \subset \{1, 2, \ldots, n\}$ satisfying $\mathbb{P}(\Omega^i_J) > 0$. Hence, in view of (5) and the above observations we get

$$\begin{aligned}
\delta^{-1/4} &\leq \sum_J \mathbb{P}_{\Omega^i_J}\Big\{\Big|\sum_{j=1}^n ((a_{ij})_H - (a'_{ij})_H) y_j\Big| > s\Big\} \mathbb{P}(\Omega^i_J) \\
&\leq \sum_{J \in L_i} \mathbb{P}(\Omega^i_J) + 2\tau \sum_{J \notin L_i} \mathbb{P}(\Omega^i_J) \\
&\leq 2\tau + \mathbb{P}(\mathcal{E}_i),
\end{aligned}$$

implying $\mathbb{P}(\mathcal{E}_i) \geq \delta^{-1/3}$.

We have noted that the events $\mathcal{E}_i$ ($i = 1, 2, \ldots, N$) are independent and $\mathbb{P}(\mathcal{E}_i) \geq \delta^{-1/3}$ for each i. Now, setting

$$\mathcal{E} = \{\omega \in \Omega : |\{i \in \{1, 2, \ldots, N\} : \omega \in \mathcal{E}_i\}| \geq N\delta^{-1/2}\},$$

we obtain by Bernstein's (or Hoeffding's) inequality $\mathbb{P}(\mathcal{E}) \geq 1 - \exp(-w_9 N)$, where $w_9 > 0$ depends only on δ. Finally, we will show that $\mathcal{E} \subset \bigcup_{W \in M} \Omega_W \cup \Omega^0$

for a set Ω^0 of zero probability measure. Define $\Omega^0 = \bigcup_W \Omega_W$, where the union is taken over all W such that $\mathbb{P}(\Omega_W) = 0$. Fix any $\omega \in \mathcal{E} \setminus \Omega^0$ and let $\tilde{W} \subset \{1,2,\dots,N\} \times \{1,2,\dots,n\}$ be such that $\omega \in \Omega_{\tilde{W}}$. In view of the definition of $\mathcal{E}$ and the events $\mathcal{E}_i$, there are indices $i_1 < i_2 < \cdots < i_k$ ($k \geq N\delta^{-1/2}$) such that $w \in \Omega_{J_q}^{i_q}$ for all $q = 1,2,\dots,k$, where $J_q = \{j : (i_q, j) \in \tilde{W}\}$ and

$$\mathcal{Q}_{\Omega_{J_q}^{i_q}}\left(\sum_{j=1}^{n} (a_{i_q j})_H y_j, \frac{s}{2}\right) \leq 1 - \tau, \quad q = 1,2,\dots,k.$$

Note that, in view of the independence of the entries of A and the relation between $\Omega_{\tilde{W}}$ and $\Omega_{J_q}^{i_q}$, the conditional distribution of the sum $\sum_{j=1}^{n} (a_{i_q j})_H y_j$ given event $\Omega_{\tilde{W}}$ is the same as its conditional distribution given $\Omega_{J_q}^{i_q}$. Hence,

$$\mathcal{Q}_{\Omega_{\tilde{W}}}\left(\sum_{j=1}^{n} (a_{i_q j})_H y_j, \frac{s}{2}\right) = \mathcal{Q}_{\Omega_{J_q}^{i_q}}\left(\sum_{j=1}^{n} (a_{i_q j})_H y_j, \frac{s}{2}\right) \leq 1 - \tau, \quad q = 1,2,\dots,k.$$

The last formula implies that $\tilde{W} \subset M$, so $\omega \in \bigcup_{W \in M} \Omega_W$. The proof is complete. ∎

Next, we combine the result of Lemma 9 with Corollary 6:

Lemma 10: *Let N, n, δ, H, A, A', y and s be exactly as in Lemma 9 and B be a non-random $N \times n$ matrix. Then*

$$\mathbb{P}\{\mathrm{d}(A_H y, V_{A,B}(H,E)) \leq s h_{10}\sqrt{N}\} \leq 2\exp(-w_{10}N),$$

where $E = \mathrm{span}\{e_j\}_{j \in \mathrm{supp} y}$ and $h_{10} > 0$, $w_{10} > 0$ depend only on δ.

Proof. Let M and τ be defined as in Lemma 9 and take any $W \in M$. Let

$$m = \left|\left\{i \in \{1,2,\dots,N\} : \mathcal{Q}_{\Omega_W}\left(\sum_{j=1}^{n} (a_{ij})_H y_j, \frac{s}{2}\right) \leq 1 - \tau\right\}\right|.$$

By the definition of M, we have $m \geq N\delta^{-1/2} \geq \sqrt{\delta} n$, hence, taking $d = m - n$ and $\ell = 4(C_4 C_5)^2/\tau$, by Corollary 6, for $\kappa = \delta^{-1/2} - \delta^{-1}$ and any fixed n-dimensional subspace $F \subset \mathbb{R}^N$ we obtain

$$\mathbb{P}_{\Omega_W}\{\mathrm{d}(A_H y, F) \leq \frac{s}{2\ell}\sqrt{\kappa N}\} \leq 2^{-\kappa N/\ell}.$$

Now, consider the random subspace $V_{A,B}(H,E) = (A+B)(E^{\perp}) + (A_{\overline{H}} + B)(E)$. Let us remark that $(A+B)(E^{\perp})$ is just the linear span of columns of $A+B$ whose indices do not belong to the support of y, and, similarly, $(A_{\overline{H}} + B)(E)$ is the span of those columns of $A_{\overline{H}} + B$ whose indices belong to the support of y.

By Lemma 8, $V_{A,B}(H,E)$ and the vector $A_H y$ are conditionally independent given Ω_W, hence the above estimate immediately implies

$$\mathbb{P}_{\Omega_W}\{\mathrm{d}(A_H y, V_{A,B}(H,E)) \le \frac{s}{2\ell}\sqrt{\kappa N}\} \le 2^{-\kappa N/\ell}.$$

Since the relation holds for all $W \in M$, in view of Lemma 9 we obtain

$$\begin{aligned}\mathbb{P}\{\mathrm{d}(A_H y, V_{A,B}(H,E)) \le \frac{s}{2\ell}\sqrt{\kappa N}\} &\le 2^{-\kappa N/\ell}\,\mathbb{P}\Big(\bigcup_{W\in M}\Omega_W\Big) + 1 - \mathbb{P}\Big(\bigcup_{W\in M}\Omega_W\Big)\\ &\le 2^{-\kappa N/\ell} + \exp(-w_9 N),\end{aligned}$$

and the result follows. ∎

Finally, we can prove the main result of the section:

Proposition 11: *Let $\delta > 1$, $n, N \in \mathbb{N}$, $N \ge \delta n$ and let $A = (a_{ij})$ be an $N \times n$ random matrix with i.i.d. entries and B be any non-random $N \times n$ matrix. Further, for some $d, r > 0$ let H be a Borel subset of $\mathbb{R}$ such that $H = H_1 \cup H_2$ for disjoint Borel sets H_1, H_2 with*

$$\mathrm{d}(H_1, H_2) \ge d \quad \text{and} \quad \min(\mathbb{P}\{a_{11} \in H_1\}, \mathbb{P}\{a_{11} \in H_2\}) \ge r.$$

For arbitrary $t > 0$ define

$$h_{11} = \frac{1-\delta^{-1/4}}{C_5}\sqrt{\frac{r}{8}}td$$

and let $y \in \mathbb{R}^n$ be a vector satisfying $\|y\| \ge t$, $\|y\|_\infty \le \frac{2h_{11}}{d}$ and $E = \mathrm{span}\{e_j\}_{j\in\mathrm{supp}\,y}$. Then

$$\mathbb{P}\{\mathrm{d}(A_H y, V_{A,B}(H,E)) \le h_{10}h_{11}\sqrt{N}\} \le 2\exp(-w_{10}N).$$

Proof. Let $A' = (a'_{ij})$ be an $N \times n$ random matrix having the same distribution as A such that 2-dimensional vectors (a_{ij}, a'_{ij}) $(1 \le i \le N,\ 1 \le j \le n)$ are i.i.d. and for any admissible i and j the variables a_{ij} and a'_{ij} are conditionally i.i.d. given event $\{\omega \in \Omega : a_{ij}(\omega) \in H\}$ and identical on $\{\omega \in \Omega : a_{ij}(\omega) \in \overline{H}\}$. For every $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, n$, in view of formula (4) for the joint distribution we get

$$\begin{aligned}\mathbb{P}\{|(a_{ij})_H - (a'_{ij})_H| \ge d\} &\ge \mathbb{P}\{a_{ij} \in H_1 \text{ and } a'_{ij} \in H_2\} + \mathbb{P}\{a_{ij} \in H_2 \text{ and } a'_{ij} \in H_1\}\\ &\ge r.\end{aligned}$$

Since $(a_{ij})_H - (a'_{ij})_H$ is symmetrically distributed, the above relation implies $\mathcal{Q}((a_{ij})_H - (a'_{ij})_H, \frac{d}{2}) \le 1 - \frac{r}{2}$. Clearly, $h_{11} \ge \frac{d|y_j|}{2}$ for every coordinate y_j of the vector y, hence by Theorem 5 for all $i = 1, 2, \ldots, N$ we have

$$\begin{aligned}
&\mathbb{P}\Big\{\Big|\sum_{j=1}^{n}((a_{ij})_H - (a'_{ij})_H)y_j\Big| \le h_{11}\Big\} \\
&\le \mathcal{Q}\Big(\sum_{j=1}^{n}((a_{ij})_H - (a'_{ij})_H)y_j, h_{11}\Big) \\
&\le C_5 h_{11}\Big(\frac{1}{4}\sum_{j=1}^{n}\Big(1 - \mathcal{Q}\Big(((a_{ij})_H - (a'_{ij})_H)y_j, \frac{|y_j|d}{2}\Big)\Big)(y_j d)^2\Big)^{-1/2} \\
&\le C_5 h_{11}\Big(\frac{r}{8}\sum_{j=1}^{n}(y_j d)^2\Big)^{-1/2} \\
&\le \frac{C_5 h_{11}}{td}\sqrt{\frac{8}{r}} = 1 - \delta^{-1/4}.
\end{aligned}$$

Thus, vector y satisfies condition (5) with $s := h_{11}$. Then, by Lemma 10,

$$\mathbb{P}\{\mathrm{d}(A_H y, V_{A,B}(H,E)) \le h_{10}h_{11}\sqrt{N}\} \le 2\exp(-w_{10}N). \qquad \blacksquare$$

4. Decomposition of S^{n-1} and proof of Theorem 1

Recall that in Section 2 we defined $S_p^{n-1}(\theta)$ as the set of θ-peaky vectors, that is, unit vectors in $\mathbb{R}^n$ whose ℓ_∞^n-norm is at least θ. We say that a vector $y \in S^{n-1}$ is m-**sparse** if $|\operatorname{supp} y| \le m$. Next, $y \in S^{n-1}$ is **almost** m-**sparse**, if there is a subset $J \subset \{1, 2, \ldots, n\}$ of cardinality at most m, such that $\|y\chi_J\| \ge 1/2$. The set of all almost m-sparse vectors shall be denoted by $S_a^{n-1}(m)$.

In our proof of Theorem 1, we represent S^{n-1} as the union of three subsets:

$$S^{n-1} = S_p^{n-1}(\theta) \cup (S_a^{n-1}(\sqrt{N}) \setminus S_p^{n-1}(\theta)) \cup (S^{n-1} \setminus S_a^{n-1}(\sqrt{N})),$$

where θ is a function of the parameters β and δ of the theorem. Then the smallest singular value of $A + B$ can be estimated by bounding separately $\inf_y \|Ay + By\|$ over each of the three subsets.

The reasons for such a representation of S^{n-1} are purely technical: Proposition 11 proved in the previous section handles vectors with a sufficiently small ℓ_∞^n-norm, so instead we use Proposition 7 to deal with the set $S_p^{n-1}(\theta)$. Further, the separate treatment of almost $\sqrt{N}$-sparse vectors is convenient because, on the one hand, the construction of the set H corresponding to $S_a^{n-1}(\sqrt{N})\backslash S_p^{n-1}(\theta)$ is trivial compared to $S^{n-1}\backslash S_a^{n-1}(\sqrt{N})$; on the other hand, vectors from $S^{n-1} \setminus S_a^{n-1}(\sqrt{N})$ have a useful geometric property (Lemma 16) which the almost sparse vectors generally do not possess. We note that the set $S_a^{n-1}(\sqrt{N})$ in the covering of S^{n-1} can be replaced with $S_a^{n-1}(N^\kappa)$ for any constant power $\kappa \in (0,1)$; this would only affect the constants in the final estimate.

In our representation of S^{n-1}, we follow an idea from [7], where the unit sphere was split into sets of "close to sparse" and "far from sparse" vectors. A similar splitting was also employed in [14], [15], where the terms "compressible" and "incompressible" were used instead. On the other hand, our "borderline" $\sqrt{N}$ is smaller by the order of magnitude than in the mentioned papers.

The next elementary lemma shall be used in conjunction with Proposition 3.

LEMMA 12: *There is a universal constant $C_{12} > 0$ with the following property: Let $n, m \in \mathbb{N}$ with $m \leq n$, $\varepsilon \in (0,1]$, $S \subset S^{n-1}$ and let $T \subset B_2^n$ consist of m-sparse vectors and satisfy*

$$\text{(8)} \qquad \text{for any } y \in S \text{ there is } x = x(y) \in T \text{ with } y\chi_{\operatorname{supp} x} = x.$$

Then there is a finite set $\mathcal{N} \subset T$ of cardinality at most $(\frac{C_{12}n}{\varepsilon m})^m$ such that for any $y \in S$ there is $y' = y'(y) \in \mathcal{N}$ with $\|y\chi_{\operatorname{supp} y'} - y'\| \leq \varepsilon$.

PROPOSITION 13 (Vectors from $S_a^{n-1}(\sqrt{N})$ with a small ℓ_∞^n-norm): *For any $\gamma > 0$ and $\delta > 1$ there are $N_{13} \in \mathbb{N}$ and $h_{13} > 0$ depending only on γ and δ with the following property: Let*

$$\theta_{13} = \frac{1-\delta^{-1/4}}{C_5}\sqrt{\frac{\gamma}{8}},$$

$N \geq \max(N_{13}, \delta n)$, $z \in \mathbb{R}$ and let A be an $N \times n$ random matrix with i.i.d. entries such that

$$\min(\mathbb{P}\{z - \sqrt{N} \leq a_{11} \leq z - 1\}, \mathbb{P}\{z + 1 \leq a_{11} \leq z + \sqrt{N}\}) \geq \gamma.$$

Then for the set $S = S_a^{n-1}(\sqrt{N}) \setminus S_p^{n-1}(\theta_{13})$ and any non-random $N \times n$ matrix B we have

$$\mathbb{P}\{\inf_{y \in S} \|Ay + By\| \leq h_{13}\sqrt{N}\} \leq \exp(-w_{10}N/2).$$

Proof. Fix any $\gamma > 0$ and $\delta > 1$ and define $d := 2$, $r := \gamma$, $t := \frac{1}{2}$; let h_{11} be as in Proposition 11 and $N_{13} = N_{13}(\gamma, \delta)$ be the smallest integer greater than $\frac{2}{h_{10}h_{11}}$ such that for all $N \geq N_{13}$

$$2(C_{12}N)^{3\sqrt{N}} \leq \exp(w_{10}N/2).$$

Now, take any $n \in N$ and $N \geq \max(N_{13}, \delta n)$; let z and A satisfy conditions of the proposition and B be any non-random $N \times n$ matrix. We will assume that S is non-empty. Without loss of generality, $z = 0$ (otherwise, we replace A, B with $A - z\mathbf{1}_{\mathbf{N}\times\mathbf{n}}$, $B + z\mathbf{1}_{\mathbf{N}\times\mathbf{n}}$). Define $H_1 = [-\sqrt{N}, -1]$, $H_2 = [1, \sqrt{N}]$, $H = H_1 \cup H_2$. Obviously, $\mathrm{d}(H_1, H_2) = d$ and $\min(\mathbb{P}\{a_{11} \in H_1\}, \mathbb{P}\{a_{11} \in H_2\}) \geq r$. Let $T \subset B_2^n$ be the set of $\sqrt{N}$-sparse vectors with the Euclidean norm at least $\frac{1}{2}$ and the maximal norm at most θ_{13}. Clearly, T and S satisfy (8), hence, by Lemma 12, there is a finite subset $\mathcal{N} \subset T$ of cardinality at most $(C_{12}N)^{3\sqrt{N}}$ such that for any $y \in S$ there is $y' = y'(y) \in \mathcal{N}$ with $\|y\chi_{\mathrm{supp}y'} - y'\| \leq N^{-2}$.

Let $E_{y'} = \mathrm{span}\{e_j\}_{j \in \mathrm{supp}y'}$ $(y' \in \mathcal{N})$ and define an event

$$\mathcal{E} = \{\omega \in \Omega : \mathrm{d}(A_H(\omega)y', V_{A,B}(H, E_{y'})(\omega)) > h_{10}h_{11}\sqrt{N} \text{ for all } y' \in \mathcal{N}\}.$$

In view of Proposition 11, the upper estimate for $|\mathcal{N}|$ and the definition of N_{13}, we get

$$\mathbb{P}(\mathcal{E}) \geq 1 - 2|\mathcal{N}| \exp(-w_{10}N) \geq 1 - \exp(-w_{10}N/2).$$

Take any $\omega \in \mathcal{E}$ and define $D_1 = A_H(\omega)$, $D_2 = A_{\overline{H}}(\omega) + B$, $D = D_1 + D_2$. Since all entries of D_1 are bounded by $\sqrt{N}$ by absolute value, we get $\|D_1\| \leq N^{3/2}$; next, for every $y' \in \mathcal{N}$

$$\mathrm{d}(D_1 y', D(E_{y'}^{\perp}) + D_2(E_{y'})) > h_{10}h_{11}\sqrt{N}$$

(note that $D(E_{y'}^{\perp}) + D_2(E_{y'}) = V_{A,B}(H, E_{y'})(\omega)$). Hence, by Proposition 3, we obtain

$$\inf_{y \in S} \|Dy\| > h_{10}h_{11}\sqrt{N} - N^{-1/2} \geq \frac{1}{2}h_{10}h_{11}\sqrt{N}.$$

Finally, applying the above argument to all $\omega \in \mathcal{E}$, we get the result. ∎

As we noted before, construction of the set H corresponding to

$$S^{n-1} \setminus S_a^{n-1}(\sqrt{N})$$

is not so trivial as in the case of almost $\sqrt{N}$-sparse vectors. The reason is that in general the set $S^{n-1} \setminus S_a^{n-1}(\sqrt{N})$ is much larger than $S_a^{n-1}(\sqrt{N})$, and we have to apply more delicate arguments to get a satisfactory probabilistic estimate. The construction of H for the set of "far from $\sqrt{N}$-sparse" vectors is contained in the following lemma:

Lemma 14: *Let ξ be a random variable such that for some $z \in \mathbb{R}$, $\gamma > 0$, $N \in \mathbb{N}$ we have*

$$\min(\mathbb{P}\{z-\sqrt{N} \le \xi \le z-1\}, \mathbb{P}\{z+1 \le \xi \le z+\sqrt{N}\}) \ge \gamma.$$

Then there exists an integer $\ell \in [0, \lfloor \log_2 \sqrt{N} \rfloor]$, $\lambda \in \mathbb{R}$ and disjoint Borel sets $H_1, H_2 \subset [-2^{\ell+2}; 2^{\ell+2}]$ such that $\mathrm{d}(H_1, H_2) \ge 2^\ell$,

$$\min(\mathbb{P}\{\xi-\lambda \in H_1\}, \mathbb{P}\{\xi-\lambda \in H_2\}) \ge c_{14}\gamma 2^{-\ell/8}$$

and $\mathbb{E}(\xi-\lambda)_H = 0$ for $H = H_1 \cup H_2$ and a universal constant $c_{14} > 0$.

Proof. Without loss of generality we can assume that $z = 0$. Let

$$c_{14} = \Big(\sum_{m=0}^{\infty} 2^{-m/8}\Big)^{-1}.$$

Then, by the conditions on ξ, there are $\ell_1, \ell_2 \in \{0, 1, \ldots, \lfloor \log_2 \sqrt{N} \rfloor\}$ such that

$$\mathbb{P}\{\xi \in [-2^{\ell_1+1}, -2^{\ell_1}]\} \ge c_{14}\gamma 2^{-\ell_1/8}; \quad \mathbb{P}\{\xi \in [2^{\ell_2}, 2^{\ell_2+1}]\} \ge c_{14}\gamma 2^{-\ell_2/8}.$$

Now, define λ as the conditional expectation of ξ given the event

$$\mathcal{M} = \{\omega \in \Omega : \xi(\omega) \in [-2^{\ell_1+1}, -2^{\ell_1}] \cup [2^{\ell_2}, 2^{\ell_2+1}]\},$$

i.e.,

$$\lambda = \mathbb{P}(\mathcal{M})^{-1} \int_{\mathcal{M}} \xi(\omega) d\,\omega.$$

Let $H_1 = -\lambda + [-2^{\ell_1+1}, -2^{\ell_1}]$ and $H_2 = -\lambda + [2^{\ell_2}, 2^{\ell_2+1}]$. Note that necessarily $\lambda \in [-2^{\ell_1+1}, 2^{\ell_2+1}]$, hence $H_1, H_2 \subset [-2^{\ell+2}, 2^{\ell+2}]$ for $\ell = \max(\ell_1, \ell_2)$. Obviously, $\mathrm{d}(H_1, H_2) \ge 2^\ell$ and for $H = H_1 \cup H_2$

$$\mathbb{E}(\xi-\lambda)_H = \int_{\{\xi-\lambda \in H\}} (\xi(\omega)-\lambda) d\,\omega = \int_{\mathcal{M}} (\xi(\omega)-\lambda) d\,\omega = 0.$$

Finally,

$$\begin{aligned}\min(&\mathbb{P}\{\xi-\lambda \in H_1\}, \mathbb{P}\{\xi-\lambda \in H_2\}) \\ &= \min(\mathbb{P}\{\xi \in [-2^{\ell_1+1}, -2^{\ell_1}]\}, \mathbb{P}\{\xi \in [2^{\ell_2}, 2^{\ell_2+1}]\}) \\ &\ge c_{14}\gamma 2^{-\ell/8}. \qquad \blacksquare\end{aligned}$$

Let us recall a folklore estimate of the norm of a random matrix with bounded mean zero entries (see, for example, [12, Proposition 2.4]):

Lemma 15: *Let $W = (w_{ij})$ be an $N\times n$ ($N \ge n$) random matrix with i.i.d. mean zero entries; $R > 0$ and assume that $|w_{ij}| \le R$ a.s. Then for a universal constant $C_{15} > 0$*

$$\mathbb{P}\{\|W\| \ge C_{15} R\sqrt{N}\} \le \exp(-N).$$

The following lemma highlights a useful property of the vectors from $S^{n-1} \setminus S_a^{n-1}(\sqrt{N})$:

Lemma 16: *For any integer $N \ge n \ge m \ge 1$ and any $y \in S^{n-1} \setminus S_a^{n-1}(\sqrt{N})$ there is a set $J = J(y) \subset \{1, 2, \ldots, n\}$ such that $|J| \le m$, $\|y\chi_J\| \ge \frac{1}{2}\sqrt{\frac{m}{n}}$ and $\|y\chi_J\|_\infty \le \frac{1}{\lfloor N^{1/4}\rfloor}$.*

Proof. Take any $N \ge n \ge m \ge 1$ and $y = (y_1, y_2, \ldots, y_n) \in S^{n-1} \setminus S_a^{n-1}(\sqrt{N})$ and let

$$J'(y) = \Big\{j \in \{1, 2, \ldots, n\} : |y_j| \le \frac{1}{\lfloor N^{1/4}\rfloor}\Big\}.$$

Obviously, $|J'| \ge n - \sqrt{N} > 0$ and, since y is not almost $\sqrt{N}$-sparse,

$$\|y\chi_{J'}\| \ge \sqrt{3/4}.$$

Let $\{J'_1, J'_2, \ldots, J'_p\}$ be any partition of J' into pairwise disjoint subsets of cardinality at most m with $p \le \lceil n/m\rceil$. Then, clearly, for some $q \in \{1, 2, \ldots, p\}$, $\|y\chi_{J_q}\| \ge \|y\chi_{J'}\|/\sqrt{p} > \frac{1}{2}\sqrt{\frac{m}{n}}$. Setting, $J(y) = J_q$, we get the result. ∎

Proposition 17 (The set $S^{n-1} \setminus S_a^{n-1}(\sqrt{N})$): *For any $\gamma > 0$, $\delta > 1$ there are $N_{17} \in \mathbb{N}$ and $h_{17} > 0$ depending only on γ and δ with the following property: Let $N \ge \max(N_{17}, \delta n)$ and let A be an $N \times n$ random matrix with i.i.d. entries such that*

$$\min(\mathbb{P}\{z - \sqrt{N} \le a_{11} \le z - 1\}, \mathbb{P}\{z + 1 \le a_{11} \le z + \sqrt{N}\}) \ge \gamma$$

for some $z \in \mathbb{R}$. Then for any non-random $N \times n$ matrix B and the set

$$S = S^{n-1} \setminus S_a^{n-1}(\sqrt{N})$$

we have

$$\mathbb{P}\{\inf_{y\in S} \|Ay + By\| \le h_{17}\sqrt{N}\} \le \exp(-w_{10}N/2).$$

Proof. Fix any $\gamma > 0$ and $\delta > 1$. To make the notation more compact, denote $f_0 := \frac{(1-\delta^{-1/4})\sqrt{c_{14}\gamma}}{C_5}$ and let $\tau_0 = \tau_0(\gamma, \delta)$ be the largest number in $(0, 1]$ such that for all $s \geq 0$

$$\Big(\frac{16\sqrt{8}C_{12}C_{15}2^{s/2}}{h_{10}f_0\tau_0^{3/2}}\Big)^{2^{-s/4}\tau_0} \leq \exp(w_{10}/4)$$

(it is not difficult to see that τ_0 is well defined). Then, take $N_{17} = N_{17}(\gamma, \delta)$ to be the smallest positive integer such that for all $N \geq N_{17}$

$$\text{(9)} \qquad \frac{1}{\lfloor N^{1/4} \rfloor} \leq \frac{f_0\sqrt{\tau_0}}{4\sqrt{8}} N^{-3/16} \quad \text{and} \quad \frac{48\sqrt{8N}C_{12}C_{15}}{h_{10}f_0\tau_0^{3/2}} \leq \exp(w_{10}N/4).$$

Let $N \geq N_{17}$, $N \geq \delta n$ and let A be an $N \times n$ random matrix with entries satisfying conditions of the lemma and B be any non-random $N \times n$ matrix.

By Lemma 14, there is an integer $\ell \in [0, \lfloor \log_2 \sqrt{N} \rfloor]$, $\lambda \in \mathbb{R}$ and disjoint Borel sets $H_1, H_2 \subset [-2^{\ell+2}, 2^{\ell+2}]$ such that $\mathrm{d}(H_1, H_2) \geq 2^\ell$,

$$\min(\mathbb{P}\{a_{11} - \lambda \in H_1\}, \mathbb{P}\{a_{11} - \lambda \in H_2\}) \geq c_{14}\gamma 2^{-\ell/8}$$

and $\mathbb{E}(a_{11} - \lambda)_H = 0$ for $H = H_1 \cup H_2$. Denote

$$\tilde{A} = A - \lambda \mathbf{1_{N\times n}}, \quad \tilde{B} = B + \lambda \mathbf{1_{N\times n}}$$

and let

$$R := 2^{\ell+2},\ d := 2^\ell,\ r := c_{14}\gamma 2^{-\ell/8},\ m := \Big\lceil \frac{\tau_0 n}{2^{\ell/4}} \Big\rceil,\ t := \frac{1}{2}\sqrt{\frac{m}{n}},\ \varepsilon := \frac{h_{10}h_{11}}{2C_{15}R},$$

where h_{11} is defined as in Proposition 11. Assume that S is non-empty and let $T \subset B_2^n$ consist of all m-sparse vectors $y \in B_2^n$ with $\|y\| \geq t$ and $\|y\|_\infty \leq \frac{2h_{11}}{d}$. The first inequality in (9) and a simple calculation show that $\frac{1}{\lfloor N^{1/4} \rfloor} \leq \frac{2h_{11}}{d}$. Hence, in view of Lemma 16, T is non-empty and satisfies (8). By Lemma 12, there is a finite subset $\mathcal{N} \subset T$ of cardinality at most $(\frac{nC_{12}}{m\varepsilon})^m$ such that for any $y \in S$ there is $y' = y'(y) \in \mathcal{N}$ with $\|y\chi_{\mathrm{supp} y'} - y'\| \leq \varepsilon$.

For each $y' \in \mathcal{N}$ denote

$$E_{y'} = \mathrm{span}\{e_j\}_{j \in \mathrm{supp} y'}.$$

By Proposition 11,

$$\mathbb{P}\{\mathrm{d}(\tilde{A}_H y', V_{\tilde{A},\tilde{B}}(H, E_{y'})) \leq h_{10}h_{11}\sqrt{N}\} \leq 2\exp(-w_{10}N).$$

Define an event

$$\mathcal{E} = \{\omega \in \Omega : \mathrm{d}(\tilde{A}_H(\omega)y', V_{\tilde{A},\tilde{B}}(H, E_{y'})(\omega)) > h_{10}h_{11}\sqrt{N} \text{ for all } y' \in \mathcal{N} \text{ and } \|\tilde{A}_H(\omega)\| \leq C_{15}R\sqrt{N}\}.$$

By the above probability estimates and Lemma 15,

$$\mathbb{P}(\mathcal{E}) \ge 1 - \exp(-N) - 2|\mathcal{N}| \exp(-w_{10}N) \ge 1 - \exp(-N) - 2\Big(\frac{C_{12}n}{m\varepsilon}\Big)^m \exp(-w_{10}N).$$

Using the definition of ε, m, τ_0 and the second inequality in (9), we can estimate the probability as

$$\begin{aligned}\mathbb{P}(\mathcal{E}) &\ge 1 - 3\Big(\frac{8C_{12}C_{15}2^{\ell+\ell/4}}{\tau_0 h_{10}h_{11}}\Big)^{2^{-\ell/4}\tau_0 n+1} \exp(-w_{10}N) \\ &\ge 1 - 3\Big(\frac{16\sqrt{8}C_{12}C_{15}2^{\ell/2}}{h_{10}f_0\tau_0^{3/2}}\Big)^{2^{-\ell/4}\tau_0 n+1} \exp(-w_{10}N) \\ &\ge 1 - \exp(-w_{10}N/2).\end{aligned}$$

Take any $\omega \in \mathcal{E}$ and define

$$D_1 = \tilde{A}_H(\omega), \quad D_2 = \tilde{A}_{\overline{H}}(\omega) + \tilde{B}, \quad D = A(\omega) + B(\omega) = D_1 + D_2.$$

Then $\|D_1\| \le C_{15}R\sqrt{N}$ and for every $y' \in \mathcal{N}$ we have

$$\mathrm{d}(D_1 y', D(E_{y'}^{\perp}) + D_2(E_{y'})) > h_{10}h_{11}\sqrt{N}.$$

Hence, by Proposition 3 and the definition of ε, we get

$$\inf_{y\in S} \|Dy\| > h_{10}h_{11}\sqrt{N} - \varepsilon C_{15}R\sqrt{N} = \frac{1}{2}h_{10}h_{11}\sqrt{N} \ge \frac{h_{10}f_0\sqrt{\tau_0}}{4\sqrt{8}}\sqrt{N}.$$

Finally, applying the above argument to the entire set $\mathcal{E}$, we obtain the result. ∎

Proof of Theorem 1. In view of the trivial identity $\mathcal{Q}(a_{ij}, \alpha) = \mathcal{Q}(a_{ij}/\alpha, 1)$, it is enough to prove the theorem for $\alpha = 1$. Fix any $\delta > 0$ and $\beta > 0$, let $\gamma = \beta/4$ and let $N_0 = N_0(\beta, \delta)$ be the smallest integer such that $N_0 \ge \max(N_{13}, N_{17})$ and for all $N \ge N_0$

$$N \le \exp(w_7 N/2) \quad \text{and} \quad 3 \le \exp(\min(w_7, w_{10})N/4).$$

Take any $N, n \in \mathbb{N}$ with $N \ge \max(N_0, \delta n)$, let $A = (a_{ij})$ be an $N \times n$ random matrix with i.i.d. entries satisfying $\mathcal{Q}(a_{11}, 1) \le 1 - \beta$ and let B be any non-random $N \times n$ matrix. By the right-continuity of the cdf of a_{11}, there is $z \in \mathbb{R}$ such that

$$\mathbb{P}\{a_{11} \le z - 1\} \ge \frac{\beta}{2} \quad \text{and} \quad \mathbb{P}\{a_{11} < z - 1\} \le \frac{\beta}{2}.$$

Then

$$\mathbb{P}\{a_{11} \ge z + 1\} \ge 1 - \mathbb{P}\{a_{11} < z - 1\} - \mathcal{Q}(a_{11}, 1) \ge \frac{\beta}{2}.$$

Let us consider three cases.

(1) $\mathbb{P}\{z+1 \le a_{11} \le z+\sqrt{N}\} \le \gamma$. Then

$$\mathcal{Q}(a_{11}, \sqrt{N}/8) \le \mathcal{Q}(a_{11}, (\sqrt{N}-1)/2) \le 1-\gamma.$$

Obviously, any vector on S^{n-1} is $N^{-1/2}$-peaky. Then, applying Proposition 7 with the "scaling factor" $\sqrt{N}/8$, we get

$$\begin{aligned}\mathbb{P}\{s_n(A+B) \le h_7\sqrt{N}/8\} &= \mathbb{P}\{\inf_{y\in S^{n-1}} \|Ay+By\| \le h_7\sqrt{N}/8\}\\ &\le n\exp(-w_7 N)\\ &\le \exp(-w_7 N/2).\end{aligned}$$

(2) $\mathbb{P}\{z-\sqrt{N} \le a_{11} \le z-1\} \le \gamma$. Treated as above.

(3) $\min(\mathbb{P}\{z-\sqrt{N} \le a_{11} \le z-1\}, \mathbb{P}\{z+1 \le a_{11} \le z+\sqrt{N}\}) \ge \gamma$. Define θ_{13} as in Proposition 13. By Proposition 7 for peaky vectors,

$$\mathbb{P}\{\inf_{y\in S_p^{n-1}(\theta_{13})} \|Ay+By\| \le h_7\theta_{13}\sqrt{N}\} \le n\exp(-w_7 N) \le \exp(-w_7 N/2).$$

By Propositions 13 and 17 for

$$S = S_a^{n-1}(\sqrt{N}) \setminus S_p^{n-1}(\theta_{13}) \quad \text{and} \quad S' = S^{n-1} \setminus S_a^{n-1}(\sqrt{N})$$

we have

$$\begin{aligned}\mathbb{P}\{\inf_{y\in S} \|Ay+By\| \le h_{13}\sqrt{N}\} &\le \exp(-w_{10}N/2);\\ \mathbb{P}\{\inf_{y\in S'} \|Ay+By\| \le h_{17}\sqrt{N}\} &\le \exp(-w_{10}N/2).\end{aligned}$$

Combining the estimates we get, for $h = \min(h_7\theta_{13}, h_{13}, h_{17})$,

$$\begin{aligned}\mathbb{P}\{s_n(A+B) \le h\sqrt{N}\} &\le \exp(-w_7 N/2) + 2\exp(-w_{10}N/2)\\ &\le \exp(-\min(w_7, w_{10})N/4).\end{aligned}$$

This completes the proof. ∎

Acknowledgement. I would like to thank my supervisor Prof. N. Tomczak-Jaegermann for valuable suggestions that helped improve the structure of the proof.

References

[1] R. Adamczak, A. Litvak, A. Pajor and N. Tomczak-Jaegermann, *Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles*, Journal of the American Mathematical Society **23** (2010), 535–561.

[2] R. Adamczak, A. Litvak, A. Pajor and N. Tomczak-Jaegermann, *Sharp bounds on the rate of convergence of the empirical covariance matrix*, Comptes Rendus Mathématique. Académie des Sciences. Paris **349** (2011), 195–200.

[3] Z. D. Bai and Y. Q. Yin, *Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix*, Annals of Probability **21** (1993), 1275–1294.

[4] O. Guédon, A. Litvak, A. Pajor and N. Tomczak-Jaegermann, *Restricted isometry property for random matrices with heavy-tailed columns*, Comptes Rendus Mathématique. Académie des Sciences. Paris **352** (2014), 431–434.

[5] V. Koltchinskii and S. Mendelson, *Bounding the smallest singular value of a random matrix without concentration*, International Mathematics Research Notices, (2015), doi: 10.1093/imrn/rnv096.

[6] P. Lévy, *Théorie de l'addition des variables aléatoires*, second edition, Gauthier-Villars, Paris, 1954.

[7] A. E. Litvak, A. Pajor, M. Rudelson and N. Tomczak-Jaegermann, *Smallest singular value of random matrices and geometry of random polytopes*, Advances in Mathematics **195** (2005), 491–523.

[8] A. E. Litvak S. Spektor, *Quantitative version of a Silverstein's result*, in *Geometric Aspects of Functional Analysis*, Lecture Notes in Mathematics **2116** (2014), 335–340.

[9] S. Mendelson and G. Paouris, *On the singular values of random matrices*, Journal of the European Mathematical Society **16** (2014), 823–834.

[10] R. I. Oliveira, *The lower tail of random quadratic forms, with applications to ordinary least squares and restricted eigenvalue properties*, arXiv:1312.2903.

[11] B. A. Rogozin, *On the increase of dispersion of sums of independent random variables*, (Russian) Teorija Verojatnostĭ i ee Primenenija **6** (1961), 106–108.

[12] M. Rudelson and R. Vershynin, *Non-asymptotic theory of random matrices: extreme singular values*, in *Proceedings of the International Congress of Mathematicians. Volume III*, Hindustan Book Agency, New Delhi, 2010, pp. 1576–102.

[13] M. Rudelson and R. Vershynin, *Small ball probabilities for linear images of high dimensional distributions*, International Mathematics Research Notices, (2014), doi: 10.1093/imrn/rnu243.

[14] M. Rudelson and R. Vershynin, *Smallest singular value of a random rectangular matrix*, Communications on Pure and Applied Mathematics **62** (2009), 1707–1739.

[15] M. Rudelson and R. Vershynin, *The Littlewood–Offord problem and invertibility of random matrices*, Advances in Mathematics **218** (2008), 600–633.

[16] A. Sankar, D. A. Spielman and S.-H. Teng, *Smoothed analysis of the condition numbers and growth factors of matrices*, SIAM Journal on Matrix Analysis and Applications **28** (2006), 446–476.

[17] D. A. Spielman and S.-H. Teng, *Smoothed analysis of algorithms*, in *Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002)*, Higher Ed. Press, Beijing, 2002, pp. 597–606.

[18] N. Srivastava and R. Vershynin, *Covariance estimation for distributions with* $2+\varepsilon$ *moments*, Annals of Probability **41** (2013), 3081–3111.

[19] T. Tao and V. Vu, *Inverse Littlewood–Offord theorems and the condition number of random discrete matrices*, Annals of of Mathematics **169** (2009), 595–632.

[20] T. Tao and V. Vu, *Smooth analysis of the condition number and the least singular value*, Mathematics of Computation **79** (2010), 2333–2352.

[21] T. Tao and V. Vu, *The condition number of a randomly perturbed matrix*, in *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, ACM, New York, 2007, pp. 248–255.

[22] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices*, in *Compressed Sensing: Theory and Applications*, Cambridge University Press, Cambridge, 2012, pp. 210–268.