NO-GAPS DELOCALIZATION FOR GENERAL RANDOM MATRICES

MARK RUDELSON AND ROMAN VERSHYNIN

Abstract. We prove that with high probability, every eigenvector of a random matrix is delocalized in the sense that any subset of its coordinates carries a non-negligible portion of its ℓ_2 norm. Our results pertain to a wide class of random matrices, including matrices with independent entries, symmetric and skew-symmetric matrices, as well as some other naturally arising ensembles. The matrices can be real and complex; in the latter case we assume that the real and imaginary parts of the entries are independent.

Contents

1. Introduction

While eigenvalues of random matrices have been extensively studied since 1950-s (see [1, 3, 35] for introduction), less is known about eigenvectors of random matrices. For matrices whose distributions are invariant under unitary or orthogonal transformations, the picture is trivial: their normalized eigenvectors are uniformly distributed over the unit Euclidean sphere. Examples of such random matrices include the classical Gaussian Unitary Ensemble (GUE), Gaussian Orthogonal Ensemble (GOE) and Ginibre ensembles. All entries of these matrices are normal, and either all of them are independent (in Ginibre ensemble) or independence holds modulo symmetry (in GUE and GOE).

Guided by the ubiquitous universality phenomenon in random matrix theory (see [38, 40, 19, 11]), we can anticipate that the eigenvectors behave in a similar way for a much broader class of random matrices. Thus, for a general $n \times n$ random matrix A with independent entries we may expect that the normalized eigenvectors are approximately uniformly distributed on the unit sphere. The same

Date: September 16, 2016.
Partially supported by NSF grants DMS 1161372, 1265782, 1464514, and USAF Grant FA9550-14-1-0009.

should hold for a general *Wigner matrix* A, a symmetric random matrix with independent entries on and above the diagonal.

The uniform distribution on the unit sphere has several remarkable properties. Showing that the eigenvectors of general random matrices have these properties, too, became a focus of attention in the recent years [17, 18, 12, 13, 6, 14, 15, 16, 41, 44, 5, 7, 33]. One of such properties is *delocalization in the sup-norm*. For a random vector v uniformly distributed on the unit sphere, a quick check reveals that no coefficients can be too large; in particular $\|v\|_\infty = O(\sqrt{\log n}/\sqrt{n})$ holds with high probability. Establishing a similar delocalization property for eigenvectors of random matrices is a challenging task. For eigenvectors of Hermitian random matrices, a weaker bound $\|v\|_\infty = O(\log^\gamma n/\sqrt{n})$, with $\gamma = O(1)$, was shown by Erdös et. al. [17, 18] using spectral methods. Later, Vu and Wang [44] obtained the optimal exponent $\gamma = 1/2$ for most eigenvectors (those corresponding to the bulk of the spectrum). Recently, the authors of the current paper established delocalization for random matrices with all independent entries by developing a completely different, geometric approach [33].

1.1. **No-gaps delocalization.** In the present paper, we will address a different natural delocalization property. Examining a random vector uniformly distributed on the sphere, we may notice that its mass (the ℓ_2 norm) is more or less evenly spread over the coordinates. There are no "gaps" in the sense that all subsets $J \subset [n]$ carry a non-negligible portion of the mass.

The goal of this paper is to establish this property for eigenvectors of random matrices. Formally, we would like to show that with high probability, for any eigenvector v, any $\varepsilon \in (0,1)$, and any subset of coordinates $J \subset [n]$ of size at least εn, one has

$$\Big(\sum_{j \in J} |v_j|^2\Big)^{1/2} \ge \phi(\varepsilon)\|v\|_2,$$

where $\phi : (0,1) \to (0,1)$ is some nice function. We call this phenomenon *no-gaps delocalization*.

One may wonder about the relation of the no-gaps delocalization to the delocalization in the sup-norm we mentioned before. As is easy to see, neither of these two properties implies the other. They offer complementary insights into the behavior of the coefficients of the eigenvectors – one property rules out peaks and the other rules out gaps.

The need for no-gaps delocalization arises naturally in problems of spectral graph theory. A similar notion appeared in the pioneering work of Dekel et. al. [9]. The desirability of establishing no-gaps delocalization was emphasized in a paper of Arora and Bhaskara [2], where a similar but weaker property was proved for a fixed subset J. Very recently, Eldan et. al. [10] established a weaker form of no-gaps delocalization for the Laplacian of an Erdös-Rényi graph with $\varepsilon > 1/2$ and the function ϕ depending on ε and n. This delocalization has been used to prove a version of a conjecture of Chung on the influence of adding or deleting edges on the spectral gap of an Erdös-Rényi graph. For shifted Wigner matrices and one-element sets J, the no-gaps delocalization was proved by Nguyen et. al. [25] with $\phi(1/n) = (1/n)^C$ for some absolute constant C.

In the present paper, we prove the no-gaps delocalization for a wide set of ensembles of random matrices including matrices with independent entries, symmetric and skew-symmetric random matrices, and others. Explicitly, we make the following assumption about possible dependencies among the entires.

Assumption 1.1 (Dependences of entries). Let A be an $N \times n$ random matrix. Assume that for any $i, j \in [n]$, the entry A_{ij} is independent of the rest of the entries except possibly A_{ji}. We also assume that the real part of A is random and the imaginary part is fixed.

Note that Assumtion 1.1 implies the following important independence property, which we will repeatedly use later: for any $J \subset [N]$, the entries of the submatrix $A_{J \times J^c}$ are independent.

Fixing the imaginary part in Assumtion 1.1 allows us to handle real random matrices. This assumption can also be arranged for complex matrices with independent real and imaginary parts, once we condition on the imaginary part. One can even consider a more general situation where the real parts of the entries conditioned on the imaginary parts have variances bounded below.

We will also assume $\|A\| = O(\sqrt{n})$ with high probability. This natural condition holds, in particular, if the entries of A have mean zero and bounded fourth moments [23]. To make this rigorous, we fix a number $M \geq 1$ and introduce the boundedness event

$$\mathcal{B}_{A,M} := \left\{ \|A\| \leq M\sqrt{n} \right\}. \tag{1.1}$$

1.2. **Main results.** Let us start with the simpler case where matrix entries have continuous distributions. This will allow us to present the method in the most transparent way, without having to navigate numerous obstacles that arise for discrete distributions.

Assumption 1.2 (Continuous distributions). We assume that the real parts of the matrix entries have densities bounded by some number $K \geq 1$.

Under Assumptions 1.1 and 1.2, we show that every subset of at least eight coordinates carries a non-negligible part of the mass of any eigenvector. This is summarized in the following theorem.

Theorem 1.3 (Delocalization: continuous distributions)**.** *Let A be an $n \times n$ random matrix which satisfies Assumptions 1.1 and 1.2. Choose $M \geq 1$ such that the boundedness event $\mathcal{B}_{A,M}$ holds with probability at least $1/2$. Let $\varepsilon \in (8/n, 1/2)$ and $s > 0$. Then, conditionally on $\mathcal{B}_{A,M}$, the following holds with probability at least $1 - (Cs)^{\varepsilon n}$. Every eigenvector v of A satisfies*

$$\|v_I\|_2 \geq (\varepsilon s)^6 \|v\|_2 \quad \text{for all } I \subset [n],\ |I| \geq \varepsilon n.$$

Here $C = C(K, M) \geq 1$.

The restriction $\varepsilon < 1/2$ can be easily removed, see Remark 1.6 below.

Note that we do not require any moments for the matrix entries, so heavy-tailed distributions are allowed. However, the boundedness assumption formalized by (1.1) implicitly yields some upper bound on the tails. Indeed, if the entries of A are i.i.d. and mean zero, then $\|A\| = O(\sqrt{n})$ can only hold if the fourth moments of entries are bounded [4].

Further, we do not require that the entries of A have mean zero. Therefore, adding to A any fixed matrix of norm $O(\sqrt{n})$ does not affect our results.

Extending Theorem 1.3 to general, possibly discrete distributions, is a challenging task. We are able to do this for matrices with identically distributed entries and under the mild assumption that the distributions of entries are not too concentrated near a single number.

Assumption 1.4 (General distribution of entries). We assume that the real parts of the matrix entries are distributed identically with a random variable ξ that satisfies

$$\sup_{u \in \mathbb{R}} \mathbb{P}\left\{ |\xi - u| \leq 1 \right\} \leq 1 - p, \quad \mathbb{P}\left\{ |\xi| > K \right\} \leq p/2 \quad \text{for some } K, p > 0. \tag{1.2}$$

Among many examples of discrete random variables ξ satisfying Assumption 1.4, the most prominent one is the symmetric *Bernoulli* random variable ξ, which takes values -1 and 1 with probability $1/2$ each.

With Assumption 1.2 replaced by Assumption 1.4, we can prove the no-gaps delocalization result, which we summarize in the following theorem.

Theorem 1.5 (Delocalization: general distributions)**.** *Let A be an $n \times n$ random matrix which satisfies Assumptions 1.1 and 1.4. Choose $M \geq 1$ such that the boundedness event $\mathcal{B}_{A,M}$ holds with* *probability at least 1/2. Let $\varepsilon \geq 1/n$ and $s \geq c_1\varepsilon^{-7/6}n^{-1/6} + e^{-c_2/\sqrt{\varepsilon}}$. Then, conditionally on $\mathcal{B}_{A,M}$, the following holds with probability at least $1-(Cs)^{\varepsilon n}$. Every eigenvector v of A satisfies*

$$\|v_I\|_2 \geq (\varepsilon s)^6 \|v\|_2 \quad \textit{for all } I \subset [n],\ |I| \geq \varepsilon n.$$

Here $c_k = c_k(p,K,M) > 0$ for $k = 1,2$ and $C = C(p,K,M) \geq 1$.

Remark 1.6. The restriction $s < 1/c_3$ making the theorem meaningful implies that $\varepsilon \in (c_4 n^{-1/7}, c_5)$ for some $c_4 > 0$ and $c_5 < 1$. The upper bound, however, can be easily removed. If $\varepsilon \geq c_5$, then delocalization event

$$\|v_I\|_2 \geq c_6\|v\|_2 \quad \text{for all } I \subset [n],\ |I| \geq \varepsilon n$$

holds with probability at least $1 - e^{-c_7 n}$. This follows by applying Theorem 1.5 with a sufficiently small constant $\varepsilon = c_7$ which would allow to choose $s = e^{-1}c_3$.

The restrictions on ε and s can be significantly relaxed; see the end of Section 6. We did not attempt to optimize these bounds, striving for clarity of the argument in lieu of more precise estimates.

Acknowledgement. We are grateful to Stephen Ge for pointing out an inaccuracy in the previous version of the argument, and to the anonymous referees whose suggestions led to a clearer presentation.

2. Outline of the argument

Our approach to Theorems 1.3 and 1.5 is based on reducing delocalization to invertibility of random matrices. We will now informally explain this reduction, which is quite flexible and can be applied for many classes of random matrices.

2.1. **Reduction of delocalization to invertibility.** Let us argue by contradiction. Suppose there exists a *localized* unit eigenvector v of A corresponding to some eigenvalue λ, which means that[1] $(A-\lambda)v = 0$ and

$$\|v_I\|_2 = o(1) \quad \text{for some index subset } I \subset [n],\ |I| = \varepsilon n. \tag{2.1}$$

Let us decompose the matrix $B := A - \lambda$ into two sub-matrices, B_I that consists of columns indexed by I and B_{I^c} with columns indexed by I^c. Then

$$0 = Bv = B_I v_I + B_{I^c} v_{I^c}. \tag{2.2}$$

To estimate the norm of $B_I v_I$, note that the operator norm of B can be bounded as

$$\|B_I\| \leq \|B\| \leq 2\|A\| = O(\sqrt{n}) \quad \text{with high probability,}$$

where we used the boundedness event (1.1). Combining with (2.1), we obtain

$$\|B_I v_I\|_2 = o(\sqrt{n}).$$

But the identity (2.2) implies that the norms of $B_I v_I$ and $B_{I^c} v_{I^c}$ are the same, thus

$$\|B_{I^c} v_{I^c}\|_2 = o(\sqrt{n}). \tag{2.3}$$

Since v is a unit vector and v_I has a small norm, the norm of v_{I^c} is close to 1. Then (2.3) implies that the matrix B_{I^c} is *not well invertible* on its range. Formally, this can be expressed as a bound on the smallest singular value:

$$s_{\min}(B_{I^c}) = o(\sqrt{n}). \tag{2.4}$$

Recall that B_{I^c} is an $n \times (n - \varepsilon n)$ random matrix. Thus we reduced delocalization to quantitative invertibility of almost square random matrices. (A minor difficulty here is that the eigenvalue λ

[1] For convenience of notation, we skip the identity symbol thus writing $A - \lambda$ for $A - \lambda I$.

depends on A, but we will be able to approximate λ by a fixed, non-random number using a simple discretization argument.)

2.2. **Invertibility of random matrices.** A standard expectation in the non-asymptotic random matrix theory is that random matrices are well invertible, and the bad event (2.4) should not hold, see e.g. [42]. For example, if the $n\times(n-\varepsilon n)$ random matrix $H = B_{I^c}$ had all independent standard normal entries, then we would have the desired lower bound

$$s_{\min}(H) = \Omega(\sqrt{n}) \quad \text{with high probability.} \tag{2.5}$$

This follows from the argument of Y. Gordon [20, 21], see [8]. Actually, this fact holds for distributions far more general than Gaussian, and in particular for many discrete distributions [24, 30]. Handling discrete distributions in the invertibility problems is considerably more challenging than continuous ones. Recent successes in these problems were based on understanding the interaction of probability with *arithmetic structure*, which was quantified via generalized arithmetic progressions in [37, 36] and approximate least common denominators (LCD) in [28, 30, 43]; see [38, 31] for background and references.

Nevertheless, there are significant difficulties in our situation that prevent us from deducing (2.5) for $H = B_{I^c}$ from any previous work. Let us mention some of these difficulties.

2.2.1. *Lack of independence.* Not all entries of A (and thus of H) may be independent. As we recall from Assumption 1.1, we are looking for ways to control symmetric and non-symmetric matrices simultaneously. This makes it necessary to extract rectangular blocks of independent entries from matrix H and modify the definition of the LCD adapting it to this block extraction.

2.2.2. *Small exceptional probability required.* We need that the delocalization result, and thus the invertibility bound (2.5), hold uniformly over all index subsets I of size εn. Since there are $\binom{n}{\varepsilon n} \sim \varepsilon^{-\varepsilon n}$ such sets, we would need the probability of non-invertibility (2.4) to be at most $\varepsilon^{\varepsilon n}$. While this is possible to achieve for real matrices with all independent entries [30], such small exceptional probabilities (smaller than $e^{-\varepsilon n}$) may not come automatically for the general case.

2.2.3. *Complex entries.* Results of the type (2.5) which hold with the probability we need are available only for real matrices; see in particular [31, 42, 27]. Since eigenvalues λ even of real matrices may be complex, we must work with complex random matrices. Extending the known results to complex matrices is non-trivial. Indeed, in order to preserve the matrix-vector multiplication, we replace a complex $n \times N$ random matrix $B = R + iT$ by the real $2n \times 2N$ random matrix $\left[\begin{smallmatrix} R & -T \\ T & R \end{smallmatrix}\right]$. The real and imaginary parts R and T each appear twice in this matrix, which causes extra dependences of the entries. Moreover, we encounter a major problem while trying to apply the covering argument to show that the least common denominator of the subspace orthogonal to a certain set of columns of H is large. Indeed, since we have to consider a real $2n \times 2N$ matrix, we will have to construct a net in a subset of the real sphere of dimension $2N$. The size of such net is exponential in the dimension. On the other hand, the number of independent rows of R is only n, so the small ball probability will be exponential in terms of n. As $n < N$, the union bound would not be applicable.

To overcome this difficulty, we introduce a stratification of the complex sphere, partitioning it according to the correlation between the real and the imaginary parts of vectors. This stratification, combined with a modified definition of the least common denominator, allows us to obtain stronger small ball probability estimates for weakly correlated vectors in Section 10. Yet the set of weakly correlated vectors has a larger complexity, which is expressed in the size of the nets. The cardinality of the nets has to be accurately estimated in Section 11. These two effects, the improvement of the small ball probability estimate and the increase of the complexity, work against each other. In Section 12, we show that they exactly balance each other, making it possible to apply the union bound.

2.3. **Organization of the argument.** After discussing basic background material in Section 3, we present a formal reduction of delocalization to invertibility in Section 4. The rest of the paper will focus on invertibility of random matrices. Section 5 covers continuous distributions; the main result there is Invertibility Theorem 5.1, from which we quickly deduce Delocalization Theorem 1.3. These sections are relatively simple and can be read independently of the rest of the paper.

Invertibility of random matrices with general distributions is considerably more difficult. We address this problem in Sections 6 – 13. The main result there is Invertibility Theorem 6.1, from which we quickly deduce Delocalization Theorem 1.5.

Our general approach to invertibility is based on the method from [28, 30] which developed upon the previous work [26, 37], see the surveys [38, 31, 27]. We reduce proving invertibility to the *distance problem*, where we seek a lower bound on $\operatorname{dist}(Z, E)$ where Z is a random vector with independent coordinates and E is an independent random subspace in $\mathbb{R}^N$. If we choose E to be a hyperplane (subspace of codimension one), we obtain an important class of examples in the distance problem, namely *sums of independent random variables*.

In Section 7 we study small ball probabilities for sums of real-valued independent random variables, as well as their higher dimensional versions $\operatorname{dist}(Z, E)$. These probabilities are controlled by the *arithmetic structure* of $E^\perp$, which we quantify via so-called *least common denominator* (LCD) of $E^\perp$. The larger LCD, the more $E^\perp$ is arithmetically unstructured, and the better are small ball probabilities for $\operatorname{dist}(Z, E)$. We formalize this relation in the very general Theorem 7.5, and then we specialize in Sections 7.3 and 7.4 to sums of independent random variables and distances to subspaces.

In Section 8, we state our main bound on the distance between random vectors and subspaces; this is Theorem 8.1. In order to deduce this result from the small ball probability bounds of Section 7, two things need to be done: (a) transfer the problem from complex to real, and (b) show that random subspaces are arithmetically unstructured, i.e. the LCD of $E^\perp$ is large. The transfer to a real problem is done in Section 8.1, and then our main focus becomes the structure of subspaces.

By the nature of our problem, the subspaces $E^\perp$ will be the kernels of random matrices. The analysis of such kernels starts in Section 9. We show there that all vectors in $E^\perp$ are *incompressible*, which means that they are not localized on a small fraction of coordinates.

Unfortunately, in the process of transferring the problem from complex to real in Section 8.1 introduces extra dependences among the entries of the random matrix. In Section 10 we adjust our results on small ball probabilities so they are not destroyed by those dependences. We find that these probabilities are controlled not only on LCD but also by *real-imaginary correlations* of the vectors in $E^\perp$.

Recall that our goal is to show that all vectors in $E^\perp = \ker(B)$ are unstructured, i.e. they have large LCD. We would obtain this if we can lower-bound Bz for all vectors with small LCD. For a fixed z, a lower bound follows from the small ball probability results of Section 10. To make the bound uniform, it is enough to run a union bound over a good net of the set of vectors with small LCD. We construct a good *net for level sets of LCD and real-imaginary correlations* in Section 11. Informally, small LCD or small correlation impose strong constraints, which make it possible to construct a smaller net than based on the trivial (volume-based) argument.

After this major step, the argument can be wrapped up relatively easily. In Section 12 we finalize the distance problem. We combine the small ball probability results with the fact that the random subspace are unstructured, and deduce Theorem 8.1.

In Section 13 we finalize the invertibility problem for general distributions; here we deduce Theorem 6.1. This is done by modifying the argument for continuous distributions in Section 5 using the non-trivial distance bound Theorem 8.1 for general distributions.

3. Notation and preliminaries

Throughout the paper, by $C, c, C_1, \dots$ we denote constants that may depend only on the parameters K and p that control the distributions of matrix entries in Assumptions 1.2 and 1.4 and the parameter M that controls the matrix norm in (1.1).

We denote by $S_{\mathbb{R}}^{n-1}$ and $S_{\mathbb{C}}^{n-1}$ the unit spheres of $\mathbb{R}^n$ and $\mathbb{C}^n$ respectively. We denote by $B(a, r)$ the Euclidean ball in $\mathbb{R}^n$ centered at a point a and with radius r. The unit sphere of a subspace E will be denoted S_E, and the orthogonal projection onto a subspace E by P_E.

Given an $n \times m$ matrix A and an index sets $J \subset [n]$, by A_J we denote the $n \times |J|$ sub-matrix of A obtained by including the columns indexed by J. Similarly, for a vector $z \in \mathbb{C}^n$, by z_J we denote the vector in $\mathbb{C}^J$ which consists of the coefficients indexed by J. For simplicity, we will routinely skip taking the integer part of expressions for cardinalities of various sets.

3.1. **Concentration function.**

The concept of *concentration function* has been introduced by P. Lévy and studied in probability theory for several decades, see [31] for the classical and recent history.

Definition 3.1 (Concentration function)**.** *Let Z be a random vector taking values in $\mathbb{C}^n$. The concentration function of X is defined as*

$$\mathcal{L}(Z, t) = \sup_{u \in \mathbb{C}^n} \mathbb{P}\left\{\|Z - u\|_2 \le t\right\}, \quad t \ge 0.$$

The concentration function gives a uniform upper bound on the *small ball probabilities* for X. We defer a detailed study of concentration function for sums of independent random variables to Section 7. Let us mention here only one elementary restriction property.

Lemma 3.2 (Small ball probabilities: restriction)**.** *Let $\xi_1, \dots, \xi_N$ be independent random variables and $a_1, \dots, a_N$ be real numbers. Then, for every subset of indices $J \subset [N]$ and every $t \ge 0$ we have*

$$\mathcal{L}\Big(\sum_{j \in J} a_j \xi_j, t\Big) \le \mathcal{L}\Big(\sum_{j=1}^{N} a_j \xi_j, t\Big).$$

Proof. This bound follows easily by conditioning on the random variables ξ_j with $j \notin J$ and absorbing their contribution into a fixed vector u in the definition of the concentration function. □

We will also use a simple and useful tensorization property which goes back to [26, 28].

Lemma 3.3 (Tensorization)**.** *Let $Z = (Z_1, \dots, Z_n)$ be a random vector in $\mathbb{C}^n$ with independent coordinates. Assume that there exists numbers $t_0, M \ge 0$ such that*

$$\mathcal{L}(Z_j, t) \le M(t + t_0) \quad \text{for all } j \text{ and } t \ge 0.$$

Then

$$\mathcal{L}(Z, t\sqrt{n}) \le [CM(t + t_0)]^n \quad \text{for all } t \ge 0.$$

Proof. By translation, we can assume without loss of generality that $u = 0$ in the definition of concentration function. Thus we want to bound the probability

$$\mathbb{P}\left\{\|Z\|_2 \le t\sqrt{n}\right\} = \mathbb{P}\left\{\sum_{j=1}^{n} |Z_j|^2 \le t^2 n\right\}.$$

Rearranging the terms, using Markov's inequality and then independence, we can bound this probability by

$$\mathbb{P}\left\{n - \frac{1}{t^2}\sum_{j=1}^{n} |Z_j|^2 > 0\right\} \le \mathbb{E}\exp\left(n - \frac{1}{t^2}\sum_{j=1}^{n} |Z_j|^2\right) = e^n \prod_{j=1}^{n} \mathbb{E}\exp(-|Z_j|^2/t^2). \tag{3.1}$$

To bound each expectation, we use the distribution integral formula followed by a change of variables. Thus

$$\mathbb{E}\exp(-|Z_j|^2/t^2) = \int_0^1 \mathbb{P}\left\{\exp(-|Z_j|^2/t^2) > x\right\} dx = \int_0^\infty 2ye^{-y^2}\,\mathbb{P}\left\{|Z_j| < ty\right\} dy.$$

By assumption, we have $\mathbb{P}\{|Z_j| < ty\} \le M(ty + t_0)$. Substituting this into the integral and evaluating it, we obtain

$$\mathbb{E}\exp(-|Z_j|^2/t^2) \le CM(t + t_0).$$

Finally, substituting this into (3.1), we see that the probability in question is bounded by $e^n[CM(t+t_0)]^n$. This completes the proof of the lemma. □

Remark 3.4 (Fixed shift). Note that the argument above establishes a somewhat stronger version of Lemma 3.3. Assume that for some $u = (u_1, \ldots, u_n) \in \mathbb{C}^n$ we have

$$\mathbb{P}\{|Z_j - u_j| \le t\} \le M(t + t_0) \quad \text{for all } j \text{ and } t \ge 0.$$

Then

$$\mathbb{P}\left\{\|Z - u\| \le t\sqrt{n}\right\} \le [CM(t + t_0)]^n \quad \text{for all } t \ge 0.$$

4. Reduction of delocalization to invertibility of random matrices

In this section, we show how to deduce delocalization from quantitative invertibility of random matrices. We outlined this reduction in Section 2.1 and will now make it formal. For simplicity of notation, we shall assume that $\varepsilon n/2 \in \mathbb{N}$, and we introduce the *localization* event

$$\mathrm{Loc}(A, \varepsilon, \delta) := \left\{\exists \text{ eigenvector } v \in S_{\mathbb{C}}^{n-1},\ \exists I \subset [n],\ |I| = \varepsilon n : \|v_I\|_2 < \delta\right\}.$$

Since we assume in Theorem 1.3 that the boundedness event $\mathcal{B}_{A,M}$ holds with probability at least 1/2, the conclusion of that theorem can be stated as follows:

$$\mathbb{P}\left\{\mathrm{Loc}(A, \varepsilon, (\varepsilon s)^6) \text{ and } \mathcal{B}_{A,M}\right\} \le (cs)^{\varepsilon n}. \tag{4.1}$$

The following proposition reduces proving delocalization results like (4.1) to an invertibility bound.

Proposition 4.1 (Reduction of delocalization to invertibility)**.** *Let A be an $n \times n$ random matrix with arbitrary distribution. Let $M \ge 1$ and $\varepsilon, p_0, \delta \in (0, 1/2)$. Assume that for any number $\lambda_0 \in \mathbb{C}$, $|\lambda_0| \le M\sqrt{n}$, and for any set $I \subset [n]$, $|I| = \varepsilon n$, we have*

$$\mathbb{P}\left\{s_{\min}\big((A - \lambda_0)_{I^c}\big) \le 8\delta M\sqrt{n} \text{ and } \mathcal{B}_{A,M}\right\} \le p_0. \tag{4.2}$$

Then

$$\mathbb{P}\left\{\mathrm{Loc}(A, \varepsilon, \delta) \text{ and } \mathcal{B}_{A,M}\right\} \le 5\delta^{-2}(e/\varepsilon)^{\varepsilon n} p_0.$$

Proof. Assume both the localization event and the boundedness event $\mathcal{B}_{A,M}$ hold. Using the definition of $\mathrm{Loc}(A, \varepsilon, \delta)$, choose a localized eigenvalue-eigenvector pair (λ, v) and an index subset I. Decomposing the eigenvector as

$$v = v_I + v_{I^c}$$

and multiplying it by $A - \lambda$, we obtain

$$0 = (A - \lambda)v = (A - \lambda)_I v_I + (A - \lambda)_{I^c} v_{I^c}.$$

By triangle inequality, this yields

$$\|(A - \lambda)_{I^c} v_{I^c}\|_2 = \|(A - \lambda)_I v_I\|_2 \le (\|A\| + |\lambda|)\|v_I\|_2.$$

By the localization event $\mathrm{Loc}(A, \varepsilon, \delta)$, we have $\|v_I\|_2 \le \delta$. By the boundedness event $\mathcal{B}_{A,M}$ and since λ is an eigenvalue of A, we have $|\lambda| \le \|A\| \le M\sqrt{n}$. Therefore

$$\|(A - \lambda)_{I^c} v_{I^c}\|_2 \le 2M\delta\sqrt{n}. \tag{4.3}$$

This happens for some λ in the disc $\{z \in \mathbb{C} : |z| \le M\sqrt{n}\}$. We will now run a covering argument in order to fix λ. Let $\mathcal{N}$ be a $(2M\delta\sqrt{n})$-net of that disc. One can construct $\mathcal{N}$ so that

$$|\mathcal{N}| \le \frac{5}{\delta^2}.$$

Choose $\lambda_0 \in \mathcal{N}$ so that $|\lambda_0 - \lambda| \le 2M\delta\sqrt{n}$. By (4.3), we have

$$\|(A - \lambda_0)_{I^c} v_{I^c}\|_2 \le 4M\delta\sqrt{n}. \tag{4.4}$$

Since $\|v_I\|_2 \le \delta \le 1/2$, we have $\|v_{I^c}\|_2 \ge \|v\|_2 - \|v_I\|_2 \ge 1/2$. Therefore, (4.4) implies that

$$s_{\min}((A - \lambda_0)_{I^c}) \le 8M\delta\sqrt{n}. \tag{4.5}$$

Summarizing, we have shown that the events $\mathrm{Loc}(A, \varepsilon, \delta)$ and $\mathcal{B}_{A,M}$ imply the existence of a subset $I \subset [n]$, $|I| = \varepsilon n$, and a number $\lambda_0 \in \mathcal{N}$, such that (4.5) holds. Furthermore, for fixed I and λ_0, assumption (4.2) states that (4.5) together with $\mathcal{B}_{A,M}$ hold with probability at most p_0. So by the union bound we conclude that

$$\mathbb{P}\left\{\mathrm{Loc}(A, \varepsilon, \delta) \text{ and } \mathcal{B}_{A,M}\right\} \le \binom{n}{\varepsilon n} \cdot |\mathcal{N}| \cdot p_0 \le \left(\frac{e}{\varepsilon}\right)^{\varepsilon n} \cdot \frac{5}{\delta^2} \cdot p_0.$$

This completes the proof of the proposition. □

5. Invertibility for continuous distributions

The reduction we made in the previous section puts invertibility of random matrices into the spotlight. Our goal becomes to establish invertibility property (4.2). In this section we do this for matrices for continuous distributions.

Theorem 5.1 (Invertibility: continuous distributions)**.** *Let A be an $n \times n$ random matrix satisfying the assumptions of Theorem 1.3. Let $M \ge 1$, $\varepsilon \in (0, 1)$, and let $I \subset [n]$ be any fixed subset with $|I| = \varepsilon n$. Then for any $t > 0$, we have*

$$\mathbb{P}\left\{s_{\min}(A_{I^c}) \le t\sqrt{n} \text{ and } \mathcal{B}_{A,M}\right\} \le (CKMt^{0.4}\varepsilon^{-1.4})^{\varepsilon n/2}.$$

Before we pass to the proof of this result, let us first see how it implies delocalization.

5.1. Deduction of Delocalization Theorem 1.3.

Let A be a matrix as in Theorem 1.3. We are going to use Proposition 4.1, so let us choose λ_0 and I as in that proposition and try to check the invertibility condition (4.2). Observe that the shifted matrix $A - \lambda_0$ still satisfies the assumptions of Theorem 5.1, and $\mathcal{B}_{A,M}$ implies $\mathcal{B}_{A-\lambda_0,2M}$ because $|\lambda_0| \le M\sqrt{n}$. So we can apply Theorem 5.1 for $A - \lambda_0$ and with $2M$, which yields

$$\mathbb{P}\left\{s_{\min}(A - \lambda_0 \,\mathrm{Id})_{I^c} \le t\sqrt{n} \text{ and } \mathcal{B}_{A,M}\right\} \le p_0,$$

for $t \ge 0$, where $p_0 = (CKMt^{0.4}\varepsilon^{-1.4})^{\varepsilon n/2}$. Therefore invertibility condition (4.2) holds for $\delta = t/8M$ and p_0. We will make sure that $\delta \le 1/2$ shortly, which is required in Proposition 4.1. Applying that proposition, we conclude that

$$\mathbb{P}\left\{\mathrm{Loc}(A, \varepsilon, t/8M) \text{ and } \mathcal{B}_{A,M}\right\} \le 5\left(\frac{8M}{t}\right)^2 (e/\varepsilon)^{\varepsilon n} p_0$$

for $t \ge 0$. Set $t = 8M(\varepsilon s)^6$ and substitute the value of p_0; this gives

$$\mathbb{P}\left\{\mathrm{Loc}(A, \varepsilon, (\varepsilon s)^6) \text{ and } \mathcal{B}_{A,M}\right\} \le (C(K, M)s)^{\varepsilon n}$$

for $s \ge 0$. It remains to check the promised condition $\delta \le 1/2$. We can assume that $s \le 1$ without loss of generality, since otherwise the probability bound in Theorem 1.3 is trivial. Since $\varepsilon < 1/2$ by assumption and $\delta = (\varepsilon s)^6$ by the choice above, we are done. This completes the proof of Theorem 1.3. □

The proof of Theorem 5.1 will occupy the rest of this section.

5.2. **Decomposition of the matrix.** To make the proof of Theorem 5.1 more convenient, let us change notation slightly. Instead of proving the theorem for an $n \times n$ matrix A and a block I of size $|I| = \varepsilon n$, we will do it for a $(1+2\varepsilon)n \times (1+2\varepsilon)n$ matrix A and a block I of size $|I| = 2\varepsilon n$. The desired conclusion in this case will be

$$\mathbb{P}\left\{s_{\min}(A_{I^c}) \le t\sqrt{n} \text{ and } \mathcal{B}_{A,M}\right\} \le (CKMt^{0.4}\varepsilon^{-1.4})^{\varepsilon n}. \tag{5.1}$$

Without loss of generality, we can assume that I is the interval of the *last* $2\varepsilon n$ indices.

Let us decompose A_{I^c} as follows:

$$\bar{A} := A_{I^c} = \begin{bmatrix} B \\ G \end{bmatrix}, \tag{5.2}$$

where B and G are rectangular matrices of size $(1+\varepsilon)n \times n$ and $\varepsilon n \times n$ respectively. By Assumption 1.1, the random matrices B and G are independent, and moreover all entries of G are independent.

We are going to show that either $\|Bx\|_2$ or $\|Gx\|_2$ is nicely bounded below for every vector $x \in S_{\mathbb{C}}^{n-1}$. To control B, we use the second negative moment identity to bound the Hilbert-Schmidt norm of the pseudo-inverse of B. We deduce from it that most singular values of B are not too small – namely, all but $0.1\varepsilon n$ singular values are bounded below by $\gtrsim \sqrt{\varepsilon n}$. It follows that B is nicely bounded below when restricted onto a subspace of codimension $0.1\varepsilon n$. (This subspace is formed by the corresponding singular vectors.) Next, we condition on B and we use G to control the remaining $0.1\varepsilon n$ dimensions. A simple covering argument shows that G is nicely bounded below when restricted to a subspace of dimension $0.1\varepsilon n$. Therefore, either B or G is nicely bounded below on the entire space, and thus A is nicely bounded below on the entire space as well.

We will now pass to a detailed proof of Theorem 5.1.

5.3. **Distances between random vectors and subspaces.** In this section we start working toward bounding B below on a large subspace. We quickly reduce this problem to a control of the distance between a random vector (a column of B) and a random subspace (the span of the rest of the columns). We then prove a lower bound for this distance.

5.3.1. *Negative second moment identity.* The negative second moment identity [39, Lemma A.4] expresses the Hilbert-Schmidt norm of the pseudo-inverse of B as follows:

$$\sum_{j=1}^{n} s_j(B)^{-2} = \sum_{i=1}^{n} \operatorname{dist}(B_j, H_j)^{-2}$$

where $s_j(B)$ denote the singular values of B, B_j denote the columns of B, and $H_j = \operatorname{span}(B_k)_{k \neq j}$.

To bound the sum above, we will establish a lower bound on the distance between the random vector $B_j \in \mathbb{C}^{(1+\varepsilon)n}$ and random subspace $H_j \subseteq \mathbb{C}^{(1+\varepsilon)n}$ of complex dimension $n-1$.

5.3.2. *Enforcing independence of vectors and subspaces.* Let us fix j. If all entries of B are independent, then B_j and H_j are independent. However, Assumption 1.1 leaves a possibility for B_j to be correlated with j-th row of B. This means that B_j and H_j may be dependent, which would complicate the distance computation.

There is a simple way to remove the dependence by projecting out the j-th coordinate. Namely, let $B'_j \in \mathbb{C}^{(1+\varepsilon)n-1}$ denote the vector B_j with j-th coordinate removed, and let $H'_j = \operatorname{span}(B'_k)_{k \neq j}$. We note the two key facts. First, B'_j and H'_j are independent by Assumption 1.1. Second,

$$\operatorname{dist}(B_j, H_j) \ge \operatorname{dist}(B'_j, H'_j), \tag{5.3}$$

since the distance between two vectors can only decrease after removing a coordinate.

Summarizing, we have

$$\sum_{j=1}^{n} s_j(B)^{-2} \geq \sum_{i=1}^{n} \operatorname{dist}(B'_j, H'_j)^{-2}. \tag{5.4}$$

Recall that $B'_j \in \mathbb{C}^{(1+\varepsilon)n-1}$ is a random vector with independent entries whose real parts have densities bounded by K (by Assumptions 1.1 and 1.2); and H'_j is an independent subspace of $\mathbb{C}^{(1+\varepsilon)n-1}$ of complex dimension $n-1$.

We are looking for a lower bound for the distances $\operatorname{dist}(B'_j, H'_j)$. It is convenient to represent them via the orthogonal projection of B'_j onto $(H'_j)^\perp$:

$$\operatorname{dist}(B'_j, H'_j) = \|P_{E_j} B'_j\|_2, \quad \text{where} \quad E_j = (H'_j)^\perp. \tag{5.5}$$

5.3.3. *Transferring the problem from $\mathbb{C}$ to $\mathbb{R}$.* We will now transfer the distance problem from the complex to the real field. To this end, we define the operation $z \mapsto \widetilde{z}$ that makes complex vectors real in the obvious way:

$$\text{for } z = x + iy \in \mathbb{C}^N, \text{ define } \widetilde{z} = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^{2N}.$$

Similarly, we can make a complex subspace $E \subset \mathbb{C}^N$ real by defining

$$\widetilde{E} = \{\widetilde{z} : \ z \in E\} \subset \mathbb{R}^{2N}.$$

Note that this operation doubles the dimension of E.

Let us record two properties that follow straight from this definition.

Lemma 5.2 (Elementary properties of operation $x \mapsto \widetilde{x}$)**.** *1. For a complex subspace E and a vector z, one has*

$$\widetilde{P_E z} = P_{\widetilde{E}} \widetilde{z}.$$

2. For a complex-valued random vector X and $r \geq 0$, one has

$$\mathcal{L}(\widetilde{X}, r) = \mathcal{L}(X, r).$$

Recall that the second part of this lemma is about the concentration function $\mathcal{L}(X, r)$ we introduced in Section 3.

After applying the operation $z \mapsto \widetilde{z}$ to the random vector B'_j in (5.4), we encounter a problem. Since the imaginary part of B'_j is fixed by Assumption 1.1, only half of the coordinates of $\widetilde{B'_j}$ will be random, and that will not be enough for us. The following lemma solves this problem by randomizing all coordinates.

Lemma 5.3 (Randomizing all coordinates)**.** *Consider a random vector $Z = X + iY \in \mathbb{C}^N$ whose imaginary part $Y \in \mathbb{R}^N$ is fixed. Set $\widehat{Z} = \binom{X_1}{X_2} \in \mathbb{R}^{2N}$ where X_1 and X_2 are independent copies of X. Let E be a subspace of $\mathbb{C}^N$. Then*

$$\mathcal{L}(P_E Z, r) \leq \mathcal{L}(P_{\widetilde{E}} \widehat{Z}, 2r)^{1/2}, \quad r \geq 0.$$

Proof. Recalling the definition of the concentration function, in order to bound $\mathcal{L}(P_E Z, r)$ we need to choose arbitrary $a \in \mathbb{C}^N$ and find a uniform bound on the probability

$$p := \mathbb{P}\left\{\|P_E Z - a\|_2 \leq r\right\}.$$

By assumption, the random vector $Z = X + iY$ has fixed imaginary part Y. So it is convenient to express the probability as

$$p = \mathbb{P}\left\{\|P_E X - b\|_2 \leq r\right\}$$

where $b = a - P_E(iY)$ is fixed. Let us rewrite this identity using independent copies X_1 and X_2 of X as follows:

$$p = \mathbb{P}\left\{\|P_E X_1 - b\|_2 \le r\right\} = \mathbb{P}\left\{\|P_E(iX_2) - ib\|_2 \le r\right\}.$$

(The last equality follows trivially by multiplying by i inside the norm.) By independence of X_1 and X_2 and using triangle inequality, we obtain

$$\begin{aligned} p^2 &= \mathbb{P}\left\{\|P_E X_1 - b\|_2 \le r \text{ and } \|P_E(iX_2) - ib\|_2 \le r\right\} \\ &\le \mathbb{P}\left\{\|P_E(X_1 + iX_2) - b - ib\|_2 \le 2r\right\} \\ &\le \mathcal{L}(P_E(X_1 + iX_2), 2r). \end{aligned}$$

Further, using part 2 and then part 1 of Lemma 5.2, we see that

$$\mathcal{L}(P_E(X_1 + iX_2), 2r) = \mathcal{L}(P_{\widetilde{E}}(\widetilde{X_1 + iX_2}), 2r) = \mathcal{L}(P_{\widetilde{E}}\widehat{Z}, 2r).$$

Thus we showed that $p^2 \le \mathcal{L}(P_{\widetilde{E}}\widehat{Z}, 2r)$ uniformly in a. By definition of the concentration function, this completes the proof. □

5.3.4. *Bounding the distances below.* We are ready to control the distances appearing in (5.5).

Lemma 5.4 (Distance between random vectors and subspaces)**.** *For every $j \in [n]$ and $\tau > 0$, we have*

$$\mathbb{P}\left\{\operatorname{dist}(B'_j, H'_j) < \tau\sqrt{\varepsilon n}\right\} \le (CK\tau)^{\varepsilon n}. \tag{5.6}$$

Proof. Representing the distances via projections of B'_j onto the subspaces $E_j = (H'_j)^\perp$ as in (5.5), and using the definition of the concentration function, we have

$$p_j := \mathbb{P}\left\{\operatorname{dist}(B'_j, H'_j) < \tau\sqrt{\varepsilon n}\right\} \le \mathcal{L}(P_{E_j} B'_j,\, \tau\sqrt{\varepsilon n}).$$

Recall that B'_j and E_j are independent, and let us condition on E_j. Lemma 5.3 implies that

$$p_j \le \mathcal{L}(P_{\widetilde{E_j}}\widehat{Z},\, \tau\sqrt{\varepsilon n})^{1/2}$$

where $\widehat{Z}$ is a random vector with independent coordinates that have densities bounded by K.

Recall that H'_j has codimension εn; thus E_j has dimension εn and $\widetilde{E_j}$ has dimension $2\varepsilon n$. We can use a bound on the small ball probability from [32], which states that the density of $P_{\widetilde{E_j}}\widehat{Z}$ is bounded by $(CK)^{2\varepsilon n}$. Integrating the density over a ball of radius $2\tau\sqrt{\varepsilon n}$ in the subspace $\widetilde{E_j}$ that has volume $(C\tau)^{2\varepsilon n}$, we conclude that

$$\mathcal{L}(P_{\widetilde{E_j}}\widehat{Z},\, \tau\sqrt{\varepsilon n}) \le (CK\tau)^{2\varepsilon n}.$$

It follows that

$$p_j \le (CK\tau)^{\varepsilon n},$$

as claimed. The proof of Lemma 5.4 is complete. □

5.4. B is bounded below on a large subspace E^+.

5.4.1. *Plugging the distance bound into second moment inequality.* In order to substitute the bound (5.6) into the negative second moment inequality (5.4), let us recall some classical facts about the weak L^p norms. The weak L^p norm of a random variable Y is defined as

$$\|Y\|_{p,\infty} = \sup_{t>0} t \cdot (\mathbb{P}\{|Y| > t\})^{1/p}.$$

This is not a norm but is equivalent to a norm if $p > 1$. In particular, the weak triangle inequality holds:

$$\|\sum_i Y_i\|_{p,\infty} \le C(p) \sum_i \|Y_i\|_{p,\infty} \tag{5.7}$$

where $C(p)$ is bounded above by an absolute constant for $p \ge 2$, see [34], Theorem 3.21.

The bound (5.6) means that $Y_i := \operatorname{dist}(B_i, H_i)^{-2}$ are in weak L^p for $p = \varepsilon n/2$, and that $\|Y_i\|_{p,\infty} \le C^2K^2/\varepsilon n$. Since by assumption $p \ge 2$, the weak triangle inequality (5.7) yields $\| \sum_{i=1}^n Y_i\|_{p,\infty} \le C_2^2 K^2/\varepsilon$. This in turn means that

$$\mathbb{P}\left\{\sum_{i=1}^n \operatorname{dist}(B_i, H_i)^{-2} > \frac{1}{\tau^2 \varepsilon}\right\} \le (C_2 K\tau)^{\varepsilon n}, \quad \tau > 0.$$

Therefore, by the second negative moment identity (5.4), the event

$$\mathcal{E}_1 := \left\{\sum_{i=1}^n s_i(B)^{-2} \le \frac{1}{\tau^2 \varepsilon}\right\} \tag{5.8}$$

is likely: $\mathbb{P}((\mathcal{E}_1)^c) \le (C_2 K\tau)^{\varepsilon n}$.

5.4.2. *A large subspace E^+ on which B is bounded below.* Fix a parameter $\tau > 0$ for now, and assume that the event (5.8) occurs. By Markov's inequality, for any $\delta > 0$ we have

$$\Big|\big\{i : \; s_i(B) \le \delta\sqrt{n}\big\}\Big| = \Big|\big\{i : \; s_i(B)^{-2} \ge \frac{1}{\delta^2 n}\big\}\Big| \le \frac{\delta^2 n}{\tau^2 \varepsilon}.$$

Let $c \in (0,1)$ be a small absolute constant. Choosing $\delta = c\tau\varepsilon$, we have

$$\Big|\big\{i : \; s_i(B) \le c\tau\varepsilon\sqrt{n}\big\}\Big| \le c\varepsilon n. \tag{5.9}$$

Let $v_i(B)$ be the right singular vectors of B, and consider the (random) orthogonal decomposition $\mathbb{C}^n = E^- \oplus E^+$, where

$$E^- = \operatorname{span}\{v_i(B) : \; s_i(B) \le c\tau\varepsilon\sqrt{n}\}, \quad E^+ = \operatorname{span}\{v_i(B) : \; s_i(B) > c\tau\varepsilon\sqrt{n}\}.$$

Inequality (5.9) means that $\dim_{\mathbb{C}}(E^-) \le c\varepsilon n$.

Let us summarize. We obtained that the event

$$\mathcal{D}_{E^-} := \big\{\dim(E^-) \le c\varepsilon n\big\} \text{ satisfies } \mathbb{P}((\mathcal{D}_{E^-})^c) \le (C_2 K\tau)^{\varepsilon n}, \tag{5.10}$$

so E^- is likely to be a small subspace and E^+ a large subspace. Moreover, by definition, B is nicely bounded below on E^+:

$$\inf_{x \in S_{E^+}} \|Bx\|_2 \ge c\tau\varepsilon\sqrt{n}. \tag{5.11}$$

5.5. **G is bounded below on the small complementary subspace E^-.** Recall that the subspaces E^+ and E^- are determined by the sub-matrix B, so these subspaces are independent of G by Assumption 1.1. Let us fix B so that $\dim(E^-) \le c\varepsilon n$; recall this is a likely event by (5.10).

Note that G is an $\varepsilon n \times n$ random matrix with independent entries. We are going to show that G is well bounded below when restricted onto the fixed subspace E^-. This can be done by a standard covering argument, where a lower bound is first proved for a fixed vector, then extended to a δ-net of the sphere by a union bound, and finally to the whole sphere by approximation.

5.5.1. *Lower bounds on a fixed vector.*

Lemma 5.5 (Lower bound for a fixed row and vector)**.** *Let G_j denote the j-th row of G. Then for each j, $z \in S_{\mathbb{C}}^{n-1}$, and $\theta \geq 0$, we have*

$$\mathbb{P}\left\{|\langle G_j, z\rangle| \leq \theta\right\} \leq C_0 K\theta. \tag{5.12}$$

Proof. Fix j and consider the random vector $Z = G_j$. Expressing Z and z in terms of their real and imaginary parts as

$$Z = X + iY, \quad z = x + iy,$$

we can write the inner product as

$$\langle Z, z\rangle = [\langle X, x\rangle - \langle Y, y\rangle] + i\,[\langle X, y\rangle + \langle Y, x\rangle]\,.$$

Since z is a unit vector, either x or y has norm at least $1/2$. Assume without loss of generality that $\|x\|_2 \geq 1/2$. Dropping the imaginary part, we obtain

$$|\langle Z, z\rangle| \geq |\langle X, x\rangle - \langle Y, y\rangle|\,.$$

Recall that the imaginary part Y is fixed by Assumption 1.1. Thus

$$\mathbb{P}\left\{|\langle Z, z\rangle| \leq \theta\right\} \leq \mathcal{L}(\langle X, x\rangle, \theta). \tag{5.13}$$

We can express $\langle X, x\rangle$ in terms of the coordinates of X and x as the sum

$$\langle X, x\rangle = \sum_{k=1}^{n} X_k x_k.$$

Since X_k are the real parts of independent entries of G, Assumptions 1.1 and 1.2 imply that X_k are independent random variables with densities bounded by K. Recalling that $\sum_{k=1}^{n} x_k^2 \geq 1/2$, we can apply a known result about the densities of sums of independent random variables, see [32]. It states that the density of $\sum_{k=1}^{n} X_k x_k$ is bounded by CK. It follows that

$$\mathcal{L}(\langle X, x\rangle, \theta) \leq CK\theta. \tag{5.14}$$

Substituting this into (5.13) completes the proof of Lemma 5.5. ☐

Lemma 5.6 (Lower bound for a fixed vector)**.** *For each $x \in S_{\mathbb{C}}^{n-1}$ and $\theta > 0$, we have*

$$\mathbb{P}\left\{\|Gx\|_2 \leq \theta\sqrt{\varepsilon n}\right\} \leq (C_0 K\theta)^{\varepsilon n}.$$

Proof. We can represent $\|Gx\|_2^2$ as a sum of independent random variables $\sum_{j=1}^{\varepsilon n} |\langle G_j, x\rangle|^2$. Each of the terms $\langle G_j, x\rangle$ satisfies (5.12). Then the conclusion follows from Tensorization Lemma 3.3. ☐

5.5.2. *Lower bound on a subspace.*

Lemma 5.7 (Lower bound on a subspace)**.** *Let $M \geq 1$ and $\mu \in (0,1)$. Let E be a fixed subspace of $\mathbb{C}^n$ of dimension at most $\mu\varepsilon n$. Then, for every $\theta > 0$, we have*

$$\mathbb{P}\left\{\inf_{x \in S_E} \|Gx\|_2 < \theta\sqrt{\varepsilon n} \text{ and } \mathcal{B}_{G,M}\right\} \leq \left[CK(M/\sqrt{\varepsilon})^{2\mu}\theta^{1-2\mu}\right]^{\varepsilon n}. \tag{5.15}$$

Proof. Let $\delta \in (0,1)$ to be chosen later. Since the dimension of $\widetilde{E} \subset \mathbb{R}^{2n}$ is at most $2\mu\varepsilon n$, standard volume considerations imply the existence of a δ-net $\mathcal{N} \subset S_E$ with

$$|\mathcal{N}| \leq \left(\frac{3}{\delta}\right)^{2\mu\varepsilon n}. \tag{5.16}$$

Assume that the event in the left hand side of (5.15) occurs. Choose $x \in S_E$ such that $\|Gx\|_2 < \theta\sqrt{\varepsilon n}$. Next, choose $x_0 \in \mathcal{N}$ such that $\|x - x_0\|_2 \leq \delta$. By triangle inequality and using $\mathcal{B}_{G,M}$, we obtain

$$\|Gx_0\|_2 \leq \|Gx\|_2 + \|G\| \cdot \|x - x_0\|_2 \leq \theta\sqrt{\varepsilon n} + M\sqrt{n} \cdot \delta$$

Choosing $\delta := \theta\sqrt{\varepsilon}/M$, we conclude that $\|Gx_0\|_2 \le 2\theta\sqrt{\varepsilon n}$.

Summarizing, we obtained that the probability of the event in the left hand side of (5.15) is bounded by

$$\mathbb{P}\left\{\exists x_0 \in \mathcal{N} :\ \|Gx_0\|_2 \le 2\theta\sqrt{\varepsilon n}\right\}.$$

By Lemma 5.6 and a union bound, this in turn is bounded by

$$|\mathcal{N}| \cdot (2C_0K\theta)^{\varepsilon n} \le \Big(\frac{3M}{\theta\sqrt{\varepsilon}}\Big)^{2\mu\varepsilon n} \cdot (2C_0K\theta)^{\varepsilon n},$$

where we used (5.16) and our choice of δ. Rearranging the terms completes the proof. □

5.5.3. *Conclusion: G is bounded below on a large subspace E^-.* We can apply Lemma 5.7 for the subspace $E = E^-$ constructed in Section 5.4.2. We do this conditionally on B, for a fixed choice of E^- that satisfies[2] $\mathcal{D}_{E^-}$, thus for $\mu \le c < 0.05$. This yields the following.

Lemma 5.8 (G is bounded below on E^-)**.** *For every $\theta > 0$, we have*

$$\mathbb{P}\left\{\inf_{x\in S_{E^-}} \|Gx\|_2 < \theta\sqrt{\varepsilon n} \text{ and } \mathcal{D}_{E^-} \text{ and } \mathcal{B}_{G,M}\right\} \le \left(CKM^{0.1}\varepsilon^{-0.05}\theta^{0.9}\right)^{\varepsilon n}.$$

5.6. **Proof of invertibility.**

5.6.1. *Decomposing invertibility.* The following lemma reduces invertibility of $\bar{A}$ to invertibility of B on E^+ and G on E^-.

Lemma 5.9 (Decomposition)**.** *Let A be an $m \times n$ matrix. Let us decompose A as*

$$A = \begin{bmatrix} B \\ G \end{bmatrix}, \quad B \in \mathbb{C}^{m_1\times n},\ G \in \mathbb{C}^{m_2\times n},\ m = m_1 + m_2.$$

Consider the orthogonal decomposition $\mathbb{C}^n = E^- \oplus E^+$ where E^- and E^+ are eigenspaces[3] of B^*B. Denote

$$s_A = s_{\min}(A),\ s_B = s_{\min}(B|_{E^+}),\ s_G = s_{\min}(G|_{E^-}).$$

Then

$$s_A \ge \frac{s_B s_G}{4\|A\|}. \tag{5.17}$$

Proof. Let $x \in S^{n-1}$. We consider the orthogonal decomposition

$$x = x^- + x^+, \quad x^- \in E^-,\ x^+ \in E^+.$$

We can also decompose Ax as

$$\|Ax\|_2^2 = \|Bx\|_2^2 + \|Gx\|_2^2.$$

Let us fix a parameter $\lambda \in (0, 1/2)$ and consider two cases.

Case 1: $\|x^+\|_2 \ge \lambda$. Then

$$\|Ax\|_2 \ge \|Bx\|_2 \ge \|Bx^+\|_2 \ge s_B \cdot \lambda.$$

Case 2: $\|x^+\|_2 < \lambda$. In this case, $\|x^-\|_2 = \sqrt{1 - \|x^+\|_2^2} \ge 1/2$. Thus

$$\begin{aligned}\|Ax\|_2 &\ge \|Gx\|_2 \ge \|Gx^-\|_2 - \|Gx^+\|_2 \\ &\ge \|Gx^-\|_2 - \|G\| \cdot \|x^+\|_2 \ge s_G \cdot \frac{1}{2} - \|G\| \cdot \lambda.\end{aligned}$$

[2]Recall that $\mathcal{D}_{E^-}$ is the likely event defined in (5.10).

[3]In other words, E^- and E^+ are the spans of two disjoint subsets of right singular vectors of B.

Using that $\|G\| \le \|A\|$, we conclude that

$$s_A = \inf_{x \in S^{n-1}} \|Ax\|_2 \ge \min\Big(s_B \cdot \lambda,\ s_G \cdot \frac{1}{2} - \|A\| \cdot \lambda\Big).$$

Optimizing the parameter λ, we conclude that

$$s_A \ge \frac{s_B s_G}{2(s_B + \|A\|)}.$$

Using that s_B is bounded by $\|A\|$, we complete the proof. □

5.6.2. *Proof of the Invertibility Theorem 5.1.* We apply Lemma 5.9 for the matrix $\bar{A}$ and the decomposition (5.2) and obtain

$$s_B s_G \le 4\|\bar{A}\| s_{\bar{A}}.$$

Since $\bar{A}$ is a sub-matrix of A, we have $\|\bar{A}\| \le M\sqrt{n}$ on the event $\mathcal{B}_{A,M}$. Further, (5.11) yields the bound $s_B \ge c\tau\varepsilon\sqrt{n}$. It follows that

$$\begin{aligned} \mathbb{P}\left\{s_{\bar{A}} < t\sqrt{n} \text{ and } \mathcal{B}_{A,M}\right\} &\le \mathbb{P}\left\{s_G < \frac{4Mt}{c\tau\varepsilon} \cdot \sqrt{n} \text{ and } \mathcal{B}_{A,M}\right\} \\ &\le \mathbb{P}\left\{s_G < \frac{4Mt}{c\tau\varepsilon^{3/2}} \cdot \sqrt{\varepsilon n} \text{ and } \mathcal{D}_{E^-} \text{ and } \mathcal{B}_{A,M}\right\} + \mathbb{P}((\mathcal{D}_{E^-})^c). \end{aligned} \tag{5.18}$$

The last line prepared us for an application of Lemma 5.8. Using this lemma along with the trivial inclusion $\mathcal{B}_{A,M} \subseteq \mathcal{B}_{G,M}$ and the estimate (5.10) of the probability of $(\mathcal{D}_{E^-})^c$, we bound the quantity in (5.18) by

$$\left[CKM^{0.1}\varepsilon^{-0.05}\Big(\frac{4Mt}{c\tau\varepsilon^{3/2}}\Big)^{0.9}\right]^{\varepsilon n} + (C_2 K\tau)^{\varepsilon n}.$$

This bound holds for all $\tau, t > 0$. Choosing $\tau = \sqrt{t}$ and rearranging the terms, we obtain

$$\mathbb{P}\left\{s_{\bar{A}} < t\sqrt{n} \text{ and } \mathcal{B}_{A,M}\right\} \le \left[CKM\varepsilon^{-1.4}t^{0.45}\right]^{\varepsilon n} + (CKt^{0.5})^{\varepsilon n}.$$

This implies the desired conclusion (5.1). Theorem 5.1 is proved. □

6. Invertibility for general distributions: statement of the result

We are now passing to random matrices whose entries may have general, possibly discrete distributions; our goal being Delocalization Theorem 1.5. Recall that in Section 2.1 we described on the informal level how delocalization can be reduced to invertibility of random matrices; Proposition 4.1 formalizes this reduction. This prepares us to state an invertibility result for general random matrices, whose proof will occupy the rest of this paper.

Theorem 6.1 (Invertibility: general distributions)**.** *Let A be an $n \times n$ random matrix satisfying the assumptions of Theorem 1.5. Let $M \ge 1$, $\varepsilon \in (1/n, c)$, and let $I \subset [n]$ be any fixed subset with $|I| = \varepsilon n$. Then for any*

$$t \ge \frac{c}{\varepsilon n} + e^{-c/\sqrt{\varepsilon}}, \tag{6.1}$$

we have

$$\mathbb{P}\left\{s_{\min}(A_{I^c}) \le t\sqrt{n} \text{ and } \mathcal{B}_{A,M}\right\} \le \left(Ct^{0.4}\varepsilon^{-1.4}\right)^{\varepsilon n/2}.$$

The constant C in the inequality above depends on M and the parameters p and K appearing in Assumption 1.4.

Delocalization Theorem 1.5 follows from Theorem 6.1 along the same lines as in Section 5.1. As in that section, for a given s we set $t = 8M(\varepsilon s)^6$, which leads to the particular form of the restriction on s in Theorem 1.5. This restriction, as well as the probability estimate, can be improved by tweaking various parameters throughout the proof of Theorem 6.1. They can be further and more significantly improved by taking into account the arithmetic structure in the small ball probability estimate, instead of disregarding it in Section 10. We refrained from pursuing these improvements in order to avoid overburdening the paper with technical calculations.

7. Small ball probabilities via least common denominator

In this section, which may have an independent interest, we relate the sums of independent random variables and random vectors to the arithmetic structure of their coefficients.

To see the relevance of this topic to invertibility of random matrices, we could try to extend the argument we gave Section 5 to general distributions. Most of the argument would go through. However, a major difficulty occurs when we try to estimate the distance between a random vector X and a fixed subspace H. For discrete distributions, $\mathrm{dist}(X, H) = \|P_{H^\perp} X\|_2$ can no longer be bounded below as easily as we did in Lemma 5.4. The source of difficulty can be best seen if we consider the simple example where H is the hyperplane orthogonal to the vector $(1, 1, 0, 0, \dots, 0)$ and X is the random Bernoulli vector (whose coefficients are independent and take values 1 and -1 with probability each). In this case, $\mathrm{dist}(X, H) = 0$ with probability $1/2$. Even if we exclude zeros by making H orthogonal to $(1, 1, 1, 1, \dots, 1)$, the distance would equal zero with probability $\sim 1/\sqrt{n}$, thus polynomially rather than exponentially fast in n.

The problem with these examples is that $H^\perp$ had rigid arithmetic structure. In Section 7.1, we will show how to quantify arithmetic structure with a notion of approximate least common denominator (LCD). In Section 7.2 will also provide bounds on sums of independent random vectors in terms of LCD. Finally, in Sections 7.3 and 7.4 we will specialize these bounds for sums of independent random variables and projections of random vectors (and in particular, for distances to subspaces).

7.1. The least common denominator.

An approximate concept of least common denominator (LCD) was proposed in [28] to quantify the arithmetic structure of vectors; this idea was developed in [30, 29, 43], see also [31]. Here we will use the version of LCD from [43]. We emphasize that throughout this section we consider *real* vectors and matrices.

Definition 7.1 (Least common denominator). *Fix $L > 0$. For a* vector $v \in \mathbb{R}^N$*, the least common denominator (LCD) is defined as*

$$D(v) = D(v, L) = \inf \left\{ \theta > 0 : \ \mathrm{dist}(\theta v, \mathbb{Z}^N) < L \sqrt{\log_+ \frac{\|\theta v\|_2}{L}} \right\}.$$

For a matrix $V \in \mathbb{R}^{m \times N}$*, the least common denominator is defined as*

$$D(V) = D(V, L) = \inf \left\{ \|\theta\|_2 : \ \theta \in \mathbb{R}^m, \ \mathrm{dist}(V^\mathsf{T} \theta, \mathbb{Z}^N) < L \sqrt{\log_+ \frac{\|V^\mathsf{T} \theta\|_2}{L}} \right\}.$$

Remark 7.2. The definition of LCD for vectors is a special case of the definition for matrices with $m = 1$. This can be seen by considering a vector $v \in \mathbb{R}^N$ as a $1 \times N$ matrix.

Remark 7.3. In applications, we will typically choose $L \sim \sqrt{m}$, so for vectors we usually choose $L \sim 1$.

Before relating the concept of LCD to small ball probabilities, let us pause to note a simple but useful lower bound for LCD. To state it, for a given matrix V we let $\|V\|_\infty$ denote the maximum

Euclidean norm of the columns of V. Note that for vectors ($1 \times N$ matrices), this quantity is the usual ℓ_∞ norm.

Proposition 7.4 (Simple lower bound for LCD)**.** *For every matrix V and $L > 0$, one has*

$$D(V, L) \geq \frac{1}{2\|V\|_\infty}.$$

Proof. By definition of LCD, it is enough to show that for $\theta \in \mathbb{R}^m$, the inequality

$$\operatorname{dist}(V^\mathsf{T}\theta, \mathbb{Z}^N) < L\sqrt{\log_+ \frac{\|V^\mathsf{T}\theta\|_2}{L}} \tag{7.1}$$

implies $\|\theta\|_2 \geq 1/(2\|V\|_\infty)$. Assume the contrary, that there exists θ which satisfies (7.1) but for which

$$\|\theta\|_2 < \frac{1}{2\|V\|_\infty}. \tag{7.2}$$

We can use Cauchy-Schwartz inequality and (7.2) to bound all coordinates $\langle V_j, \theta \rangle$ of the vector $V^\mathsf{T}\theta$ as follows:

$$|\langle V_j, \theta \rangle| \leq \|V\|_\infty \|\theta\|_2 < \frac{1}{2}, \quad j = 1, \ldots, N.$$

(Here $V_j \in \mathbb{R}^m$ denote the columns of the matrix V.) This bound means that each coordinate of $V^\mathsf{T}\theta$ is closer to zero than to any other integer. Thus the vector $V^\mathsf{T}\theta$ itself is closer (in the ℓ_2 norm) to the origin than to any other integer vector in $\mathbb{Z}^N$. This implies that

$$\operatorname{dist}(V^\mathsf{T}\theta, \mathbb{Z}^N) = \|V^\mathsf{T}\theta\|_2.$$

Substituting this into (7.1) and dividing both sides by L, we obtain

$$u \leq \sqrt{\log_+ u} \quad \text{where} \quad u = \|V^\mathsf{T}\theta\|_2 / L.$$

But this inequality has no solutions for $u > 0$. This contradiction completes the proof. □

7.2. **Small ball probabilities via LCD.** The following theorem relates small ball probabilities to arithmetic structure, which is measured by LCD. It is a general version of results from [28, 30, 43].

Theorem 7.5 (Small ball probabilities via LCD)**.** *Consider a random vector $\xi = (\xi_1, \ldots, \xi_N)$, where ξ_k are i.i.d. copies of a real-valued random variable ξ satisfying (1.2). Consider a matrix $V \in \mathbb{R}^{m \times N}$. Then for every $L \geq \sqrt{m/p}$ we have*

$$\mathcal{L}(V\xi, t\sqrt{m}) \leq \frac{(CL/\sqrt{m})^m}{\det(VV^\mathsf{T})^{1/2}} \Big(t + \frac{\sqrt{m}}{D(V)}\Big)^m, \quad t \geq 0. \tag{7.3}$$

Proof. We shall apply Esseen's inequality for the small ball probabilities of a general random vector $Y \in \mathbb{R}^m$. It states that

$$\mathcal{L}(Y, \sqrt{m}) \leq C^m \int_{B(0,\sqrt{m})} |\phi_Y(\theta)| \, d\theta \tag{7.4}$$

where $\phi_Y(\theta) = \mathbb{E} \exp(2\pi i \langle \theta, Y \rangle)$ is the characteristic function of Y and $B(0, \sqrt{m})$ is the Euclidean ball centered at the origin and with radius $\sqrt{m}$.

Let us apply Esseen's inequality for $Y = t^{-1} V\xi$, assuming without loss of generality that $t > 0$. Denoting the columns of V by V_k, we express

$$\langle \theta, Y \rangle = \sum_{k=1}^N t^{-1} \langle \theta, V_k \rangle \xi_k.$$

By independence of ξ, this yields

$$\phi_Y(\theta) = \prod_{k=1}^N \phi_k(t^{-1}\langle \theta, V_k\rangle), \quad \text{where} \quad \phi_k(\tau) = \mathbb{E}\exp(2\pi i \tau \xi_k)$$

are the characteristic functions of ξ_k. Therefore, Esseen's inequality (7.4) yields

$$\mathcal{L}(V\xi, t\sqrt{m}) = \mathcal{L}(Y, \sqrt{m}) \le C^m \int_{B(0,\sqrt{m})} \prod_{k=1}^N \big|\phi_k(t^{-1}\langle \theta, V_k\rangle)\big|\, d\theta. \tag{7.5}$$

Now we evaluate the characteristic functions that appear in this integral. First we apply a standard symmetrization argument. Let ξ' be an independent copy of ξ, and consider the random vector $\bar{\xi} := \xi - \xi'$. Its coordinates $\bar{\xi}_k$ are i.i.d. random variables with symmetric distribution. It follows that

$$|\phi_k(\tau)|^2 = \mathbb{E}\exp(2\pi i \tau \bar{\xi}_1) = \mathbb{E}\cos(2\pi\tau\bar{\xi}_1) \quad \text{for all } k.$$

Using the inequality $x \le \exp(-\frac{1}{2}(1-x^2))$ that is valid for all $x \ge 0$, we obtain

$$|\phi_k(\tau)| \le \exp\Big[-\frac{1}{2}\,\mathbb{E}\big[1-\cos(2\pi\tau\bar{\xi}_1)\big]\Big]. \tag{7.6}$$

The assumptions on ξ_k imply that the event $\{1 \le |\bar{\xi}_1| \le K\}$ holds with probability at least $p/2$. Denoting by $\bar{\mathbb{E}}$ the conditional expectation on that event, we obtain

$$\mathbb{E}\big[1-\cos(2\pi\tau\bar{\xi}_1)\big] \ge \frac{p}{2}\bar{\mathbb{E}}\big[1-\cos(2\pi\tau\bar{\xi}_1)\big] \ge 4p\cdot\bar{\mathbb{E}}\min_{q\in\mathbb{Z}}|\tau\bar{\xi}_1 - q|^2 \tag{7.7}$$

where the last bound follows from the elementary inequality

$$1-\cos(2\pi y) \ge 8\min_{q\in\mathbb{Z}}|y-q|^2, \quad y \in \mathbb{R}.$$

Substituting this inequality into (7.6) and then back into (7.5), we further derive

$$\begin{aligned}\mathcal{L}(V\xi, t\sqrt{m}) &\le C^m \int_{B(0,\sqrt{m})} \exp\Big[-2p\,\bar{\mathbb{E}}\sum_{k=1}^N \min_{q_k\in\mathbb{Z}} \big|t^{-1}\bar{\xi}_1\langle\theta, V_k\rangle - q_k\big|^2\Big]\, d\theta \\ &= C^m \int_{B(0,\sqrt{m})} \exp\big(-2pf(\theta)^2\big)\, d\theta \end{aligned}\tag{7.8}$$

where

$$f(\theta)^2 := \bar{\mathbb{E}}\min_{q\in\mathbb{Z}^N}\Big\|t^{-1}\bar{\xi}_1 V^{\mathsf{T}}\theta - q\Big\|_2^2 = \bar{\mathbb{E}}\,\mathrm{dist}(t^{-1}\bar{\xi}_1 V^{\mathsf{T}}\theta, \mathbb{Z}^N)^2.$$

The least common denominator $D(V,L)$ will help us estimate the distance to the integer lattice that appears in the definition of $f(\theta)$. Let us first assume that

$$t \ge t_0 := \frac{2K\sqrt{m}}{D(V,L)}, \tag{7.9}$$

or equivalently that $D(V,L) \ge 2K\sqrt{m}/t$. Then for any θ appearing in the integral (7.8), that is for $\theta \in B(0,\sqrt{m})$, one has

$$\|t^{-1}\bar{\xi}_1\theta\|_2 \le Kt^{-1}\sqrt{m} < D(V).$$

(Here we used that $|\bar{\xi}_1| \le K$ holds on the event over which the conditional expectation $\bar{\mathbb{E}}$ is taken.) By the definition of $D(V)$, this implies that

$$\mathrm{dist}(V^{\mathsf{T}}(t^{-1}\bar{\xi}_1\theta), \mathbb{Z}^N) \ge L\sqrt{\log_+ \frac{\|V^{\mathsf{T}}(t^{-1}\bar{\xi}_1\theta)\|_2}{L}}.$$

Recalling the definition of f and using that $|\bar{\xi}_1| \geq 1$ on the event over which the conditional expectation $\bar{\mathbb{E}}$ is taken, we obtain

$$f(\theta)^2 \geq L^2 \log_+ \frac{\|V^\mathsf{T}\theta\|_2}{Lt}.$$

where in the second inequality we use that $|\bar{\xi}_1| \geq 1$ on the event over which the conditional expectation $\bar{\mathbb{E}}$ is taken. Substituting this bound into (7.8), we obtain

$$\mathcal{L}(V\xi, t\sqrt{m}) \leq C^m \int_{B(0,\sqrt{m})} \exp\Big(- 2pL^2 \log_+ \frac{\|V^\mathsf{T}\theta\|_2}{Lt}\Big)\, d\theta.$$

One can estimate this integral in a standard way.

Let us get rid of V in the integrand by an appropriate change of variable. Using a singular value decomposition of V, one can replace $V^\mathsf{T}\theta$ by $\Sigma\theta$ where $\Sigma \in \mathbb{R}^{m\times m}$ is a diagonal matrix with singular values of V on the diagonal. Next, we change variables to $\Sigma\theta/Lt = z$. Since $\det \Sigma = \det(VV^\mathsf{T})^{1/2}$, this yields

$$\mathcal{L}(V\xi, t\sqrt{m}) \leq \frac{(CLt)^m}{\det(VV^\mathsf{T})^{1/2}} \int_{\mathbb{R}^m} \exp\Big(- 2pL^2 \log_+ \|z\|_2\Big)\, dz. \tag{7.10}$$

We evaluate the integral by breaking it into two parts:

$$\int_{\mathbb{R}^m} \exp\Big(- 2pL^2 \log_+ \|z\|_2\Big)\, dz = \int_{B(0,1)} 1\, dz + \int_{B(0,1)^c} \|z\|_2^{-2pL^2}\, dz. \tag{7.11}$$

Let us start with the second integral. Passing to the polar coordinates $(r,\phi) \in \mathbb{R}_+ \times S^{m-1}$ where $dz = r^{m-1}\, dr\, d\phi$, we obtain for any $q \geq 0$ that

$$\int_{B(0,1)^c} \|z\|_2^{-q}\, dz = \int_1^\infty dr \int_{S^{m-1}} r^{-q} r^{m-1}\, d\phi = \sigma_{m-1}(S^{m-1}) \int_1^\infty r^{m-q-1}\, dr,$$

where $\sigma_{m-1}(S^{m-1})$ is the surface area of the unit sphere. Recall that

$$\sigma_{m-1}(S^{m-1}) = \frac{2\pi^{m/2}}{\Gamma(m/2)} \leq \Big(\frac{C}{\sqrt{m}}\Big)^m$$

and that

$$\int_1^\infty r^{m-q-1}\, dr = \frac{1}{q-m} \leq 1 \quad \text{for} \quad q \geq 2m.$$

This yields

$$\int_{B(0,1)^c} \|z\|_2^{-q}\, dz \leq \Big(\frac{C}{\sqrt{m}}\Big)^m \quad \text{for} \quad q \geq 2m.$$

We use this bound for $q = 2pL^2$, where $q \geq 2m$ by assumption. It follows that the integral over $B(0,1)^c$ in (7.11) is bounded by $(C/\sqrt{m})^m$. Moreover, the integral over $B(0,1)$ equals the volume of the unit ball $B(0,1)$, which is also bounded by $(C/\sqrt{m})^m$. Thus the right hand side of (7.11) is bounded by $2(C/\sqrt{m})^m$. Substituting it into (7.10), we obtain

$$\mathcal{L}(V\xi, t\sqrt{m}) \leq \frac{2(CLt)^m}{\det(VV^\mathsf{T})^{1/2}} \Big(\frac{C}{\sqrt{m}}\Big)^m. \tag{7.12}$$

This completes the proof in the case where $t \geq t_0$ as specified in (7.9).

In the opposite case where $t \leq t_0$, it is enough to use that $\mathcal{L}(V\xi, t\sqrt{m}) \leq \mathcal{L}(V\xi, t_0\sqrt{m})$ and apply the inequality (7.12) for t_0. This completes the proof of Theorem 7.5. $\square$

7.3. **Special cases: sums of independent random variables.** Let us state an immediate consequence of Theorem 7.5 in the important special case where $m = 1$. In this case, $V\xi$ becomes a sum of independent random variables.

Corollary 7.6 (Small ball probabilities for sums)**.** *Let ξ_k be i.i.d. copies of a real-valued random variable ξ satisfying (1.2). Let $a = (a_1, \ldots, a_N) \in \mathbb{R}^n$. Then for every $L \geq 1/\sqrt{p}$ we have*

$$\mathcal{L}\Big(\sum_{k=1}^{N} a_k \xi_k, t\Big) \leq \frac{CL}{\|a\|_2}\Big(t + \frac{1}{D(a,L)}\Big), \quad t \geq 0.$$

This corollary was proved in [43]; similar versions appeared in [28, 30]. It gives a non-trivial probability bound when the coefficient vector a is sufficiently unstructured, i.e. when $D(a, L)$ is large enough. In the situation where no information is known about the structure of a, the following result can be useful.

Lemma 7.7 (Small ball probabilities: a simple bound)**.** *Let ξ_k be independent random variables satisfying (1.2), and let a_j be real numbers such that $\sum_{j=1}^N a_j^2 = 1$. Then*

$$\mathcal{L}\Big(\sum_{k=1}^{N} a_k \xi_k, c\Big) \leq 1 - c'$$

where c and c' are positive numbers that may only depend on p and K.

Proof. We will consider separately the cases where a has a large coordinate and where it does not. Assume first that

$$\|a\|_\infty \geq \frac{1}{4CL} =: \nu$$

where C is the constant appearing in Corollary 7.6. Choose a coordinate k_0 such that $|a_{k_0}| = \|a\|_\infty$. Applying Lemma 3.2, we obtain

$$\mathcal{L}\Big(\sum_{k=1}^{N} a_k \xi_k, \nu\Big) \leq \mathcal{L}\Big(a_{k_0}\xi_{k_0}, \nu\Big) \leq \mathcal{L}(\xi_{k_0}, 1) \leq 1 - p.$$

In the opposite case where $\|a\|_\infty < \nu$, Proposition 7.4 implies $D(a, L) \geq 1/(2\nu)$. Combining this with Corollary 7.6, we obtain

$$\mathcal{L}\Big(\sum_{j=1}^{N} a_j \xi_j, \nu\Big) \leq CL \cdot 3\nu \leq \frac{3}{4},$$

which completes the proof. □

7.4. **Special cases: projections of random vectors.** Another class of examples where Theorem 7.5 is useful is for projections of a random vector ξ onto a fixed subspace E of $\mathbb{R}^N$. Equivalently, this result allows us to estimate the distances between random vectors and fixed subspaces, since $\mathrm{dist}(X, H) = \|P_{H^\perp} X\|_2$.

To deduce such estimates, we will make the matrix V in Theorem 7.5 encode an orthogonal projection onto E. Let us pause to interpret the LCD of such matrix V as the LCD of the subspace E itself.

Definition 7.8 (LCD of a subspace)**.** *Fix $L > 0$. For a subspace $E \subseteq \mathbb{R}^N$, the least common denominator is defined as*

$$D(E) = D(E, L) = \inf\{D(v, L) : \; v \in S_E\}.$$

By now, we have defined LCD of vectors, matrices, and subspaces. The following lemma relates them together.

Lemma 7.9 (LCD of subspaces vs. matrices)**.** *Let E be a subspace of $\mathbb{R}^N$. Then*

(1) $D(E) = \inf\left\{\|x\|_2 :\ x \in E,\ \operatorname{dist}(x, \mathbb{Z}^N) < L\sqrt{\log_+ \frac{\|x\|_2}{L}}\right\}$.

(2) Let $U \in \mathbb{R}^{N\times m}$ be a matrix such that $U^\mathsf{T} U = I_m$ and $\operatorname{Im}(U) = E$. Then $D(E) = D(U^\mathsf{T})$.

Proof. The first part follows directly from the definition. To prove the second part, note that according to Definition 7.1 we have

$$D(U^\mathsf{T}) = \inf\left\{\|\theta\|_2 :\ \theta \in \mathbb{R}^m,\ \operatorname{dist}(U\theta, \mathbb{Z}^N) < L\sqrt{\log_+ \frac{\|U\theta\|_2}{L}}\right\}.$$

Let us change variable to $x = U\theta$. The assumptions on U imply that $\|x\|_2 = \|\theta\|_2$ and as θ runs over $\mathbb{R}^m$, x runs over $\operatorname{Im}(U) = E$. We finish by applying the first part of this lemma. □

The following corollary is a version of a result from [30].

Corollary 7.10 (Small ball probabilities for projections)**.** *Consider a random vector $\xi = (\xi_1, \dots, \xi_N)$, where ξ_k are i.i.d. copies of a real-valued random variable ξ satisfying (1.2). Let E be a subspace of $\mathbb{R}^N$ with $\dim(E) = m$, and let P_E denote the orthogonal projection onto E. Then for every $L \geq \sqrt{m/p}$ we have*

$$\mathcal{L}(P_E\xi, t\sqrt{m}) \leq \Big(\frac{CL}{\sqrt{m}}\Big)^m \Big(t + \frac{\sqrt{m}}{D(E,L)}\Big)^m, \quad t \geq 0. \tag{7.13}$$

Proof. Choose a matrix $U \in \mathbb{R}^{N\times m}$ so that $U^\mathsf{T} U = I_m$ and $UU^\mathsf{T} = P_E$. Then U acts as an isometric embedding from $\mathbb{R}^m$ into $\mathbb{R}^N$, i.e. $\|Ux\|_2 = \|x\|_2$ for all $x \in \mathbb{R}^m$. This yields

$$\mathcal{L}(P_E\xi, t\sqrt{m}) = \mathcal{L}(UU^\mathsf{T}\xi, t\sqrt{m}) = \mathcal{L}(U^\mathsf{T}\xi, t\sqrt{m}).$$

We apply Theorem 7.5 for $V = U^\mathsf{T}$ and note that $\det(VV^\mathsf{T}) = \det(U^\mathsf{T} U) = \det(I_m) = 1$. Thus $\mathcal{L}(P_E\xi, t\sqrt{m})$ gets bounded by the same quantity as in the right hand side of (7.13) except for $D(V)$. It remains to use Lemma 7.9, which yields $D(V) = D(U^\mathsf{T}) = D(E)$. □

8. Distances between random vectors and subspaces: statement of the result

Our next goal is to prove a lower bound for the distance between independent random vectors and subspaces. For continuous distributions, this was achieved in Lemma 5.4. Doing this for general, possibly discrete, distributions, is considerably more difficult. The following result is a version of Lemma 5.4 for general distributions.

Theorem 8.1 (Distance between random vectors and subspaces)**.** *Let $H \in \mathbb{C}^{N\times n}$ be a random matrix which satisfies Assumptions 1.1[4] and 1.4, and assume that $n = (1-\varepsilon)N$ for some $\varepsilon \in (2/n, c)$. Let $Z \in \mathbb{C}^N$ be a random vector independent of H, and whose coordinates are i.i.d. random variables satisfying the same distributional assumptions as specified in Assumption 1.4. Then*

$$\mathbb{P}\left\{\operatorname{dist}(Z, \operatorname{Im}(H)) \leq \tau\sqrt{\varepsilon N} \text{ and } \mathcal{B}_{H,M}\right\} \leq \left[C\Big(\tau + \frac{1}{\sqrt{\varepsilon N}} + e^{-c/\sqrt{\varepsilon}}\Big)\right]^{\varepsilon N}, \quad \tau \geq 0.$$

[4]Assumption 1.1 is formulated for square random matrices. For rectangular matrices, one of the entries A_{ij} or A_{ji} may not exist. In this case, we assume that the other entry is independent of the rest.

A version of this theorem was proved in [30] in the simpler situation where the entries of H are real-valued and all independent. In this simpler case, [30] gives the following optimal bound:

$$\mathbb{P}\left\{\operatorname{dist}(Z, \operatorname{Im}(H)) \le \tau\sqrt{\varepsilon N}\right\} \le (C\tau)^{\varepsilon N} + e^{-cN}.$$

We do not know if the same bound can be proved in the setting of Theorem 8.1.

To prove Theorem 8.1, we will first reduce it to a problem over reals – much like we did in Section 5.3.3. Then, expressing the distance $\operatorname{dist}(Z, \operatorname{Im}(H))$ as the norm of the projection of Z onto $\operatorname{Im}(H)^\perp = \ker(H^*)$, we should be able to apply Corollary 7.10. However, for the resulting probability bound (7.13) to be meaningful, we would need to show that the least common denominator $D(\ker(H^*), L)$ is large, or in other words, that the subspace H is unstructured. This will be a major step in the argument. Eventually we will achieve this in Section 12, which will allow us to quickly finalize the proof of Theorem 8.1.

In preparation for the proof of Theorem 8.1, let us express the distance we need to estimate as follows:

$$\operatorname{dist}(Z, \operatorname{Im}(H)) = \|P_{\operatorname{Im}(H)^\perp} Z\|_2 = \|P_{\ker(B)} Z\|_2, \quad \text{where } B = H^* \in \mathbb{C}^{n\times N}. \tag{8.1}$$

Our goal is to show that $\ker(B)$ is arithmetically unstructured.

8.1. **Transferring the problem from $\mathbb{C}$ to $\mathbb{R}$.** Similarly to our argument for continuous distributions, we will now transfer the distance problem from the complex to the real field. In Section 5.3.3, we introduced the operation $z \mapsto \widetilde{z}$ that makes a complex vector $z = x + iy$ in $\mathbb{C}^N$ real by defining $\widetilde{z} := \binom{x}{y} \in \mathbb{R}^{2N}$. We also introduced this operation for subspaces E of $\mathbb{C}^N$ by defining $\widetilde{E} = \{\widetilde{z} : \ z \in E\} \subset \mathbb{R}^{2N}$.

In the analysis of the distance problem for continuous distributions, we did not need to know anything about the subspaces H'_j beyond their dimensions. This time, our analysis will be sensitive to the structure of the subspace $\ker(B)$. For this purpose, we will need to transfer the matrix B from complex to real field. We can do this in a way that preserves matrix-vector multiplication as follows:

$$\text{For } B = R + iT \in \mathbb{C}^{n\times N}, \text{ define } \widetilde{B} = \begin{bmatrix} R & -T \\ T & R \end{bmatrix} \in \mathbb{R}^{2n\times 2N}. \tag{8.2}$$

We already observed two elementary properties of the operation $z \mapsto \widetilde{z}$ in Lemma 5.2; let us record one more straightforward fact.

Lemma 8.2 (Elementary property of operation $x \mapsto \widetilde{x}$)**.** *For a complex matrix B and a vector z, one has $\widetilde{Bz} = \widetilde{B}\widetilde{z}$, and consequently $\widetilde{\ker(B)} = \ker(\widetilde{B})$.*

Let us return to the distance problem (8.1). Applying Lemma 5.3 for $E = \ker B$, we conclude that

$$\mathcal{L}(P_{\ker(B)} Z, r) \le \mathcal{L}(P_{\widetilde{\ker(B)}} \widehat{Z}, 2r)^{1/2}.$$

Using the interpretation of distance as norm of projection in (8.1), we can summarize the first step toward the proof of Theorem 8.1. We showed that

$$\mathbb{P}\left\{\operatorname{dist}(Z, \operatorname{Im}(H)) \le \tau\sqrt{\varepsilon N}\right\} \le \mathcal{L}(P_{\widetilde{\ker(B)}} \widehat{Z}, 2\tau\sqrt{\varepsilon N})^{1/2}. \tag{8.3}$$

Recall that here, according to Lemma 8.2, $\widetilde{\ker B} = \ker \widetilde{B}$, where $\widetilde{B}$ is the random matrix from (8.2) and $\widehat{Z} \in \mathbb{R}^{2N}$ is a random vector. Specifically, T is a fixed $n \times N$ matrix and R is an $n \times N$ random matrix, which satisfies the structural and distributional requirements of Assumptions 1.1 and 1.4 (except that R entirely real). The coordinates of the random vector $\widehat{Z}$ are i.i.d. copies of a real random variable ξ satisfying (1.2).

9. Kernels of random matrices are incompressible

9.1. **Compressible and incompressible vectors.** Before we can show that the kernel of B consists of arithmetically unstructured vectors, we will prove a much simpler result. It states that the kernel of B consists of *incompressible vectors* – those whose mass is not concentrated on a small number of coordinates. The partition of the space into compressible and incompressible vectors has been instrumental in arguments leading to invertibility random matrices, see [28, 30, 43]. A similar splitting of the sphere was used in [24], and in some form the idea of using such partitions goes back to [22].

Definition 9.1 (Compressible and incompressible vectors)**.** *Let $c_0, c_1 \in (0,1)$ be two constants. A vector $z \in \mathbb{C}^N$ is called* sparse *if $|\operatorname{supp}(z)| \le c_0 N$. A vector $z \in S_{\mathbb{C}}^{N-1}$ is called* compressible *if x is within Euclidean distance c_1 from the set of all sparse vectors. A vector $z \in S_C^{N-1}$ is called* incompressible *if it is not compressible. The sets of compressible and incompressible vectors in $S_{\mathbb{C}}^{N-1}$ will be denoted by* Comp *and* Incomp *respectively.*

The definition above depends on the choice of the constants c_0, c_1. These constants will be chosen in Proposition 9.4 and remain fixed throughout the paper.

As we already announced, our goal in this section is to prove that, with high probability, the kernel of B consists entirely of incompressible vectors. We will deduce this by providing a uniform lower bound for $\|Bz\|_2$ for all compressible vectors z.

9.2. **Relating $\|Bz\|_2$ to a sum of independent random variables.** Let us fix a vector z for now. We would like to express $\|Bz\|_2^2$ as a sum of independent random variables, and then to use bounds on small ball probabilities from Section 7. Using the real version $\widetilde{B}$ of the matrix B, and the real version $\widetilde{z}$ of the vector z we introduced in Section 8.1, we can write

$$\|Bz\|_2^2 = \|\widetilde{B}\widetilde{z}\|_2^2 = \|Rx + Ty\|_2^2 + \|Ry - Tx\|_2^2. \tag{9.1}$$

Let us fix a subset $J \subset [N]$. Dropping the coefficients of the vectors $Rx+Ty$ and $Ry-Tx$ indexed by J, we obtain

$$\|Bz\|_2^2 \ge \|R_{J^c \times [n]} x + a\|_2^2 + \|R_{J^c \times [n]} y - b\|_2^2,$$

where $a = T_{J^c \times [n]} y$ and $b = T_{J^c \times [n]} x$ are fixed vectors.

Further, let us decompose $R_{J^c \times [n]} = R_{J^c \times J} + R_{J^c \times J^c}$, where $R_{J^c \times I}$ is obtained from the matrix $R_{J^c \times [n]}$ by replacing the columns that do not belong to I by zeros. Assumption 1.1 implies that these two components are independent, and moreover the first one, $R_{J^c \times J}$, has independent entries. So let us condition on the second component, $R_{J^c \times J^c}$. Absorbing its contribution into a and b, we obtain

$$\|Bz\|_2^2 \ge \|R_{J^c \times J} x + a'\|_2^2 + \|R_{J^c \times J} y - b'\|_2^2,$$

where a' and b' are fixed vectors. Expanding the matrix-vector multiplication, we arrive at the bound

$$\|Bz\|_2^2 \ge \sum_{i \in [n] \setminus J} X_i^2 + Y_i^2, \tag{9.2}$$

where

$$X_i = \sum_{j \in J} R_{ij} x_j + a'_j, \quad Y_i = \sum_{j \in J} R_{ij} y_j - b'_j \tag{9.3}$$

and a'_j and b'_j are fixed numbers. The sum in (9.2) should be convenient to control, since all R_{ij} appearing in (9.3) are independent random variables.

9.3. **A lower bound on $\|Bz\|_2$ for compressible vectors.** We start with a simple and general lower bound on $\|Bz\|_2$ for a fixed vector z.

Proposition 9.2 (Matrix acting on a fixed vector: simple bound). *Let $n \le N \le 2n$, and $B \in \mathbb{C}^{n\times N}$ be a random matrix satisfying Assumptions 1.1 and 1.4. Then for any fixed vector $z \in \mathbb{C}^N$ with $\|z\|_2 = 1$ we have*

$$\mathbb{P}\left\{\|Bz\|_2 \le c\sqrt{n}\right\} \le e^{-cn}.$$

Proof. Let $z = x + iy$, and choose J to be the set of indices of the $N/4$ largest coordinates of z. Since z is a unit vector, we have

$$\|x_J\|_2^2 + \|y_J\|_2^2 = \|z_J\|_2^2 \ge \frac{1}{4}.$$

It follows that either x_J or y_J has norm at least $1/4$. Without loss of generality, let us assume that $\|x_J\|_2 \ge 1/4$.

Dropping the terms Y_j from (9.2), we see that

$$\|Bz\|_2^2 \ge \sum_{i\in[n]\setminus J} X_j^2 \quad \text{where} \quad X_j = \sum_{j\in J} R_{ij}x_j + a'_j. \tag{9.4}$$

By Assumption 1.1, R_{ij} are i.i.d. random variables. Moreover, their distribution satisfies Assumption 1.4, so we can apply Lemma 7.7 and conclude that for each j,

$$\mathbb{P}\left\{|X_j| \le c\right\} \le 1 - c'. \tag{9.5}$$

Assume that $\|Bz\|_2^2 \le \alpha c^2 n$ where $\alpha \in (0,1)$ is a number to be chosen later. By (9.4), this yields $\sum_{i\in[n]\setminus J} X_j^2 \le \alpha c^2 n$, which in turn implies that $X_j \le c$ for at least $|[n]\setminus J| - \alpha n$ random variables X_j in this sum. Therefore, using independence we obtain

$$\mathbb{P}\left\{\|Bz\|_2^2 \le \alpha c^2 n\right\} \le \binom{|[n]\setminus J|}{\alpha n} \cdot (1-c')^{|[n]\setminus J| - \alpha n} \le \left(\frac{e}{\alpha}\right)^{\alpha n} \cdot (1-c')^{(1/2-\alpha)n}. \tag{9.6}$$

The second inequality holds if we choose α small enough so that $\alpha n \le n/4$, while $|[n]\setminus J| \ge n - N/4 \ge n/2$ by assumption. The probability bound in (9.6) can be made smaller than $e^{-\bar{c}n}$ for some $\bar{c} = \bar{c}(c') > 0$ by choosing $\alpha = \alpha(c') > 0$ sufficiently small. This completes the proof. $\square$

We are going to argue that the lower bound in Proposition 9.2 holds not only for a fixed unit vector z but also uniformly over $z \in \mathrm{Comp}$. This will follow by combining Proposition 9.2 with the following standard construction on a net for the set of compressible vectors.

Lemma 9.3 (Net for compressible vectors). *For any $\delta \in (0,1)$, there exists a $(2c_1)$-net of the set of* Comp *of cardinality at most*

$$\left(\frac{C}{c_0 c_1^2}\right)^{c_0 N}.$$

Proof. First we construct a c_1-net of the set of sparse vectors. This set is a union of coordinate subspheres of $S_{\mathbb{C}}^J$ for all sets $J \subset [N]$ of cardinality $c_0 N$. For a fixed J, the standard volume argument yields a c_1-net of $S_{\mathbb{C}}^J$ of cardinality at most $(3/c_1)^{2c_0 N}$. (This is because the dimension of this sphere over $\mathbb{R}$ is $2c_0 N$.) A union bound over $\binom{N}{c_0 N} \le (e/c_0)^{c_0 N}$ choices of J produces a c_1-net of the set of sparse vectors with cardinality at most $(C/c_0 c_1^2)^{c_0 N}$. By approximation, this is automatically a $(2c_1)$-net for the set of compressible vectors. $\square$

Proposition 9.4 (A lower bound on the set of compressible vectors)**.** *Let $B \in \mathbb{C}^{n\times N}$ be a random matrix satisfying Assumptions 1.1 and 1.4. Then one can choose constants $c_0, c_1 \in (0,1)$ in Definition 9.1 depending on p and K only, and so that*

$$\mathbb{P}\left\{ \inf_{z\in \text{Comp}} \|Bz\|_2 \le c\sqrt{n} \text{ and } \mathcal{B}_{B,M} \right\} \le e^{-cn}.$$

Proof. Let us choose $c_1 = c/(4M)$ and let $\mathcal{N}$ be a $(2c_1)$-net of the set Comp given by Lemma 9.3.

Assume the bad event in Proposition 9.4 occurs, so thus $\|Bz\|_2 \le c\sqrt{n}$ for some $z \in \text{Comp}$ and $\|B\| \le M\sqrt{n}$. Choose $z_0 \in \mathcal{N}$ such that $\|z - z_0\|_2 \le 2c_1$. By triangle inequality, we have

$$\|Bz_0\|_2 \le \|Bz\|_2 + \|B\| \, \|z - z_0\|_2 \le c\sqrt{n} + M\sqrt{N} \cdot 2c_1 \le 2c\sqrt{n}.$$

In the last inequality, we used the definition of c_1 and the fact that $N \le 2n$.

Furthermore, Proposition 9.2 states that for fixed z_0, the inequality $\|Bz_0\|_2 \le 2c\sqrt{n}$ holds with probability at most e^{-cn}. Combining this with the union bound over $z_0 \in \mathcal{N}$ and using the cardinality of $\mathcal{N}$ given by Lemma 9.3, we conclude that the bad event in Proposition 9.4 holds with probability at most

$$e^{-cn} \cdot \left(\frac{C}{c_0 c_1^2} \right)^{c_0 N}.$$

Choosing c_0 so that the last expression does not exceed $e^{-c_0 n/2}$ completes the proof. $\square$

Proposition 9.4 implies in particular that with high probability the kernel of B consists of incompressible vectors:

$$\ker B \cap S^{N-1} \subseteq \text{Incomp}.$$

10. Small ball probabilities via real-imaginary correlations

Recall that our big goal is to show that the kernel of B is unstructured, which means that all vectors in $\ker(B)$ have large LCD. We may try to approach this problem using the same general line of attack as in Section 9. Namely, we can try to bound $\|Bz\|_2$ below uniformly on the set of vectors with small LCD.

This will require us to considerably sharpen the tools we developed in Section 9 – small ball probabilities and constructions of nets. More precisely, we would like to make the probability in Proposition 9.2 exponential in $2n$ rather than n; an ideal bound for us would be

$$\mathbb{P}\left\{ \|Bz\|_2 \le t\sqrt{n} \right\} \le \left(Ct + \frac{1}{\sqrt{n}} \right)^{2n}, \quad t \ge 0. \tag{10.1}$$

A bound like this will be crucial when we combine it with a union bound over a net, just like in Section 9. But there the nets were for compressible vectors $z \in \mathbb{C}^N$. Now we will have to handle much larger sets: *the level sets of LCD*. As we will describe in Section 11, the nets of these level sets are exponential in $2N$. To control them, it is crucial to have the small probability bound that is also exponential in $2n$. (The difference between $2N$ and $2n$ is minor and can intuitively be neglected since $N = (1+\varepsilon)n$.)

At first glance, this should be possible because our problem is over $\mathbb{C}$, so the dimension there should double compared to $\mathbb{R}$. But recall that according to Assumption 1.1, the imaginary part of B is fixed, so there is no extra randomness that could help us double the exponent.

One can even come up with concrete examples where the bound (10.1) fails. Assume that the entries of B are *real* independent random variables with bounded densities, and that z is a real vector. Since the matrix R has n rows, the optimal small ball probability is

$$\mathbb{P}\left\{ \|Bz\|_2 \le t\sqrt{n} \right\} \le (Ct)^n. \tag{10.2}$$

The same is true for complex vectors z with very correlated real and imaginary parts, such as for $z = x + ix$.

These observations might lead us to the conclusion that it must be impossible to combine the small ball probabilities with nets. However, one can notice that the examples of vectors z we just considered are special. The real vectors z are contained in the N-dimensional real sphere, and this sphere has a net exponential in N rather than $2N$. The same holds for vectors of the type $z = x + ix$. So these special vectors have smaller nets, which can hopefully be balanced by the small ball probabilities like (10.2).

For other, more "typical" vectors, we might hope for stronger probability bounds. Consider, for example, the vector $z = x + iy$, where x and y have disjoint support and both have norms $\Omega(1)$. Still assuming that B is a real matrix, we then have $\|Bz\|_2^2 = \|Bx\|_2^2 + \|By\|_2^2$. The assumption of disjoint support yields implies that Bx and By are independent, and $\|Bz\|_2^2$ is thus a sum of $2n$ independent random variables (the row-vector products). So we do have a double amount of randomness here, and

$$\mathbb{P}\left\{\|Bz\|_2 \le t\sqrt{n}\right\} \le (Ct)^{2n}.$$

Such probability bounds can balance a net for the whole sphere of $\mathbb{C}^N$, which is exponential in $2N$.

Guided by these examples, we may surmise that the small ball probabilities for Bz and the cardinalities of nets for vectors z both depend on the correlation of real and the imaginary parts of z. Exploring this interaction in search for tight matching bounds for both quantities will be the main technical difficulty in proving Theorem 8.1. We will get a hold of small ball probabilities in the current section, and of cardinalities of nets in Section 11.

10.1. **Toward a more sensitive bound.** We start by representing $\|Bz\|_2^2$ as a sum of independent random variables exactly as in Section 9.2, leading up to (9.2). In a moment, we will apply Littlewood-Offord theory for each term of the sum in (9.2). To do this, we express these terms as functions of the rows of $R_{J^c \times J}$ as follows:

$$\|Bz\|_2^2 \ge \sum_{i \in [n] \setminus J} X_j^2 + Y_j^2 = \sum_{i \in [n] \setminus J} \|V_J (R_i)_J - u_i\|_2^2. \tag{10.3}$$

Here

$$V = \begin{bmatrix} x^\mathsf{T} \\ y^\mathsf{T} \end{bmatrix} \in \mathbb{R}^{2 \times N}$$

is a fixed matrix, R_i^T denotes the i-th row of R, and $u_i \in \mathbb{R}^2$ are fixed vectors.

Note that at this time we have three different ways to represent a complex vector $z \in \mathbb{C}^N$: the usual way $z = x + iy$, as a long real vector $\widetilde{z} = \binom{x}{y} \in \mathbb{R}^{2N}$, and as a $2 \times N$ real matrix V as above.

All $(R_i)_J$ in (10.3) are independent real random vectors with all independent coordinates. We can now apply Theorem 7.5 in dimension $m = 2$ and for $L = 1/\sqrt{2p}$. It yields

$$\mathbb{P}\left\{\|V_J (R_i)_J - u_i\|_2 \le t\right\} \le \frac{C}{\det(V_J V_J^\mathsf{T})^{1/2}} \Big(t + \frac{1}{D_2(V_J)}\Big)^2, \quad t \ge 0. \tag{10.4}$$

Here we use the notation $D_2(V_J)$ to emphasize that the least common denominator used in this application of Theorem 7.5 is for $2 \times |J|$ matrices, as opposed to the one for vectors which we will focus on later.

10.2. **Disregarding the arithmetic structure.** The small ball probability bound (10.4) relies on two different qualities of z. First, the arithmetic structure of z is reflected in the least common denominator $D_2(V_J)$. Second, the correlation between real and imaginary parts of z_J is measured by the term $\det(V_J V_J^\mathsf{T})^{1/2}$.

In this particular place of the argument, we may essentially disregard the arithmetic structure of z. One can get rid of $D_2(V_J)$ using Proposition 7.4, which states that

$$D_2(V_J) \geq \frac{1}{2\|V_J\|_\infty}. \tag{10.5}$$

To bound $\|V_J\|_\infty$, let us introduce a set of small coordinates as follows.

Definition 10.1 (Small coordinates)**.** *Fix $\delta \in (0,1)$ and let $z \in \mathbb{C}^N$. We will denote by $\mathrm{sm}(z)$ the set of indices of all except the δN largest (in the absolute value) coordinates of z. If some of the coordinates of z are equal, the ties are broken arbitrarily.*

If z is a unit vector in $\mathbb{C}^N$ and J is a subset of $\mathrm{sm}(z)$, a simple application of Markov's inequality yields $\|z_J\|_\infty \leq \frac{1}{\sqrt{\delta N}}$. Moreover, by definition of V, we have $\|V_J\|_\infty = \|z_J\|_\infty$. Thus

$$\|V_J\|_\infty \leq \frac{1}{\sqrt{\delta N}}.$$

Substituting this into (10.5), we conclude that

$$D_2(V_J) \geq \frac{1}{2}\sqrt{\delta N}. \tag{10.6}$$

This crude estimate leads to the appearance of the term $1/\sqrt{\varepsilon N}$ in Theorem 8.1. One can probably remove this term by involving the arithmetic structure. However, this would come at a price of a significant increase of the complexity of the argument, so we did not pursue this direction.

10.3. **Quantifying the real-imaginary correlation.** The determinant $\det(V_J V_J^\mathsf{T})^{1/2}$ measures the correlation between real and imaginary parts of z_I. For example, if the real and imaginary parts are equal to each other, then the determinant vanishes, and the small ball probability bound (10.4) becomes useless.

To make the bound as strong as possible, one would choose the subset J so that, on the one hand, it lies in $\mathrm{sm}(z)$ to ensure (10.6), and on the other hand, the determinant $\det(V_J V_J^\mathsf{T})^{1/2}$ is maximized. This motivates the following definition.

Definition 10.2 (Real-complex correlation)**.** *For $z \in \mathbb{C}^N$ and $\delta \in (0,1)$, we define*

$$d(z) = \max\left\{\det(V_J V_J^\mathsf{T})^{1/2} :\ J \subset \mathrm{sm}(z),\ |J| = \delta N\right\}.$$

Clearly, $d(z) \in (0,1)$ for any unit vector z.

Choosing J that achieves the maximum in the definition of $d(z)$ and using the bound (10.6), we conclude from (10.4) that

$$\mathbb{P}\left\{\|V_J(R_i)_J - u_i\|_2 \leq t\right\} \leq \frac{C}{d(z)}\Big(t + \frac{1}{\sqrt{\delta N}}\Big)^2, \quad t \geq 0.$$

Substituting this into (10.3) and using Tensorization Lemma 3.3, we obtain the following result.

Theorem 10.3 (Small ball probabilities via real-imaginary correlation)**.** *Let $B \in \mathbb{C}^{n\times N}$ be a random matrix satisfying Assumptions 1.1 and 1.4, and let $\delta \in (0,1)$. Then for a fixed vector $z \in \mathbb{C}^N$ with $\|z\|_2 = 1$ we have*

$$\mathbb{P}\left\{\|Bz\|_2 \leq t\sqrt{n}\right\} \leq \left[\frac{C}{d(z)}\Big(t + \frac{1}{\sqrt{\delta n}}\Big)^2\right]^{(1-\delta)n}, \quad t \geq 0.$$

10.4. **The essentially real case.** Theorem 10.3 is useful for vectors z whose real-imaginary correlations $d(z)$ are not too small. We wonder what could be done in the "essentially real" case where $d(z)$ happens to be small?

Our strategy will be different in that case. Let us first prove a version of Theorem 10.3 that is not based on $d(z)$, but where t is understandably exponential in $(1-\delta)n$ rather than $2(1-\delta)n$. Such probability bound will hold for incompressible vectors z (which were introduced Definition 9.1), and it will be stronger than the simpler but more general bound of Proposition 9.2.

Theorem 10.4 (Small ball probabilities for general incompressible vectors)**.** *Let $B \in \mathbb{C}^{n\times N}$ be a random matrix satisfying Assumptions 1.1 and 1.4, and let $\delta \in (0, c_0)$. Then for a fixed vector $z \in \mathrm{Incomp}$ we have*

$$\mathbb{P}\left\{\|Bz\|_2 \le t\sqrt{n}\right\} \le \left[\frac{C}{\sqrt{\delta}}\Big(t + \frac{1}{\sqrt{\delta n}}\Big)\right]^{(1-\delta)n}, \quad t \ge 0.$$

Proof. The argument is somewhat simpler than for Theorem 10.3. Consider the set of small coordinates $\mathrm{sm}(z)$ introduced in Definition 10.1. By definition of that set combined with Markov's inequality, and Definition 9.1 of incompressible vectors, we have

$$\|z_{\mathrm{sm}(z)}\|_\infty \le \frac{1}{\sqrt{\delta N}}, \quad \|z_{\mathrm{sm}(z)}\|_2 \ge c_1. \tag{10.7}$$

It follows that there exists a subset $J \subset \mathrm{sm}(z)$ with $|J| = \delta N$ and such that

$$\|z_J\|_\infty \le \frac{1}{\sqrt{\delta N}}, \quad \|z_J\|_2 \ge c_1\sqrt{\delta}. \tag{10.8}$$

(The first inequality is trivial, and the second can be obtained by dividing $\mathrm{sm}(z)$ into $1/\delta - 1$ blocks of coordinates of size δN each, and arguing by contradiction.)

Since $z_J = x_J + iy_J$, either the real part x_J or complex part y_J has ℓ_2-norm bounded below by $c\sqrt{\delta}/2$. Let us assume without loss of generality that x_J satisfies this, so

$$\|x_J\|_\infty \le \frac{1}{\sqrt{\delta N}}, \quad \|x_J\|_2 \ge c'\sqrt{\delta}. \tag{10.9}$$

To control $\|Bz\|_2$, we can proceed similarly to the proof of Proposition 9.2, taking as the starting point the bound

$$\|Bz\|_2^2 \ge \sum_{i\in[n]\setminus J} X_j^2 \quad \text{where} \quad X_j = \sum_{j\in J} R_{ij}x_j + a_j'. \tag{10.10}$$

For each sum defining X_j, we can apply the small ball probability bound of Corollary 7.6 with $L = 1/\sqrt{p}$. This gives

$$\mathbb{P}\left\{|X_j| \le t\right\} \le \frac{C}{\|x_J\|_2}\Big(t + \frac{1}{D(x_J)}\Big).$$

We can use the two inequalities in (10.9) to get rid of the two terms dependent on x_J. Indeed, Proposition 7.4 and the first inequality in (10.9) yield

$$D(x_J) \ge \frac{1}{2}\sqrt{\delta N}.$$

Using this and the second inequality in (10.9) gives

$$\mathbb{P}\left\{|X_j| \le t\right\} \le \frac{C}{\sqrt{\delta}}\Big(t + \frac{1}{\sqrt{\delta n}}\Big).$$

Using this bound for each term of the sum in (10.10) and applying Tensorization Lemma 3.3, we complete the proof. □

Next, we will show that for vectors with small $d(z)$, not only a δN fraction of coordinates but almost the entire real and imaginary parts are close to each other. This strong constraint intuitively means that the set of such vectors is relatively small, and we will indeed construct a small net for such vectors later.

Lemma 10.5 (Real-imaginary correlation)**.** *Let $z \in \mathbb{C}^N$ and set $I := \mathrm{sm}(z)$. Then*

$$\det(V_I V_I^\mathsf{T})^{1/2} \le \frac{Cd(z)}{\delta}.$$

Proof. The argument is based on Cauchy-Binet formula, which yields

$$\det(V_I V_I^\mathsf{T}) = \sum_{I_2 \subset I,\, |I_2|=2} \det(V_{I_2})^2, \tag{10.11}$$

where the sum is over all $\binom{|I|}{2}$ two-element subsets of I. Similarly, for each set J as in the definition of $d(z)$, that is for $J \subset I$, $|J| = \delta N$, we can expand

$$\det(V_J V_J^\mathsf{T}) = \sum_{I_2 \subset J,\, |I_2|=2} \det(V_{I_2})^2.$$

Summing over J, we get

$$\sum_{J \subset I,\, |J|=\delta N} \det(V_J V_J^\mathsf{T}) = \sum_{J \subset I,\, |J|=\delta N} \; \sum_{I_2 \subset J,\, |I_2|=2} \det(V_{I_2})^2.$$

To simplify the right hand side, note that every two-element set $I_2 \subset I$ is included in $\binom{N_0}{\delta N - 2}$ sets J, where we denote $N_0 := |I| = N - \delta N$. Therefore

$$\sum_{J \subset I,\, |J|=\delta N} \det(V_J V_J^\mathsf{T}) = \binom{N_0}{\delta N - 2} \sum_{I_2 \subset I,\, |I_2|=2} \det(V_{I_2})^2.$$

The sum in the right hand side equals $\det(V_I V_I^\mathsf{T})$ by (10.11). Each determinant $\det(V_J V_J^\mathsf{T})$ in the left hand side is bounded by $d(z)^2$ by definition. This yields

$$\binom{N_0}{\delta N} d(z)^2 \ge \binom{N_0}{\delta N - 2} \det(V_I V_I^\mathsf{T}).$$

Simplifying this inequality and using that $N_0 = N - \delta N$, we complete the proof. □

11. A net for vectors with given LCD and real-imaginary correlations

Thanks to Section 9, we can now focus on the set of incompressible vectors. Our goal is to construct a net for the set of incompressible vectors z with given least common denominator $D(\widetilde{z}_{\mathrm{sm}(z)})$ and real-imaginary correlation $d(z)$. Let us define this set formally.

Definition 11.1 (Level set for LCD and real-imaginary correlations)**.** *For $D, d > 0$, we define with the following subset of $\mathbb{C}^N$:*

$$S_{D,d} = \left\{ z \in \mathrm{Incomp} : \; D/2 < D(\widetilde{z}_{\mathrm{sm}(z)}) \le D; \; d(z) \le d \right\}.$$

Hidden in this definition are the parameters L from the definition of $D(\widetilde{z}_{\mathrm{sm}(z)})$ and δ from the definition of $d(z)$, which we assume to be fixed. When we work with level sets $S_{D,d}$, we can assume that

$$D \ge \frac{1}{2}\sqrt{\delta N}. \tag{11.1}$$

This is because by Proposition 7.4 and (10.7), we have

$$D(\widetilde{z}_{\mathrm{sm}(z)}) \geq \frac{1}{2\|\widetilde{z}_{\mathrm{sm}(z)}\|_\infty} \geq \frac{\sqrt{\delta N}}{2}.$$

A first attempt at constructing a small γ-net of the level set $S_{D,d}$ could be to use the standard volume argument. For instance, if one chooses $\gamma = \sqrt{N}/D$, the volume argument will yield a net of cardinality

$$\Big(\frac{D}{\sqrt{N}}\Big)^{2N}. \tag{11.2}$$

The exponent $2N$ appears here because the vectors are complex.

This net is too large for our purposes. Our next, refined, attempt is to leverage the information about LCD of the vectors in $S_{D,d}$. Indeed, known constructions lead to the existence of a finer net, namely with $\gamma \ll \sqrt{N}/D$, and still with approximately the same cardinality as in (11.2), see [30].

However, this net would still be too large if we try to use it in combination with the small ball probability bound given in Theorem 10.3. Our final, successful, refinement of the construction will use both LCD and the real-imaginary correlation $d(z)$ of the vectors in $S_{D,d}$. Ideally, we would hope to construct a γ-net with $\gamma \ll \sqrt{N}/D$ and with cardinality bounded by

$$\Big(\frac{D}{\sqrt{N}}\Big)^{2N} d^N. \tag{11.3}$$

The correction term d^N will allow the net to become smaller for more "real" vectors – those with stronger real-imaginary correlations.

Of course, if d is extremely small, such as for purely real vectors, the cardinality in (11.3) is too good to be true. For purely real vectors, the ideal cardinality would be the same as in (11.2) except with exponent N, that is

$$\Big(\frac{D}{\sqrt{N}}\Big)^{N}. \tag{11.4}$$

Summarizing, we hope to construct a γ-net of the level set $S_{D,d}$ for some $\gamma \ll \sqrt{N}/D$, and with cardinality bounded as in (11.4) if d is not too small (the *genuinely complex* case) as in (11.3) if d is very small (the *essentially real* case). The following theorem, which is the main result of this section, provides slightly weaker but still adequate bounds.

Theorem 11.2 (Nets for level sets)**.** *There exist constants $C, c, \bar{c} > 0$ such that the following holds. Assume that L from the definition of $D(\widetilde{z}_{\mathrm{sm}(z)})$ is such that $L \leq \bar{c}\sqrt{\delta N}$, and δ from the definition of $d(z)$ is such that $\delta \in (0, c)$. Fix $D \geq eL$, and let*

$$\gamma = \frac{L}{D}\sqrt{\log_+ \frac{D}{L}} \quad \text{and} \quad d_0 = C\delta \cdot \max\Big(\gamma, \frac{\sqrt{N}}{D}\Big). \tag{11.5}$$

1. *(Genuinely complex case). For any $d \geq d_0$, there exists a $(C\gamma)$-net of the level set $S_{D,d}$ with cardinality at most*

$$\delta^{-N+1}\gamma^{-2\delta N-1}\Big(\frac{CD}{\sqrt{N}}\Big)^{2N-2\delta N-1} d^{N-\delta N-1}.$$

2. *(Essentially real case). For any $d \leq d_0$, there exists a $(C\gamma)$-net of the level set $S_{D,d}$ with cardinality at most*

$$\delta^{-\delta N}\gamma^{-2\delta N-1}\Big(\frac{CD}{\sqrt{N}}\Big)^{N-\delta N+1}.$$

To compare this result with the ideal bounds (11.3) and (11.4), let us use it with $L \ll \sqrt{N}$. Then $\gamma \ll \sqrt{N}/D$ as we needed, and the theorem gives bounds similar to (11.3) and (11.4).

We will prove Theorem 11.2 in the next few subsections.

11.1. **Step 1: setting out the constraints.** We will first construct a net for the points in $S_{D,d}$ with given set of small coordinates; in the end we unfix this set using the union bound. So let us fix a subset of indices

$$I \subset [N] \quad \text{with} \quad |I| = N - \delta N =: N_0$$

and define the following subset of $\mathbb{C}^I$:

$$S_{D,d,I} := \{z_I : z \in S_{D,d},\, \mathrm{sm}(z) = I\}. \tag{11.6}$$

Consider a point $z_I \in S_{D,d,I}$. As before, depending on the situation, we will work with one of the three representations of z_I: via real and imaginary parts $z_I = x_I + iy_I$, via a long real vector $\widetilde{z}_I = \binom{x_I}{y_I} \in \mathbb{R}^{2N_0}$, and via the $2 \times N_0$ real matrix $V_I = \left[\begin{smallmatrix} x_I^\mathsf{T} \\ y_I^\mathsf{T} \end{smallmatrix}\right]$.

Juxtaposing the available constraints on z_I will help us to construct a small net. Let us set out what we know about z_I. We have three pieces of information.

1. Norm. We know that $\|z\|_2 = 1$ and $z \in \mathrm{Incomp}$. Since $|I^c| = \delta N \le cN$, the coordinates of z in I must have significant mass, i.e.

$$\|z_I\|_2 \ge c.$$

Since $\|z_I\|_2^2 = \|x_I\|_2^2 + \|y_I\|_2^2$, at least one of these terms is bounded below by $c^2/2$. Let us assume without loss of generality that it is the first term, which yields

$$\frac{c}{2} \le \|x_I\|_2 \le 1, \qquad \|y_I\|_2 \le 1. \tag{11.7}$$

2. Least common denominator. We know that $D(\widetilde{z}_{\mathrm{sm}(z)}) \in (D/2, D]$. By definition, this implies that there exists $\theta \in [D/2, D]$ and integer points $p, q \in \mathbb{Z}^I$ such that

$$\|\theta x_I - p\|_2 \le L\sqrt{\log_+ \frac{\theta}{L}}, \qquad \|\theta y_I - q\|_2 \le L\sqrt{\log_+ \frac{\theta}{L}}. \tag{11.8}$$

3. Real-imaginary correlation. We know that $d(z) \le d$. By Lemma 10.5, this implies that

$$\det(V_I V_I^\mathsf{T})^{1/2} \le \frac{Cd}{\delta}.$$

On the other hand, the determinant is the product of the singular values, that is

$$\det(V_I V_I^\mathsf{T})^{1/2} = s_1(V_I)\, s_2(V_I).$$

The larger singular value $s_1(V_I)$ is the operator norm of $\|V_I\|$, so it is bounded below by the norm of either of the two rows of V_I. Thus $s_1(V) \ge \|x_I\|_2 \ge c/2$ due to (11.7). This gives

$$s_2(V_I) \le C'\nu, \quad \text{where} \quad \nu := \frac{C'd}{\delta}. \tag{11.9}$$

11.2. **Step 2: an attempt at construction based on LCD.** Let us ignore for a moment the information about real-imaginary correlation, and try to construct a net for $S_{D,d,I}$ based on the least common denominator only. Dividing the inequalities in (11.8) by θ and using that $\theta \geq D/2$, we obtain

$$\Big\| x_I - \frac{p}{\theta} \Big\|_2 \leq \frac{L}{\theta} \sqrt{\log_+ \frac{\theta}{L}} \leq \frac{2L}{D} \sqrt{\log_+ \frac{D}{L}} = 2\gamma, \tag{11.10}$$

and similarly

$$\Big\| y_I - \frac{q}{\theta} \Big\|_2 \leq 2\gamma. \tag{11.11}$$

This means that x_I and y_I can be approximated by scaled integer points p/θ and q/θ, respectively.

To count the integer points p and q, let us check their norms. By triangle inequality, (11.8) implies

$$\|p\|_2 \leq \|\theta x_I\|_2 + L\sqrt{\log_+ \frac{\theta}{L}} \leq 2D + L\sqrt{\log_+ \frac{2D}{L}} \leq 3D, \tag{11.12}$$

where we used that $\|x_I\|_2 \leq 1$ and $\theta \leq D$.

We claim that the bound (11.12) is sharp within an absolute constant. To check this, use a similar reasoning to get

$$\|p\|_2 \geq \|\theta x_I\|_2 - L\sqrt{\log_+ \frac{\theta}{L}} \geq cD - L\sqrt{\log_+ \frac{D}{L}}. \tag{11.13}$$

In the second inequality we used that $\|x_I\|_2 \geq c/2$ due to (11.7) and $\theta \geq D/2$. Further, we use that $D \geq \frac{1}{2}\sqrt{\delta N}$ due to (11.1) and $L \leq \bar{c}\sqrt{\delta N}$ by assumption. Thus

$$D \geq \frac{1}{2\bar{c}} L.$$

We can choose the absolute constant $\bar{c} > 0$ so that

$$D \geq eL \tag{11.14}$$

as claimed in the statement of the theorem, and also

$$L\sqrt{\log_+ \frac{D}{L}} \leq \frac{c}{2} D.$$

Substituting into (11.13), we obtain

$$\|p\|_2 \geq \frac{c}{2} D.$$

Similarly to (11.12), we obtain

$$\|q\|_2 \leq 3D.$$

Summarizing, we have shown that z_I can be approximated by a scaled integer point $p+iq$, where both p and q have norms at most $4D$. Formally, the set

$$\mathcal{N}_I := \big\{ \alpha(p+iq) : \ \alpha \in \mathbb{R}; \ p, q \in \mathbb{Z}^I \cap B(0, 3D) \big\}$$

is a (4γ)-net of $S_{D,d,I}$. How large is this net? Since $D \geq c_0\sqrt{N}$ due to (11.1), a standard volume argument shows that the number of integer points in the real ball $B(0, 3D)$ in dimension $|I| = N_0$ is bounded by $(CD/\sqrt{N_0})^{N_0}$. Thus the number of "generators" $p + iq$ of the net $\mathcal{N}_I$ is bounded by

$$\Big(\frac{CD}{\sqrt{N_0}} \Big)^{2N_0}. \tag{11.15}$$

Further, one can easily discretize the multipliers α (we will do this later), and obtain a finite net of $S_{D,d,I}$ of cardinality similar to (11.15).

The bound (11.15) is close to the ideal result (11.3). However, it misses the d^N factor, which is understandable since we have not used the real-imaginary correlation $d(z)$ yet. Let us do this now.

11.3. **Step 3: factoring in the real-imaginary correlation.** Let us rewrite the approximation bound (11.8) in terms of the $2 \times N_0$ matrices

$$V_I = \begin{bmatrix} x_I^\mathsf{T} \\ y_I^\mathsf{T} \end{bmatrix} \quad \text{and} \quad W := \begin{bmatrix} p^\mathsf{T} \\ q^\mathsf{T} \end{bmatrix}.$$

It follows that θV_I is approximated by W in the operator norm:

$$\|\theta V_I - W\| \le 2L\sqrt{\log_+ \frac{\theta}{L}}.$$

Weyl's inequality implies that the corresponding singular values of θV_I and W are within $2L\sqrt{\log_+ \frac{\theta}{L}}$ from each other, and in particular we have

$$s_2(W) \le s_2(\theta V_I) + 2L\sqrt{\log_+ \frac{\theta}{L}}$$

Recalling from (11.9) that $s_2(V_I) \le C'\nu$ and that $\theta \le D$, we conclude that

$$s_2(W) \le CD\nu + 2L\sqrt{\log_+ \frac{D}{L}}. \tag{11.16}$$

We can interpret this inequality as saying that the vectors p and q that form the rows of W are almost collinear. Indeed, let $P_{p^\perp}$ denote the orthogonal projection in $\mathbb{R}^I$ onto the subspace orthogonal to the vector p. We claim that

$$\|P_{p^\perp} q\|_2 \le \Big(1 + \frac{\|q\|_2}{\|p\|_2}\Big)\, s_2(W). \tag{11.17}$$

To see why this inequality holds, we can express the determinant $\det(WW^\mathsf{T})^{1/2}$ in two ways – via the base times height formula and as the product of singular values:

$$\det(WW^\mathsf{T})^{1/2} = \|p\|_2 \cdot \|P_{p^\perp} q\|_2 = s_1(W)\, s_2(W). \tag{11.18}$$

The larger singular value $s_1(W)$ is the operator norm of W, which is bounded by the sum of the norms of the rows:

$$s_1(W) \le \|p\|_2 + \|q\|_2.$$

Substituting this into the identity (11.18), we obtain the bound (11.17).

To successfully apply the bound (11.17), we recall from Section 11.2 that $\|q\|_2 \le 3D$ and $\|p\|_2 \ge cD/2$, and moreover $s_2(W)$ is bounded as in (11.16). Thus we obtain

$$\|P_{p^\perp} q\|_2 \le C\Big(D\nu + L\sqrt{\log_+ \frac{D}{L}}\Big). \tag{11.19}$$

Intuitively, this means that p and q are almost collinear, with the degree of collinearity measured by the real-imaginary correlation factor $d(z)$ (which is reflected here through $\nu = C'd/\delta$).

11.4. **Step 4: construction of the net in the genuinely complex case.** We are now ready to construct a net of $S_{D,d,I}$ based on both LCD and the real-imaginary correlation. Let us start with the genuinely complex case of the theorem, where $d \ge d_0$. Using definitions of d_0 and γ in (11.5) and recalling that $\nu = C'd/\delta$, we can rewrite the inequality $d \ge d_0$ as

$$D\nu \ge CL\sqrt{\log_+ \frac{D}{L}} \quad \text{and} \quad D\nu \ge C\sqrt{N}. \tag{11.20}$$

By the first inequality, the first term dominates in the bound (11.19), and we have

$$\|P_{p^\perp} q\|_2 \le 2CD\nu. \tag{11.21}$$

Arguing as in Section 11.2, we see that the set

$$\mathcal{N}_I^{(1)} := \left\{\alpha(p+iq) : \ \alpha \in \mathbb{R};\ p,q \in \mathbb{Z}^I \cap B(0,3D);\ \|P_{p^\perp} q\|_2 \le 2CD\nu\right\}$$

is a (4γ)-net of $S_{D,d,I}$. The collinearity condition (11.21) included in this definition will allow us to bound the number of generators $p+iq$ better than before.

First, exactly as in Section 11.2, the number of possible integer points p in the definition of $\mathcal{N}_I^{(1)}$ can be bounded by the standard volume argument, and we have

$$\#\left\{p\text{'s in the definition of } \mathcal{N}_I^{(1)}\right\} \le \left|\mathbb{Z}^I \cap B(0,3D)\right| \le \Big(\frac{CD}{\sqrt{N_0}}\Big)^{N_0}. \tag{11.22}$$

Next, for a fixed p, let us count the number of possible q's that can make a generator $p+iq$. By definition of $\mathcal{N}_I^{(1)}$, any such q is an integer point in the cylinder

$$\mathcal{C}(p, 3D, 2CD\nu) \tag{11.23}$$

where we denote

$$\mathcal{C}(p,a,b) =: \left\{u \in \mathbb{R}^I : \|P_p u\|_2 \le a,\ \|P_{p^\perp} u\|_2 \le b\right\}.$$

(Here P_p denotes the orthogonal projection in $\mathbb{R}^I$ onto the line spanned by p.)

By a standard covering argument, the number of integer points in the cylinder $\mathcal{C}(p,a,b)$ is bounded by the volume of the Minkowski sum

$$\mathcal{C}(p,a,b) + Q \quad \text{where} \quad Q = \Big[-\frac{1}{2}, \frac{1}{2}\Big]^{N_0}.$$

Further, the unit cube Q is contained in the Euclidean ball $B(0,\sqrt{N_0})$, and the Minkowski sum $\mathcal{C}(p,a,b) + B(0,\sqrt{N_0})$ is clearly contained in the cylinder

$$\mathcal{C}(p, a+\sqrt{N_0}, b+\sqrt{N_0}).$$

This cylinder is a Cartesian product of an interval of length $2(a+\sqrt{N_0})$ and the Euclidean ball of radius $b+\sqrt{N_0}$ in the hyperplane orthogonal to the interval, so the volume of the cylinder can be bounded by

$$2(a+\sqrt{N_0}) \cdot \Big(\frac{C(b+\sqrt{N_0})}{\sqrt{N_0-1}}\Big)^{N_0-1}. \tag{11.24}$$

We can apply this bound to our specific cylinder (11.23) where $a = 3D$ and $b = 2CD\nu$. By (11.1), $a \ge 4c\sqrt{N_0}$, and by the second inequality in (11.20), $b \ge C\sqrt{N_0}$, which means that both $\sqrt{N_0}$ terms can be absorbed into a and b. Thus the number of integer points in the cylinder (11.23) is bounded by

$$Ca\Big(\frac{Cb}{\sqrt{N_0}}\Big)^{N_0-1} \le CD\Big(\frac{CD\nu}{\sqrt{N_0}}\Big)^{N_0-1}. \tag{11.25}$$

Summarizing, now we know the following about the generators $p+iq$ of the net $\mathcal{N}_I^{(1)}$. The number of possible points p is bounded as in (11.22). For each fixed p, the number of possible q's that can make the generator $p+iq$ is bounded by the quantity in (11.25). Thus, the total number of generators $p+iq$ is bounded by

$$\Big(\frac{CD}{\sqrt{N_0}}\Big)^{N_0} CD\Big(\frac{CD\nu}{\sqrt{N_0}}\Big)^{N_0-1}. \tag{11.26}$$

11.5. **Step 5: finalizing the genuinely complex case.** Three minor points still remain to be addressed in this case. First, the net $\mathcal{N}_I^{(1)}$ we constructed is infinite due to the real multiplier α. Second, this net controls only the coordinates that lie in I (recall the definition (11.6) of the set $S_{D,d,I}$). Third, the construction we made was for a fixed set of coordinates I. We will now take care of these issues.

11.5.1. *Discretizing the multipliers.* The first point can be addressed by discretizing the set of multipliers α in the definition of the net $\mathcal{N}_I^{(1)}$. Since $S_{D,d,I}$ is a subset of the unit ball $B(0,1)$, we may consider only the multipliers α that satisfy $\|\alpha(p+iq)\|_2 \le 1$. For a fixed generator $p+iq$, we discretize the interval of multipliers $\{\alpha \in \mathbb{R} : \|\alpha(p+iq)\|_2 \le 1\}$ by replacing it with a set of $2/\gamma$ numbers α_j that are equally spaced in that interval. The vector $\alpha(p+iq)$ can then be approximated by a vector $\alpha_i(p+iq)$ with error at most γ in the Euclidean norm.

Since $\mathcal{N}_I^{(1)}$ is a (4γ)-net of $S_{D,d,I}$, the discretization we just constructed is a (6γ)-net of $S_{D,d,I}$. Let us call this net $\mathcal{M}_I^{(1)}$. The cardinality of $\mathcal{M}_I^{(1)}$ is bounded by $(2/\gamma)$ times the number of generators of $\mathcal{N}_I^{(1)}$, which we bounded in (11.26). In other words,

$$|\mathcal{M}_I^{(1)}| \le \frac{2}{\gamma} \cdot \Big(\frac{CD}{\sqrt{N_0}}\Big)^{N_0} CD\Big(\frac{CD\nu}{\sqrt{N_0}}\Big)^{N_0-1}. \tag{11.27}$$

11.5.2. *Controlling the coordinates outside I.* The second point we need to address is that since $\mathcal{M}_I^{(1)}$ is a (6γ)-net of the set $S_{D,d,I} = \{z_I : z \in S_{D,d},\ \mathrm{sm}(z) = I\}$, this net can only control the coordinates of z in I. To control the coordinates outside I, it is enough to construct a separate γ-net for the set

$$\Big\{z_{I^c} : z \in S_{\mathbb{C}}^{N-1}\Big\}.$$

Since $|I^c| = \delta N$, a standard volume bound allows one to find such a net of cardinality at most $(5/\gamma)^{2\delta N}$. Combining the two nets, we conclude that there exists a (7γ)-net of the set

$$T_{D,d,I} = \{z : z \in S_{D,d},\ \mathrm{sm}(z) = I\}$$

of cardinality at most

$$\Big(\frac{5}{\gamma}\Big)^{2\delta N} |\mathcal{M}_I^{(1)}|.$$

11.5.3. *Unfixing the set of coordinates I.* Finally, the third point we need to address is that our construction was for a fixed set of indices I. To unfix I, we note that there is at most

$$\binom{N}{N-|I|} = \binom{N}{\delta N} \le \Big(\frac{e}{\delta}\Big)^{\delta N}$$

ways to choose I. So, combining the nets we constructed for each I, we obtain a (7γ)-net of $S_{D,d}$ of cardinality at most

$$\Big(\frac{e}{\delta}\Big)^{\delta N} \cdot \Big(\frac{5}{\gamma}\Big)^{2\delta N} |\mathcal{M}_I^{(1)}|.$$

Substituting here the bound (11.27) for $|\mathcal{M}_I^{(1)}|$ and recalling that $N_0 = N - \delta N$ and $\nu = C'd/\delta$, we can express the previous bound as

$$\Big(\frac{e}{\delta}\Big)^{\delta N} \cdot \Big(\frac{5}{\gamma}\Big)^{2\delta N} \cdot \frac{2}{\gamma} \cdot \Big(\frac{CD}{\sqrt{N}}\Big)^{N-\delta N} CD\Big(\frac{CDd}{\delta\sqrt{N}}\Big)^{N-\delta N-1}.$$

Simplifying the expression, we obtain the first part of the theorem.

11.6. **Step 6: the essentially real case.** We proceed to the essentially real case, where $d < d_0$. This means that at least one of the inequalities in (11.20) fails.

11.6.1. *Case 1.* Assume that the first inequality in (11.20) holds but the other fails, that is

$$D\nu \geq CL\sqrt{\log_+ \frac{D}{L}} \quad \text{and} \quad D\nu \leq C\sqrt{N}.$$

We proceed in the same way as in the genuinely complex case until we apply the general bound on the integer points (11.24) to our cylinder with $a = 4D$ and $b = 2CD\nu$. This is the only place where we used the second inequality in (11.20), which now fails. This means that b gets absorbed into the $\sqrt{N_0}$ term, and the the number of integer points in the cylinder (11.23) is consequently bounded by

$$Ca\Big(\frac{C\sqrt{N_0}}{\sqrt{N_0}}\Big)^{N_0-1} \leq C^N D.$$

Using this bound in place of (11.25) and arguing exactly as in the genuinely complex case, we complete the proof for this sub-case.

11.6.2. *Case 2.* The remaining sub-case is where the first inequality in (11.20) fails, that is

$$D\nu < L\sqrt{\log_+ \frac{D}{L}}. \tag{11.28}$$

Then the second term dominates in the bound (11.19), and we have

$$\|P_{p^\perp} q\|_2 \leq 2CL\sqrt{\log_+ \frac{D}{L}}. \tag{11.29}$$

Let us fix a point $z_I = x_I + iy_I$ from $S_{D,d,I}$. Since the orthogonal projection has norm one, (11.11) yields

$$\Big\|P_{p^\perp}\Big(y_I - \frac{q}{\theta}\Big)\Big\|_2 \leq 2\gamma.$$

Combining this with (11.29) and using triangle inequality, we obtain

$$\|P_{p^\perp} y_I\|_2 \leq 2\gamma + \frac{2CL}{\theta}\sqrt{\log_+ \frac{D}{L}}.$$

Recalling that $\theta \geq D$ and the definition of γ in the theorem, we obtain

$$\|P_{p^\perp} y_I\|_2 \leq C\gamma.$$

We can interpret this inequality as follows. There exists a multiplier $\beta \in \mathbb{R}$ such that

$$\|y_I - \beta p\|_2 \leq C\gamma.$$

Let us rewrite (11.10) in a similar way – there exists a multiplier $\alpha = 1/\theta \in \mathbb{R}$ such that

$$\|x_I - \alpha p\|_2 \leq \gamma.$$

Recalling that $z_I = x_I + iy_I$, it follows that

$$\|z_I - (\alpha + i\beta)p\|_2 \leq C\gamma.$$

Furthermore, recalling from (11.12) that $\|p\|_2 \leq 3D$, we conclude that the set

$$\mathcal{N}_I^{(2)} := \big\{(\alpha + i\beta)p : \ \alpha, \beta \in \mathbb{R}; \ p \in \mathbb{Z}^I \cap B(0, 3D)\big\}$$

is a $(C\gamma)$-net of $S_{D,d,I}$.

The number of generators p can be counted by a volume argument, exactly as in (11.22):

$$\#\Big\{p\text{'s in the definition of } \mathcal{N}_I^{(2)}\Big\} \leq \big|\mathbb{Z}^I \cap B(0, 3D)\big| \leq \Big(\frac{CD}{\sqrt{N_0}}\Big)^{N_0}.$$

Finally, we can discretize the multipliers $\alpha + i\beta$ and unfix the set I similarly to how we did it in Section 11.5. We obtain a (6γ)-net of $S_{D,d}$ of cardinality at most

$$\left(\frac{e}{\delta}\right)^{\delta N} \cdot \left(\frac{5}{\gamma}\right)^{2\delta N} \cdot \left(\frac{C}{\gamma}\right)^{2} \cdot \left(\frac{CD}{\sqrt{N_0}}\right)^{N_0}. \tag{11.30}$$

(To recall, the first term here comes from unfixing I, the second from controlling coordinates outside I, the third from discretizing the multipliers in the complex disc $\{\alpha + i\beta \in \mathbb{C} : \|(\alpha + i\beta)p\|_2 \leq 1\}$, and fourth from (11.30).)

Recalling that $N_0 = N - \delta N$ and $\nu = C'd/\delta$, and simplifying the expression, we prove the second part of Theorem 11.2. $\square$

12. Structure of kernels, and proof of Theorem 8.1 on the distances

In this section we will show that random subspaces, and specifically the kernels of random matrices, are arithmetically unstructured, which means that they have large LCD with high probability. The following is the main result of this section.

Theorem 12.1 (Kernels of random matrices are unstructured)**.** *Let $B \in \mathbb{C}^{n\times N}$ be a random matrix satisfying Assumptions 1.1 and 1.4, and assume that $n = (1-\varepsilon)N$ for some $\varepsilon \in (2/n, c)$. Set $L := \sqrt{\varepsilon N}$. Then,*

$$\mathbb{P}\left\{D(\widetilde{\ker B}, L) \leq \min\left(\sqrt{N}e^{c/\sqrt{\varepsilon}},\ \varepsilon N\right) \text{ and } \mathcal{B}_{B,M}\right\} \leq e^{-cN}.$$

This theorem will follow by balancing the two forces – the small ball probability estimates of Section 10 and the net for vectors with given LCD of Section 11.

It would be convenient to first state a preliminary version of Theorem 12.1 which holds for vectors with given levels of LCD $D(\widetilde{z}_{\mathrm{sm}(z)})$ and real-imaginary correlation $d(z)$. We will work here with somewhat smaller level sets than $S_{D,d}$ from Section 11. For $D, d > 0$, we consider

$$\bar{S}_{D,d} = \left\{z \in \mathrm{Incomp} :\ D/2 < D(\widetilde{z}_{\mathrm{sm}(z)}) \leq D;\ d/2 \leq d(z) \leq d\right\}.$$

Clearly, $S_{D,d}$ is the union of the sets $\bar{S}_{D,d}$ for all $d \leq d_0$.

Proposition 12.2 (Kernels and level sets)**.** *Let $B \in \mathbb{C}^{n\times N}$ be a random matrix satisfying Assumptions 1.1 and 1.4, and let $n = (1-\varepsilon)N$ for some $\varepsilon \in (2/n, c)$. Set L from the definition of $D(\widetilde{z}_{\mathrm{sm}(z)})$ to be $L := \sqrt{\varepsilon N}$. Let*

$$D \leq \min\left(\sqrt{N}e^{c/\sqrt{\varepsilon}},\ \varepsilon N\right)$$

and let d_0 be the threshold value from (11.5) for $\delta = c\sqrt{\varepsilon}$.

1. (Genuinely complex case). For any $d \in [d_0, 1]$, we have

$$\mathbb{P}\left\{\bar{S}_{D,d} \cap \ker B \neq \emptyset \text{ and } \mathcal{B}_{B,M}\right\} \leq e^{-N}. \tag{12.1}$$

2. (Essentially real case). For any $d \in [0, d_0]$, we have

$$\mathbb{P}\left\{S_{D,d} \cap \ker B \neq \emptyset \text{ and } \mathcal{B}_{B,M}\right\} \leq e^{-N}.$$

Note here that in the genuinely complex case, we use an additional stratification by $d(z)$ by considering the sets $\bar{S}_{D,d}$, while in the essentially real case, we treat the set $S_{D,d}$ in one strike.

We will prove this proposition in the next two subsections.

12.1. Proof of Proposition 12.2 in the genuinely complex case.

Let us not fix δ at this moment but rather assume that it satisfies the inequality

$$\bar{c}^{-2}\varepsilon \leq \delta \leq \sqrt{\varepsilon}. \tag{12.2}$$

The lower bound ensures that $L = \sqrt{\varepsilon N} \leq \bar{c}\sqrt{\delta N}$, which will be needed to apply Theorem 11.2.

12.1.1. *Step 1: combining the small ball probability with the net.* We fix a vector $z \in \bar{S}_{D,d}$ and apply Theorem 10.3. By the assumptions on n, we can write the conclusion of this theorem as follows:

$$\mathbb{P}\left\{\|Bz\|_2 \le t\sqrt{N}\right\} \le \left[\frac{C}{d}\Big(t + \frac{1}{\sqrt{\delta N}}\Big)^2\right]^{(1-\delta)(1-\varepsilon)N}, \quad t \ge 0. \tag{12.3}$$

Let us apply this bound for $t := \lambda\sqrt{\delta}$, where $\lambda \in (0,1)$ is a parameter whose value we choose later. If we assume that

$$\lambda\sqrt{\delta} \ge \frac{1}{\sqrt{\delta N}}, \tag{12.4}$$

then the probability bound can be expressed as

$$\mathbb{P}\left\{\|Bz\|_2 \le \lambda\sqrt{\delta N}\right\} \le \Big(\frac{C\lambda^2\delta}{d}\Big)^{(1-\delta)(1-\varepsilon)N}.$$

Next, Theorem 11.2 provides us with a $(C\gamma)$-net of $S_{D,d}$ of controlled cardinality. Clearly, the same is true for the smaller set $\bar{S}_{D,d}$. Let us denote such a net by $\mathcal{N}$. Using a union bound, we can combine the probability bound with the net as follows:

$$p := \mathbb{P}\left\{\inf_{z\in\mathcal{N}} \|Bz\|_2 \le \lambda\sqrt{\delta N}\right\} \le \Big(\frac{C\lambda^2\delta}{d}\Big)^{(1-\delta)(1-\varepsilon)N} \cdot |\mathcal{N}|.$$

Recalling the bound on $|\mathcal{N}|$ given by Theorem 11.2, we obtain

$$p \le \Big(\frac{C\lambda^2\delta}{d}\Big)^{(1-\delta)(1-\varepsilon)N} \cdot \delta^{-N+1}\gamma^{-2\delta N-1}\Big(\frac{CD}{\sqrt{N}}\Big)^{2N-2\delta N-1} d^{N-\delta N-1}. \tag{12.5}$$

Our goal is to show that $p \le e^{-N}$.

12.1.2. *Step 2: simplifying the probability bound.* We claim that d, δ and γ can be removed from (12.5) at the cost of increasing the bound by C^N.

To see this for d, note that the exponent of d in the bound (12.5) is

$$(N - \delta N - 1) - (1-\delta)(1-\varepsilon)N = (1-\delta)\varepsilon N - 1 \ge \varepsilon N/2 - 1 \ge 0.$$

Since $d \in [0,1]$, removing d can only make the bound (12.5) larger.

Similarly, the exponent of $1/\delta$ in the bound (12.5) is

$$N - 1 - (1-\delta)(1-\varepsilon)N \le 2\delta N.$$

Thus δ contributes to the bound a factor not larger than $\delta^{-2\delta N} \le C^N$.

Finally, to evaluate the contribution of γ, let us denote

$$\bar{D} := \frac{D}{L} = \frac{D}{\sqrt{\varepsilon N}}.$$

Notice in passing that $\bar{D} \ge e$ by (11.14). Recalling the definition of γ in (11.5), we have

$$\frac{1}{\gamma} = \frac{\bar{D}}{\sqrt{\log_+ \bar{D}}} \le \bar{D}.$$

Therefore, the contribution of γ to the bound (12.5) is a factor not larger than

$$\gamma^{-2\delta N-1} \le \bar{D}^{4\delta N}.$$

To bound this quantity further, we can use the assumption on D and the comparison (12.2), which imply

$$\bar{D} \le \frac{1}{\sqrt{\varepsilon}} e^{c/\sqrt{\varepsilon}} \le \frac{1}{\delta} e^{c/\delta}. \tag{12.6}$$

Therefore, γ contributes to the bound (12.5) a factor not larger than C^N.

We have shown that d, δ and γ can be removed from (12.5) at the cost of increasing the bound by C^N. In other words, we have

$$p \le \left[C\lambda^{2(1-\delta)(1-\varepsilon)} \left(C\sqrt{\varepsilon}\bar{D}\right)^{2-2\delta} \right]^N \cdot (C\sqrt{\varepsilon}\bar{D})^{-1}.$$

Further, since $\varepsilon \ge 1/N$ and $\bar{D} \ge e$ due to (11.14), the last term can be dropped at the cost of increasing the bound by C^N. Thus

$$p \le \left[C\lambda^{2(1-\delta)(1-\varepsilon)} \left(C\sqrt{\varepsilon}\bar{D}\right)^{2-2\delta} \right]^N.$$

Solving for λ, we see that the desired bound

$$p \le e^{-N}$$

holds whenever

$$\lambda \le \left(\frac{c}{\sqrt{\varepsilon}\bar{D}}\right)^{\frac{1}{1-\varepsilon}}. \tag{12.7}$$

12.1.3. *Step 3: approximation by the net.* In the previous step we showed that the event

$$\inf_{z_0 \in \mathcal{N}} \|Bz_0\|_2 \le \lambda\sqrt{\delta N}$$

holds with probability at least $1 - e^{-N}$, as long as the parameter λ satisfies (12.4) and (12.7). Let us fix a realization of the random matrix B for which this event does hold.

Fix a vector $z \in \bar{S}_{D,d}$. To finish the proof of (12.1), we need to show that $\|Bz\|_2 > 0$. Let us choose a vector $z_0 \in \mathcal{N}$ which best approximates z; by definition of $\mathcal{N}$ we have

$$\|z - z_0\|_2 \le C\gamma = \frac{C\sqrt{\log_+ \bar{D}}}{\bar{D}}.$$

Assume that the event $\mathcal{B}_{B,M}$ occurs. By triangle inequality, it follows that

$$\begin{aligned} \|Bz\|_2 &\ge \|Bz_0\|_2 - \|B\| \cdot \|z - z_0\|_2 \\ &\ge \lambda\sqrt{\delta N} - M\sqrt{N} \cdot \frac{C\sqrt{\log_+ \bar{D}}}{\bar{D}} \end{aligned}$$

This quantity is positive as we desired if λ satisfies

$$\lambda \ge \frac{C\sqrt{\log_+ \bar{D}}}{\sqrt{\delta}\bar{D}}. \tag{12.8}$$

Recall that we allow our constants to depend on M. This allows to absorb M in C in the inequality above.

12.1.4. *Step 4: final choice of the parameters.* We have shown that the conclusion (12.1) of the proposition in the genuinely complex case holds if we can choose the parameters δ and λ in such a way that they satisfy (12.2), (12.4), (12.7) and (12.8). We will now check that such a choice indeed exists.

First, let us choose λ just large enough to satisfy (12.8); thus we set

$$\lambda := \frac{C\sqrt{\log_+ \bar{D}}}{\sqrt{\delta}\bar{D}}.$$

Without loss of generality, we assume that $C \geq 1/\bar{c}$ where $\bar{c}$ appears in (12.2). To check (12.4), we use the assumption that $D \leq \varepsilon N$, which implies that $\bar{D} \leq \sqrt{\varepsilon N}$. This and the choice of λ imply

$$\lambda\sqrt{\delta} \geq \frac{1}{\bar{c}\bar{D}} \geq \frac{1}{\bar{c}\sqrt{\varepsilon N}} \geq \frac{1}{\sqrt{\delta N}},$$

where in the first inequality we used that $C \geq 1/\bar{c}$ and that $\bar{D} \geq e$, and the last inequality follows from (12.2). This proves (12.4).

It remains to check (12.7), which takes the form

$$\frac{C\sqrt{\log_+ \bar{D}}}{\sqrt{\delta}\bar{D}} \leq \Big(\frac{c}{\sqrt{\varepsilon}\bar{D}}\Big)^{\frac{1}{1-\varepsilon}}.$$

Replacing $(\sqrt{\varepsilon})^{\frac{1}{1-\varepsilon}}$ by the smaller quantity $c\sqrt{\varepsilon}$ (with a small constant $c > 0$) and rearranging the terms, we may rewrite this restriction as

$$\bar{D}^{\frac{\varepsilon}{1-\varepsilon}}\sqrt{\log_+ \bar{D}} \leq c\sqrt{\frac{\delta}{\varepsilon}}.$$

Substituting here the assumption $\bar{D} \leq \frac{1}{\varepsilon} e^{c/\varepsilon}$, we see that $\bar{D}^{\varepsilon}$ is bounded by an absolute constant, and $\log_+ \bar{D} \leq c/\sqrt{\varepsilon}$. Thus the restriction is satisfied if

$$\frac{C}{\sqrt{\varepsilon}} \leq \frac{c\delta}{\varepsilon}.$$

It remains to choose $\delta := c\sqrt{\varepsilon}$; then the restriction is satisfied, and we have verified (12.7). This finished the proof of the proposition in the genuinely complex case.

12.2. **Proof of Proposition 12.2 in the essentially real case.** The argument is similar, and even simpler, than in the genuinely complex case. We just need to use the appropriate small ball probability bound, namely Theorem 10.4 instead of (12.3), and the corresponding bound on the net – the one from the essentially real case in Theorem 11.2. This leads to the following variant of (12.5):

$$p \leq (C\lambda)^{(1-\delta)(1-\varepsilon)N} \cdot \delta^{-\delta N}\gamma^{-2\delta N-1}\Big(\frac{CD}{\sqrt{N}}\Big)^{N-\delta N+1}.$$

Like before, we can remove δ and γ, simplifying the bound to

$$p \leq \Big[C\lambda^{(1-\delta)(1-\varepsilon)}\,\big(C\sqrt{\varepsilon}\bar{D}\big)^{1-\delta}\Big]^N.$$

Then the desired bound $p \leq e^{-N}$ holds if λ is chosen so that

$$\lambda \leq \Big(\frac{c}{\sqrt{\varepsilon}\bar{D}}\Big)^{\frac{1}{1-\varepsilon}}.$$

In particular, this holds if λ satisfies the same restriction as in the genuinely complex case, namely (12.7). (To see this, note that $1 + 2\delta \geq 1/(1-\varepsilon)$ by (12.2).)

The rest of the proof is exactly as in the genuinely complex case. Proposition 12.2 is proved. □

12.3. **Proof of Theorem 12.1.** For convenience, let us denote

$$D_0 := \min(\sqrt{N}e^{c/\sqrt{\varepsilon}},\ \varepsilon N).$$

Assume that $D(\widetilde{\ker B}, L) \leq D_0$, and the event $\mathcal{B}_{B,M}$ occurs. This means that there exists $z \in S^{N-1}$ such that

$$z \in \ker B, \quad D(\widetilde{z}_{\mathrm{sm}(z)}) \leq D_0.$$

We can bound the probability of this event by considering the following cases. If z is compressible, then such event holds with probability at most $e^{-c_1 N}$ according to Proposition 9.4. Assume that

z is incompressible. In the genuinely complex case where $d(\tilde{z}) > d_0$, the vector z must belong to a level set $\bar{S}_{D,d}$ for some $D \in [c_0\sqrt{N}, D_0]$ and $d \in [d_0, 1]$. (The lower bound on D here is from (11.1).) For given D and d, the probability that such z exists is at most $e^{-c_2 N}$ by the first part of Proposition 12.2. In the remaining, essentially real case where $d(\tilde{z}) \le d_0$, the vector z must belong to a level set S_{D,d_0} for some $D \in [c_0\sqrt{N}, D_0]$. For given D and d, the probability that such z exists is at most $e^{-c_2 N}$ by to the second part of Proposition 12.2.

This reasoning shows that the probability that $D(\widetilde{\ker B}, L) \le D_0$ is bounded by

$$e^{-c_1 N} + \sum_{\substack{D \in [c_0\sqrt{N}, D_0] \\ \text{dyadic}}} \sum_{\substack{d \in [d_0, 1] \\ \text{dyadic}}} e^{-c_2 N} + \sum_{\substack{D \in [c_0\sqrt{N}, D_0] \\ \text{dyadic}}} e^{-c_3 N}. \tag{12.9}$$

The definitions of the level sets allowed us here to discretize the ranges of D and d by including only the dyadic values in the sum, namely the values of the form 2^k, $k \in \mathbb{Z}$.

It remains to bound the number the terms in the sums. The number of dyadic values in the interval $[c_0 N, D_0]$ is at most

$$\log\Big(\frac{D_0}{c_0\sqrt{N}}\Big) \le \log N$$

since $D_0 \le N$ by definition. Similarly, using the definition 11.5 of d_0, we see that the number of dyadic values in the interval $[d_0, 1]$ is at most

$$\log\Big(\frac{1}{d_0}\Big) \le \log\Big(\frac{D}{\varepsilon N}\Big) \le \log\Big(\frac{D_0}{\varepsilon N}\Big) \le \log N.$$

Therefore, the probability estimate (12.9) is bounded by

$$e^{-c_1 N} + \log^2(N) e^{-c_2 N} + \log(N) e^{-c_3 N} \le e^{-cN}.$$

This completes the proof. □

12.4. **Distances between random vectors and subspaces.** We are ready to prove Theorem 8.1 on the distances between random vectors and subspaces. It can be quickly deduced by combining small ball probability bounds we developed in Section 7 with the bound on LCD for random subspaces, namely Theorem 12.1.

Given a random vector $Y \in \mathbb{R}^k$ and an event Ω, denote

$$\mathcal{L}_\Omega(Y, r) = \sup_{y \in \mathbb{R}^k} \mathbb{P}\left\{\|Y - y\|_2 < r \text{ and } \Omega\right\}.$$

In Section 8.1, we reduced Theorem 8.1 to a problem over reals. Indeed, according to (8.3), it suffices to bound

$$p_0 := \mathcal{L}_{\mathcal{B}_{B,M}}(P_{\widetilde{\ker B}} \widehat{Z}, 2\tau\sqrt{\varepsilon N}).$$

For this, we recall that $\dim(\widetilde{\ker B}) = 2\varepsilon N$ and apply Corollary 7.10, which yields

$$p_0 \le \Big(\frac{CL}{\sqrt{\varepsilon N}}\Big)^{2\varepsilon N} \Big(\tau + \frac{\sqrt{\varepsilon N}}{D(\widetilde{\ker B}, L)}\Big)^{2\varepsilon N}. \tag{12.10}$$

Next, Theorem 12.1 states that for $L = \sqrt{\varepsilon N}$, with probability at least $1 - e^{-cN}$ we have

$$D(\widetilde{\ker B}, L) \ge \min\Big(\sqrt{N} e^{c/\sqrt{\varepsilon}},\ \varepsilon N\Big).$$

Substituting this into (12.10) and simplifying the bound, we obtain

$$p_0 \le \Big[C\Big(\tau + \frac{1}{\sqrt{\varepsilon N}} + \sqrt{\varepsilon} e^{-c/\sqrt{\varepsilon}}\Big)\Big]^{2\varepsilon N} + e^{-cN} \le \Big[C\Big(\tau + \frac{1}{\sqrt{\varepsilon N}} + e^{-c'/\sqrt{\varepsilon}}\Big)\Big]^{2\varepsilon N}.$$

This inequality combined with (8.3) completes the proof of Theorem 8.1. □

13. Proof of Theorem 6.1 on invertibility for general distributions

The strategy of the proof of Theorem 6.1 will be very close to the argument we gave for continuous distributions in Section 5. However, there are two important differences. First, the distance bound for continuous distributions given in Lemma 5.4 can not hold for general distributions; we will replace it by Theorem 8.1. Another ingredient that is not available for general distributions is the lower bound given in Lemma 5.5 and all of its consequences in Section 5.5. Instead, we will use the small ball probability bounds for general distributions that we developed in Section 7.3. Let us start with this latter task.

13.1. G is bounded below on the small subspace E^-.

Here we extend the argument of Section 5.5 to general distributions.

Lemma 13.1. *With probability at least $1 - e^{-cn}$, we have $S_{E^-} \subset \mathrm{Incomp}$.*

Proof. The definition of E^- in Section 5.4.2 implies that

$$\|Bz\|_2 \le c\tau\varepsilon\sqrt{n} \quad \text{for all } z \in S_{E^-}.$$

On the other hand, Proposition 9.4 states that with probability at least $1 - e^{-cn}$,

$$\|Bz\|_2 > c\sqrt{n} \quad \text{for all } z \in \mathrm{Comp}.$$

Since these two bounds can not hold together for the same z, it follows that the sets S_{E^-} and Comp are disjoint. This proves the lemma. □

The following result is a version of Lemma 5.5 for general distributions.

Lemma 13.2 (Lower bound for a fixed row and vector)**.** *Let G_j denote the j-th row of G. Then for each j, $z \in \mathrm{Incomp}$, and $\theta \ge 0$, we have*

$$\mathbb{P}\left\{|\langle G_j, z\rangle| \le \theta\right\} \le C\Big(\theta + \frac{1}{\sqrt{n}}\Big).$$

Proof. Fix $z = x + iy$ and let J be the set of indices of all except cn largest (in the absolute value) coordinates of z. By Markov's inequality and Definition 9.1 of incompressible vectors, we have

$$\|z_J\|_\infty \le \frac{c}{\sqrt{n}}, \quad \|z_J\|_2 \ge c.$$

Since $z_J = x_J + iy_J$, either the real part x_J or the complex part y_J has ℓ_2-norm bounded below by $c/2$. Let us assume without loss of generality that x_J satisfies this, so

$$\|x_J\|_\infty \le \frac{c}{\sqrt{n}}, \quad \|x_J\|_2 \ge c. \tag{13.1}$$

The first inequality and Proposition 7.4 imply that

$$D_1(x_J) \ge c\sqrt{n}.$$

Here we use notation $D_1(\cdot)$ for the LCD in dimension $m = 1$ and with $L \sim 1$; note that it is distinct from the LCD in dimension $m = 2$ we studied in the major part of this paper.

We proceed similarly to the proof of Lemma 5.5. Decomposing the random vector $Z := G_j$ as $Z = X + iY$, we obtain the bound (5.13):

$$\mathbb{P}\left\{|\langle Z, z\rangle| \le \theta\right\} \le \mathcal{L}(\langle X, x\rangle, \theta).$$

Further, we use the restriction property of the concentration function (Lemma 3.2) followed by the small ball probability bound (Corollary 7.6), and obtain

$$\mathcal{L}(\langle X, x\rangle, \theta) \le \mathcal{L}(\langle X_J, x_J\rangle, \theta) \le C\Big(\theta + \frac{1}{\sqrt{n}}\Big).$$

This completes the proof of the lemma. □

Using Tensorization Lemma 3.3 and Remark 3.4 exactly as we did before in Section 5.5.1, we obtain the following version of Lemma 5.6 for general distributions.

Lemma 13.3 (Lower bound for a fixed vector)**.** *For each $x \in \text{Incomp}$ and $\theta > 0$, we have*

$$\mathbb{P}\left\{\|Gx\|_2 \le \theta\sqrt{\varepsilon n}\right\} \le \left(C\theta + \frac{C}{\sqrt{n}}\right)^{\varepsilon n}.$$

In particular, if $\theta \ge 1/\sqrt{n}$, then the probability is further bounded by $(C\theta)^{\varepsilon n}$. This is similar to the bound we had in Lemma 5.6 for the continuous case. Using this observation, we deduce the following version of Lemma 5.7 for general distributions, and with the same proof.

Lemma 13.4 (Lower bound on a subspace)**.** *Let $M \ge 1$ and $\mu \in (0,1)$. Let E be a fixed subspace of $\mathbb{C}^n$ of dimension at most $\mu\varepsilon n$, and such that $S_E \subset \text{Incomp}$. Then, for every $\theta \ge 1/\sqrt{n}$, we have*

$$\mathbb{P}\left\{\inf_{x\in S_E}\|Gx\|_2 < \theta\sqrt{\varepsilon n} \text{ and } \mathcal{B}_{G,M}\right\} \le \left[C(M/\sqrt{\varepsilon})^{2\mu}\theta^{1-2\mu}\right]^{\varepsilon n}.$$

This lemma implies a lower bound on the smallest singular value of G restricted to the set S_{E^-} similar to Lemma 5.7.

Corollary 13.5. *Let $M \ge 1$ and $\mu \in (0,1)$. Then, for every $\theta \ge 1/\sqrt{n}$, we have*

$$\mathbb{P}\left\{\inf_{x\in S_{E^-}}\|Gx\|_2 < \theta\sqrt{\varepsilon n} \text{ and } S_{E^-} \subset \text{Incomp and } \mathcal{D}_{E^-}\cap\mathcal{B}_{G,M}\right\} \le \left[C(M/\sqrt{\varepsilon})^{2\mu}\theta^{1-2\mu}\right]^{\varepsilon n}.$$

13.2. **Proof of Theorem 6.1.** The argument will follow the same lines as in Section 5 for continuous distributions; here we will only indicate necessary modifications.

Without loss of generality, we may assume that

$$\varepsilon \ge n^{-0.4}. \tag{13.2}$$

Indeed, (6.1) implies that $t^{0.4}\varepsilon^{-1.4} \ge \varepsilon^{-1}n^{-0.4}$, so the statement of Theorem 6.1 becomes vacuous whenever (13.2) does not hold.

Since $\text{dist}(B_j, H_j) \ge \text{dist}(B_j', H_j')$ by (5.3), Theorem 8.1 implies that

$$\mathbb{P}\left\{\text{dist}(B_j, \text{Im}(H_j)) \le \tau\sqrt{\varepsilon n} \text{ and } \mathcal{B}_{A,M}\right\} \le (C\tau)^{\varepsilon n} \tag{13.3}$$

for any

$$\tau \ge \frac{c}{\sqrt{\varepsilon n}} + e^{-c/\sqrt{\varepsilon}} =: \tau_0.$$

Consider the random variables

$$Y_j := \left[\max\left(\text{dist}(B_j, H_J),\ \tau_0\sqrt{\varepsilon n}\right)\right]^{-2}\cdot\mathbf{1}_{\mathcal{B}_{A,M}}$$

and argue as in Section 5.4.1. We see that Y_j belong to weak L^p for $p = \varepsilon n/2$ and $\|Y_j\|_{p,\infty} \le C^2/\varepsilon n$. By weak triangle inequality, this yields

$$\mathbb{P}\left\{\sum_{j=1}^n Y_j > \frac{C}{\tau^2\varepsilon}\right\} \le (C\tau)^{\varepsilon n}, \quad \tau > 0.$$

Therefore,

$$\begin{aligned}\mathbb{P}\left\{\sum_{j=1}^n \text{dist}(B_j, H_j)^{-2} > \frac{1}{\tau^2\varepsilon} \text{ and } \mathcal{B}_{A,M}\right\} &\le (C\tau)^{\varepsilon n} + \sum_{j=1}^n \mathbb{P}\left\{\text{dist}(B_j, H_j)^{-2} \ne Y_j \text{ and } \mathcal{B}_{A,M}\right\}\\ &\le (C\tau)^{\varepsilon n} + \sum_{j=1}^n \mathbb{P}\left\{\text{dist}(B_j, H_j) < \tau_0\sqrt{\varepsilon n} \text{ and } \mathcal{B}_{A,M}\right\}.\end{aligned}$$

Using again (13.3) and then (13.2), we see that this probability can be further bounded by

$$(C\tau)^{\varepsilon n} + n(C\tau_0)^{\varepsilon n} \leq (C_1\tau)^{\varepsilon n} \quad \text{for } \tau \geq \tau_0.$$

Defining the subspaces E^+, E^- and the event $\mathcal{D}_{E^-}$ as in Section 5.4.2, we derive from this that

$$\mathbb{P}\left\{(\mathcal{D}_{E^-})^c \text{ and } \mathcal{B}_{A,M}\right\} \leq (C_1\tau)^{\varepsilon n} \quad \text{for } \tau \geq \tau_0. \tag{13.4}$$

We finish the proof as in Section 5.6.2. Set

$$\tau = \sqrt{t} \quad \text{and} \quad \theta = \frac{C\sqrt{t}}{\varepsilon^{3/2}}.$$

Then (6.1) ensures that $\tau \geq \tau_0$, so (13.4) holds. Furthermore, (6.1) and (13.2) guarantee that $\theta \geq 1/\sqrt{n}$, hence Corollary 13.5 applies. Similarly to (5.18), we use Corollary 13.5, Lemma 13.1, and (13.4) to obtain

$$\begin{aligned}
\mathbb{P}\left\{s_{\bar{A}} < t\sqrt{n} \text{ and } \mathcal{B}_{A,M}\right\} &\leq \mathbb{P}\left\{s_G < \frac{Ct}{\tau\varepsilon}\cdot\sqrt{n} \text{ and } \mathcal{B}_{A,M}\right\} \\
&\leq \mathbb{P}\left\{s_G < \theta\cdot\sqrt{\varepsilon n} \text{ and } S_{E^-} \subset \text{Incomp and } \mathcal{D}_{E^-} \cap \mathcal{B}_{A,M}\right\} \\
&\quad + \mathbb{P}\left\{S_{E^-} \not\subset \text{Incomp and } \mathcal{B}_{A,M}\right\} + \mathbb{P}\left\{(\mathcal{D}_{E^-})^c \text{ and } \mathcal{B}_{A,M}\right\} \\
&\leq \left(C\varepsilon^{-0.05}\theta^{0.9}\right)^{\varepsilon n} + e^{-cn} + (C_1\tau)^{\varepsilon n} \\
&\leq \left[C\varepsilon^{-1.4}t^{0.45}\right]^{\varepsilon n} + e^{-cn} + (Ct^{0.5})^{\varepsilon n}.
\end{aligned}$$

It remains to check that the last two terms of the expression above can be absorbed into the first one. This is obvious for the third term, and follows from (6.1) for the second one as we assumed that $\varepsilon < c$. This completes the proof of Theorem 6.1. □

References

[1] G. Anderson, A. Guionnet, O. Zeitouni, *An introduction to random matrices.* Cambridge Studies in Advanced Mathematics, 118. Cambridge University Press, Cambridge, 2010.

[2] S. Arora, A. Bhaskara, *Eigenvectors of random graphs: delocalization and nodal domains,* manuscript, 2011.

[3] Z. Bai, J. Silverstein, *Spectral analysis of large dimensional random matrices.* Second edition. Springer Series in Statistics. Springer, New York, 2010.

[4] Z. Bai, Z. D., J. Silverstein, Y. Yin, *A note on the largest eigenvalue of a large dimensional sample covariance matrix,* J. Multivariate Anal. 26 (1988), 166–168.

[5] F. Benaych-Georges, S. Péché, *Localization and delocalization for heavy tailed band matrices*, Annales Inst. H. Poincaré, to appear.

[6] C. Bordenave, A. Guionnet, *Localization and delocalization of eigenvectors for heavy-tailed random matrices*, Probab. Theory Related Fields 157 (2013), 885–953.

[7] C. Cacciapuoti, A. Maltsev, B. Schlein, *Local Marchenko-Pastur law at the hard edge of sample covariance matrices*, Journal of Mathematical Physics, to appear.

[8] K. R. Davidson, S. Szarek, *Local operator theory, random matrices and Banach spaces.* In "Handbook on the Geometry of Banach spaces," Volume 1, W.B. Johnson, J. Lindenstrauss eds., Elsevier Science 2001, pp. 317–366.

[9] Y. Dekel, J. R. Lee, N. Linial, *Eigenvectors of random graphs: nodal domains,* Random Structures Algorithms 39 (2011), 39–58.

[10] R. Eldan, M. Rácz, T. Schramm, *Braess's paradox for the spectral gap in random graphs and delocalization of eigenvectors,* submitted. ArXiv: 1504.07669

[11] L. Erdös, *Universality for random matrices and log-gases*, Current developments in mathematics 2012, 59–132, Int. Press, Somerville, MA, 2013.

[12] L. Erdös, A. Knowles, *Quantum diffusion and eigenfunction delocalization in a random band matrix model*, Commun. Math. Phys. 303 (2011), 509–554.

[13] L. Erdös, A. Knowles, *Quantum diffusion and delocalization for band matrices with general distribution*, Annales Inst. H. Poincaré 12 (2011), 1227–1319.

[14] L. Erdös, A. Knowles, H.-T. Yau, J. Yin, *Spectral statistics of Erdös-Rényi graphs I: local semicircle law*, Annals of Probability, 41 (2013), no. 3B, 2279–2375.

[15] L. Erdös, A. Knowles, H.-T. Yau, J. Yin, *Spectral statistics of Erdös-Rényi graphs II: eigenvalue spacing and the extreme eigenvalues*, Comm. Math. Phys. 314 (2012), no. 3, 587–640.
[16] L. Erdös, A. Knowles, H.-T. Yau, J. Yin, *Delocalization and diffusion profile for random band matrices*, Comm. Math. Phys., to appear.
[17] L. Erdös, B. Schlein, H.-T. Yau, *Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices*, Ann. Probab. 37 (2009), 815–852.
[18] L. Erdös, B. Schlein, H.-T. Yau, *Local semicircle law and complete delocalization for Wigner random matrices*, Comm. Math. Phys. 287 (2009), 641–655.
[19] L. Erdös, H.-T. Yau, *Universality of local spectral statistics of random matrices*, Bull. Amer. Math. Soc. (N.S.) 49 (2012), 377–414.
[20] Y. Gordon, *On Dvoretzky's theorem and extensions of Slepian's lemma.* Israel seminar on geometrical aspects of functional analysis (1983/84), II, 25 pp., Tel Aviv Univ., Tel Aviv, 1984.
[21] Y. Gordon, —em Some inequalities for Gaussian processes and applications, Israel J. Math. 50 (1985), 265–289.
[22] B. Kashin, *The widths of certain finite-dimensional sets and classes of smooth functions* Izv. Akad. Nauk SSSR Ser. Mat., 41 (1977), 334–351.
[23] R. Latala, *Some estimates of norms of random matrices,* Proc. Amer. Math. Soc. 133 (2005), 1273–1282.
[24] A. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, *Smallest singular value of random matrices and geometry of random polytopes*, Adv. Math. 195 (2005), 491–523.
[25] H. Nguyen, T. Tao, V. Vu, *Random matrices: tail bounds for gaps between eigenvalues,* submitted. ArXiv: 1504.00396
[26] M. Rudelson, *Invertibility of random matrices: norm of the inverse,* Annals of Mathematics 168 (2008), 575–600.
[27] M. Rudelson, *Recent developments in non-asymptotic theory of random matrices*, Modern aspects of random matrix theory, 83–120, Proc. Sympos. Appl. Math., 72, Amer. Math. Soc., Providence, RI, 2014.
[28] M. Rudelson, R. Vershynin, *The Littlewood-Offord Problem and invertibility of random matrices,* Advances in Mathematics 218 (2008), 600–633.
[29] M. Rudelson, R. Vershynin, *The least singular value of a random square matrix is* $O(n^{-1/2})$*,* Comptes rendus de l'Académie des sciences - Mathématique 346 (2008), 893–896.
[30] M. Rudelson, R. Vershynin, *Smallest singular value of a random rectangular matrix,* Communications on Pure and Applied Mathematics 62 (2009), 1707–1739.
[31] 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.
[32] M. Rudelson, R. Vershynin, *Small ball probabilities for linear images of high dimensional distributions,* International Mathematics Research Notices, to appear. ArXiv: 1402.4492
[33] M. Rudelson, R. Vershynin, *Delocalization of eigenvectors of random matrices with independent entries*, Duke Math. Journal, to appear. ArXiv: 1306.2887
[34] E. M. Stein, G. Weiss, *Introduction to Fourier analysis on Euclidean spaces.* Princeton Mathematical Series, No. 32. Princeton University Press, Princeton, N.J., 1971.
[35] T. Tao, *Topics in random matrix theory.* Graduate Studies in Mathematics, 132. American Mathematical Society, Providence, RI, 2012.
[36] T. Tao,V. Vu, *The condition number of a randomly perturbed matrix.* STOC'07–Proceedings of the 39th Annual ACM Symposium on Theory of Computing, 248–255, ACM, New York, 2007.
[37] T. Tao,V. Vu, *Inverse Littlewood-Offord theorems and the condition number of random discrete matrices,* Ann. of Math. (2) 169 (2009), 595–632.
[38] 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. (N.S.) 46 (2009), 377–396.
[39] T. Tao, V. Vu, *Random matrices: universality of ESDs and the circular law,* with an appendix by Manjunath Krishnapur. Ann. Probab. 38 (2010), 2023–2065.
[40] T. Tao, V. Vu, *Random matrices: the universality phenomenon for Wigner ensembles,* Modern aspects of random matrix theory, 121–172, Proc. Sympos. Appl. Math., 72, Amer. Math. Soc., Providence, RI, 2014.
[41] L. V. Tran, V. Vu, K. Wang, *Sparse random graphs: eigenvalues and eigenvectors*, Random Structures Algorithms 42 (2013), 110–134.
[42] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices.* Compressed sensing, 210–268, Cambridge Univ. Press, Cambridge, 2012. ArXiv: 1011.3027
[43] R. Vershynin, *Invertibility of symmetric random matrices,* Random Structures and Algorithms 44 (2014), 135–182. ArXiv: 1102.0300
[44] V. Vu, K. Wang, *Random weighted projections, random quadratic forms and random eigenvectors*, Random Structures and Algorithms, to appear. ArXiv: 1306.3099

Department of Mathematics, University of Michigan, 530 Church St., Ann Arbor, MI 48109, U.S.A.
E-mail address: {rudelson, romanv}@umich.edu