Delocalization of eigenvectors of random matrices

Mark Rudelson

Abstract. Let $x \in S^{n-1}$ be a unit eigenvector of an $n \times n$ random matrix. This vector is delocalized if it is distributed roughly uniformly over the real or complex sphere. This intuitive notion can be quantified in various ways. In these lectures, we will concentrate on the *no-gaps delocalization*. This type of delocalization means that with high probability, any non-negligible subset of the support of x carries a non-negligible mass. Proving the no-gaps delocalization requires establishing small ball probability bounds for the projections of random vector. Using Fourier transform, we will prove such bounds in a simpler case of a random vector having independent coordinates of a bounded density. This will allow us to derive the no-gaps delocalization for matrices with random entries having a bounded density. In the last section, we will discuss the applications of delocalization to the spectral properties of Erdős-Rényi random graphs.

1. Introduction

Let G be a symmetric random matrix with independent above the diagonal normal random entries having expectation 0 and variance 1 ($N(0,1)$ random variables). The distribution of such matrices is invariant under the action of the orthogonal group $O(n)$. Consider a unit eigenvector $v \in S^{n-1}$ of this matrix. The distribution of the eigenvector should share the invariance of the distribution of the matrix itself, so v is uniformly distributed over the real unit sphere $S_{\mathbb{R}}^{n-1}$. Similarly, if Γ is an $n \times n$ complex random matrix with independent entries whose real and imaginary part are independent $N(0,1)$ random variables, then the distribution of Γ is invariant under the action of the unitary group $U(n)$. This means that any unit eigenvector of Γ is uniformly distributed over the complex unit sphere $S_{\mathbb{C}}^{n-1}$. For a general distribution of entries, we cannot expect such strong invariance properties. Indeed, if the entries of the matrix are random variables taking finitely many values, the eigenvectors will take finitely many values as well, so the invariance is impossible. Nevertheless, as n increases, a central limit phenomenon should kick in, so the distribution of an eigenvector should be approximately uniform. This vague idea called delocalization can be made mathematically precise in a number of ways. Some of these formalizations use

2010 *Mathematics Subject Classification.* Primary 60B20; Secondary 05C80.
Key words and phrases. random matrices, eigenvectors, random graphs.
Partially supported by NSF grant, DMS-1464514.

the local structure of a vector. One can fix in advance several coordinates of the eigenvector and show that the joint distribution of these coordinates approaches the distribution of a properly normalized gaussian vector, see [6].

In these notes, we adopt a different approach to delocalization coming from the non-asymptotic random matrix theory. The asymptotic theory is concerned with establishing limit distributions of various spectral characteristics of a family of random matrices when the sizes of these matrices tend to infinity. In contrast to it, the non-asymptotic theory strives to obtain explicit, valid with high probability bounds for the matrices of a large fixed size. This approach is motivated by applications primarily to convex geometry, combinatorics, and computer science. For example, while analyzing performance of an algorithm solving a noisy linear system, one cannot let the size of the system go to infinity. An interested reader can find an introduction to the non-asymptotic theory in [21,22,27]. In this type of problems, strong probabilistic guarantees are highly desirable, since one typically wants to show that many "good" events occur at the same time. This will be the case in our analysis of the delocalization behavior as well.

We will consider the global structure of the eigenvector of a random matrix controlling all coordinates of it at once. The most classical type of such delocalization is the ℓ_∞ norm bound. If $v \in S^{n-1}$ is a random vector uniformly distributed over the unit sphere, then with high probability, all its coordinates are small. This is easy to check using the concentration of measure. Indeed, the vector v has the same distribution as $g/\|g\|_2$, where $g \in \mathbb{R}^n$ or $\mathbb{C}^n$ is the standard Gaussian vector, i.e., a vector with the independent $N(0,1)$ coordinates. By the concentration of measure, $\|g\|_2 = \sqrt{n}(1+o(1))$ with high probability. Also, since the coordinates of g are independent,

$$\mathbb{E}\,\|g\|_\infty = \mathbb{E}\max_{j\in[n]}|g_j| \leqslant C\sqrt{\log n},$$

and the measure concentration yields that $\|g\|_\infty \leqslant C'\sqrt{\log n}$ with high probability. Therefore, with high probability,

$$\|v\|_\infty \leqslant C\frac{\sqrt{\log n}}{\sqrt{n}}.$$

Here and below, $C, \bar{C}, C', c$, etc. denote absolute constants which can change from line to line, or even within the same line.

One would expect to have a similar ℓ_∞ delocalization for a general random matrix. The bound

$$\|v\|_\infty \leqslant C\frac{\log^c n}{\sqrt{n}}$$

for unit eigenvectors was proved in [12,13] for Hermitian random matrices and in [24] for random matrices all whose entries are independent. Moreover, in the case of the Hermitian random matrix with entries having more than four moments, the previous estimate has been established with the optimal power of the logarithm $c = 1/2$, see [14,28]. We will not discuss the detailed history and

the methods of obtaining the ℓ_∞ delocalization in these notes, and refer a reader to a comprehensive recent survey [20].

Instead, we are going to concentrate on a different manifestation of the delocalization phenomenon. The ℓ_∞ delocalization rules out peaks in the distribution of mass among the coordinates of a unit eigenvector. In particular, it means that with high probability, most of the mass, i.e., ℓ_2 norm of a unit eigenvector cannot be localized on a few coordinates. We will consider a complementary phenomenon, namely ruling out chasms in the mass distribution. More precisely, we aim at showing that with high probability, any non-negligible set of the coordinates of a unit eigenvector carries a relatively large mass. We call this property of lack of almost empty zones in the support of the eigenvector the *no-gaps delocalization* property.

No-gaps delocalization property holds for the eigenvectors of many natural classes of random matrices. This includes matrices, whose all entries are independent, random real symmetric and skew-symmetric matrices, random complex hermitian matrices with independent real and imaginary parts of the entries, etc. We formulate the explicit assumption on the dependencies of the entries below.

Assumption 1.0.1 (Dependencies 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.

Fixing the imaginary part in Assumption 1.0.1 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 that the operator norm of the matrix A satisfies $\|A\| = O(\sqrt{n})$ with high probability. This natural condition holds, in particular, if the entries of A have mean zero and bounded fourth moments (see, e.g., [26]). To make this rigorous, we fix a number $M \geqslant 1$ and introduce the boundedness event

$$\mathcal{B}_{A,M} := \left\{ \|A\| \leqslant M\sqrt{n} \right\}. \tag{1.0.2}$$

We will give two versions of the no-gaps delocalization theorem, for absolutely continuous entries with bounded density and for general entries. Although the second case includes the first, the results assuming bounded density are stronger and the proofs significantly easier. We formulate the first assumption explicitly.

Assumption 1.0.3 (Continuous distributions)**.** We assume that the real parts of the matrix entries have densities bounded by some number $K \geqslant 1$.

Under Assumptions 1.0.1 and 1.0.3, 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 [25].

Theorem 1.0.4 (Delocalization: continuous distributions)**.** *Let A be an $n \times n$ random matrix which satisfies Assumptions 1.0.1 and 1.0.3. Choose $M \geqslant 1$. Let $\varepsilon \in [8/n, 1)$ and $s > 0$. Then, the following event holds with probability at least*

$$1 - (Cs)^{\varepsilon n} - \mathbb{P}\left(\mathcal{B}_{A,M}^{c}\right).$$

Every eigenvector v of A satisfies

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

Here $C = C(K, M) \geqslant 1$, and $[n]$ denotes the set of all natural numbers from 1 to n.

Note that we do not require any moments for the matrix entries, so heavy-tailed distributions are allowed. However, the boundedness assumption formalized by (1.0.2) implicitly yields some upper bound on the tails.

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

Extending Theorem 1.0.4 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 1.0.5 (General distribution of entries)**.** We assume that the real parts of the matrix entries are distributed identically with a random variable ξ, that satisfies

$$\sup_{u \in \mathbb{R}} \mathbb{P}\left\{|\xi - u| \leqslant 1\right\} \leqslant 1 - p, \quad \mathbb{P}\left\{|\xi| > K\right\} \leqslant p/2 \quad \text{for some } K, p > 0. \tag{1.0.6}$$

Assumption 1.0.5 holds for any non-constant random variable with some p, K after a proper scaling. Its meaning therefore is not to restrict the class of random variables, but to introduce parameters p and K which will be used in the formulation of Theorem 1.0.7 below.

With Assumption 1.0.3 replaced by Assumption 1.0.5, we can prove a general no-gaps delocalization result [25].

Theorem 1.0.7 (Delocalization: general distributions)**.** *Let A be an $n \times n$ random matrix which satisfies Assumptions 1.0.1 and 1.0.5. Let $M \geqslant 1$. Let $\varepsilon \geqslant 1/n$ and $s \geqslant c_1 \varepsilon^{-7/6} n^{-1/6} + e^{-c_2/\sqrt{\varepsilon}}$. Then, the following event holds with probability at least*

$$1 - (Cs)^{\varepsilon n} - \mathbb{P}\left(\mathcal{B}_{A,M}^{c}\right).$$

Every eigenvector v of A satisfies

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

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

Remark 1.0.8. The assumption on s appearing in Theorem 1.0.7 forces us to consider only $\varepsilon \geqslant Cn^{-1/7}$ in contrast with Theorem 1.0.4 which yields non-trivial

results as long as $\varepsilon \geqslant 8/n$. This assumption can be probably relaxed if one replaces using Berry-Esseen Theorem in the proof of Theorem 1.0.7 in [25] by a more complicated argument based on the least common denominator.

Remark 1.0.9. The proof of Theorem 1.0.7 presented in [25] can be modified to allow an extension to random matrices shifted by a constant multiple of the all ones matrix $\mathbf{1}_n$. More precisely, for a given $\mu \in \mathbb{C}$, the event described in the theorem holds with probability at least $1 - (Cs)^{\varepsilon n} - \mathbb{P}\left(\mathcal{B}^c_{A-\mu\mathbf{1}_n,M}\right)$. This allows to consider random matrices with identically distributed entries having a non-zero expectation, in particular, with Bernoulli(p) entries for p being a constant. Moreover, tracing the proof appearing in [25], one can see that the constants c_k and C depend polynomially on p, which allows to extend no-gaps delocalization to matrices with i.i.d. Bernoulli entries for $p = \Omega(n^{-c'})$ for some absolute constant $c' \in (0,1)$.

Remark 1.0.10. The no-gaps delocalization phenomenon holds also for any unit vector which is a linear combination of eigenvectors whose eigenvalues are not too far apart, see Remark 2.1.8 for the details.

Acknowledgement These notes are based on the mini-courses given at Hebrew University of Jerusalem and at PCMI Summer School on Random Matrices. The author is grateful to Alex Samorodnitsky, Alexey Borodin, Ivan Corwin, and Alice Guionnet for their hospitality and an opportunity to present this material. The author is grateful to Feng Wei for for running problem sessions at PCMI which were an integral part of the mini-course. He would also like to thank Anirban Basak, Konstantin Tikhomirov, and Feng Wei for careful reading of the manuscript and several suggestions on improving the presentation.

2. Reduction of no-gaps delocalization to invertibility of submatrices

2.1. From no-gaps delocalization to the smallest singular value bounds The first step in proving no-gaps delocalization is pretty straightforward. Let us consider the toy case when there exists a unit eigenvector u of the matrix A with $u_j = 0$ for all $j \in J$, where J is some subset of $[n]$. If we denote the corresponding eigenvalue by λ and the submatrix of A with columns from the set J^c by A_{J^c}, then we have that $(A_{J^c} - \lambda I_{J^c})u_{J^c} = 0$ so the kernel of $A_{J^c} - \lambda I_{J^c}$ is non-trivial. Here, $A_{J^c} - \lambda I_{J^c}$ is a "tall" matrix with the number of rows larger than the number of columns. A linear operator defined by a tall rectangular random matrix with sufficiently many independent entries is an injection with high probability. This means that the event that the probability of this "toy" case should be small. This idea is not directly applicable since the random eigenvalue λ depends on all entries of the matrix A, but this obstacle is easy to circumvent by discretizing the set of plausible values of λ and considering a deterministic λ from this discretization. If the probability that $A_{J^c} - \lambda I_{J^c}$ is close to a singular matrix is small for any fixed

λ, we can use the union bound over the dicretisation along with approximation to show that, with high probability, the matrix $A_{J^c} - \lambda I_{J^c}$ has a trivial kernel for all λ from this plausible set simultaneously. This would imply the same statement for a random λ allowing us to avoid using hard to obtain information about its distribution except for a very rough bound defining the plausible set.

To implement this idea for a real setup, recall the definition of the singular values of a matrix. Let B be a real or complex $N \times n$ matrix, $N \geqslant n$. The singular values of B are defined as the square roots of eigenvalues of B^*B arranged in the decreasing order:

$$s_1(B) \geqslant s_2(B) \geqslant \ldots \geqslant s_n(B) \geqslant 0.$$

If B is real, and we consider this matrix as a linear operator $B : \mathbb{R}^n \to \mathbb{R}^N$, then the image of the Euclidean unit ball will be an ellipsoid whose semi-axes have lengths $s_1(B), \ldots, s_n(B)$. The extreme singular values have also an analytic meaning with

$$s_1(B) = \max_{x \in S^{n-1}} \|Bx\|_2 \quad \text{and}$$
$$s_n(B) = \min_{x \in S^{n-1}} \|Bx\|_2 ,$$

so $s_1(B) = \|B\|$, the operator norm of B, and $s_n(B)$ is the distance from B to the set of matrices of a rank smaller than n in the operator norm. Throughout these notes, we will also denote the smallest singular value by $s_{\min}(B)$. We will also abbreviate $A - \lambda I$ to $A - \lambda$.

Let us introduce the event that one of the eigenvectors is localized. Define the *localization event* by

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

Since we assume in Theorem 1.0.4 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:

$$\mathbb{P}\left\{ \mathrm{Loc}(A, \varepsilon, (\varepsilon s)^6) \text{ and } \mathcal{B}_{A,M} \right\} \leqslant (cs)^{\varepsilon n}. \tag{2.1.1}$$

The following proposition reduces proving a delocalization result like (2.1.1) to an invertibility bound.

Proposition 2.1.2 (Reduction of delocalization to invertibility)**.** *Let A be an $n \times n$ random matrix with arbitrary distribution. Let $M \geqslant 1$ and $\varepsilon, p_0, \delta \in (0, 1/2)$. Assume that for any number $\lambda_0 \in \mathbb{C}$, $|\lambda_0| \leqslant M\sqrt{n}$, and for any set $I \subset [n]$, $|I| = \varepsilon n$, we have*

$$\mathbb{P}\left\{ s_{\min}\big((A - \lambda_0)_{I^c}\big) \leqslant 8\delta M\sqrt{n} \text{ and } \mathcal{B}_{A,M} \right\} \leqslant p_0. \tag{2.1.3}$$

Then

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

Proof. Assume both the localization event and the boundedness event $\mathcal{B}_{A,M}$ occur. Use the definition of $\mathrm{Loc}(A, \varepsilon, \delta)$ to 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

$$0 = (A-\lambda)v = (A-\lambda)_I v_I + (A-\lambda)_{I^c} v_{I^c}. \tag{2.1.4}$$

By triangle inequality, this yields

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

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

$$\|(A-\lambda)_{I^c} v_{I^c}\|_2 \leqslant 2M\delta\sqrt{n}. \tag{2.1.5}$$

This happens for some λ in the disc $\{z \in \mathbb{C} : |z| \leqslant M\sqrt{n}\}$. 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

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

Choose $\lambda_0 \in \mathcal{N}$ so that $|\lambda_0 - \lambda| \leqslant 2M\delta\sqrt{n}$. By (2.1.5), we have

$$\|(A-\lambda_0)_{I^c} v_{I^c}\|_2 \leqslant 4M\delta\sqrt{n}. \tag{2.1.6}$$

Since $\|v_I\|_2 \leqslant \delta \leqslant 1/2$, we have $\|v_{I^c}\|_2 \geqslant \|v\|_2 - \|v_I\|_2 \geqslant 1/2$. Therefore, (2.1.6) implies that

$$s_{\min}((A-\lambda_0)_{I^c}) \leqslant 8M\delta\sqrt{n}. \tag{2.1.7}$$

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 (2.1.7) holds. Furthermore, for fixed I and λ_0, assumption (2.1.3) states that (2.1.7) together with $\mathcal{B}_{A,M}$ hold with probability at most p_0. So by the union bound we conclude that

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

This completes the proof of the proposition. □

Remark 2.1.8. A simple analysis of the proof of Proposition 2.1.2 shows that it holds not only for eigenvectors of the matrix A, but for its *approximate* eigenvectors as well. Namely, instead of the event $\mathrm{Loc}(A, \varepsilon, \delta)$ one can consider the following event

$$\widetilde{\mathrm{Loc}}(A, \varepsilon, \delta) := \Big\{\exists v \in S_{\mathbb{C}}^{n-1},\ \exists \lambda \in \mathbb{C}\ |\lambda| \leqslant M\sqrt{n}\ \exists I \subset [n],\ |I| = \varepsilon n : \\ \|(A - \lambda I)v\|_2 \leqslant M\delta\sqrt{n} \text{ and } \|v_I\|_2 < \delta\Big\}.$$

This event obeys the same conclusion as $\mathrm{Loc}(A, \varepsilon, \delta)$:

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

Indeed, equation (2.1.4) is replaced by

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

where w is a vector of a norm not exceeding $M\delta\sqrt{n}$. This in turn results in replacing $2M\delta\sqrt{n}$ by $3M\delta\sqrt{n}$ in (2.1.5) and $3M\delta\sqrt{n}$ by $4M\delta\sqrt{n}$ in (2.1.6). This observation shows, in particular, that the no-gaps delocalization phenomenon holds for any unit vector which is a linear combination of eigenvectors whose eigenvalues are at most $M\delta\sqrt{n}$ apart.

2.2. The ε-net argument. We have reduced the proof of the no-gaps delocalization to establishing quantitative invertibility of a matrix whose number of rows is larger than number of columns. This problem has been extensively studied, so before embarking on the real proof, let us check whether we can apply an elementary bound based on the discretization of the sphere. Assume for simplicity that all entries of the matrix A are real and independent, and the entries are centered and of the unit variance. We will formulate the result in a bigger generality than we need at this moment.

Lemma 2.2.1. *Let $M > 0$ and let A be an $m \times n$ matrix with real independent entries $A_{i,j}$ satisfying*

$$\mathbb{E}a_{i,j} = 0, \quad \mathbb{E}a_{i,j}^2 = 1, \quad \textit{and} \quad \mathbb{E}a_{i,j}^4 \leqslant C.$$

Let E be a linear subspace of $\mathbb{R}^n$ of dimension

$$k = \dim(E) < c\frac{m}{\log(2+n/m)}.$$

Then with probability at least $1 - \exp(-c'm) - \mathbb{P}\left\{\mathcal{B}^c_{A,M}\right\}$, all vectors $x \in E$ satisfy

$$\|Ax\|_2 \geqslant c\sqrt{m}.$$

The parameters c, c' here may depend on C.

Recall the definition of the ε-net. Let (X, d) be a metric space, and let $\varepsilon > 0$. A set $\mathcal{N} \subset X$ is called an ε-net for a set $V \subset X$ if for any $x \in V$, there exists $y \in \mathcal{N}$ with $d(x, y) < \varepsilon$. We will consider ε-nets for various subsets of $\mathbb{R}^n$ in the Euclidean metric below. These nets are useful in discretization of continuous structures. For instance, it is easy to show that if $\mathcal{N} \subset S^{n-1}$ is a $(1/2)$-net in the unit sphere S^{n-1}, then the operator norm of an $m \times n$ matrix A and the maximum of the Euclidean norm of Ax over the net are commensurate:

$$\max_{x\in\mathcal{N}} \|Ax\|_2 \leqslant \|A\| \leqslant 2\max_{x\in\mathcal{N}} \|Ax\|_2 .$$

The proof of Lemma 2.2.1 is based on the ε-net argument. To implement it, we need an elementary lemma.

Lemma 2.2.2. *Let $\varepsilon \in (0, 1]$ and let $V \subset S^{k-1}_{\mathbb{R}}$ be any set. The set V contains an ε-net of cardinality at most $(1 + 2/\varepsilon)^k$.*

The proof of Lemma 2.2.2 relies on a simple volume comparison. Notice that the balls of radius $\varepsilon/2$ centered at the points of the net are disjoint. On the other

hand, the union of these balls is contained in the ball of radius $(1+\varepsilon/2)$ centered at 0. We leave the details to a reader.

Proof of Lemma 2.2.1. Let $\varepsilon > 0$. It is enough to prove the norm bound for all vectors of $V := E \cap S^{n-1}$. Since the dimension of E is k, this set admits an ε-net $\mathcal{N}$ of cardinality $(1+2/\varepsilon)^k$. Let $y \in \mathcal{N}$, and let $z_j = (Ay)_j$ be the j-th coordinate of the vector Ay.

The Paley–Zygmund inequality asserts that a random variable $Y \geqslant 0$ satisfies

$$\mathbb{P}\left\{Y > t\right\} \geqslant \frac{(\mathbb{E}Y - t)^2}{\mathbb{E}Y^2} \quad \text{for any } t \in (0, \mathbb{E}Y).$$

If $Y = z_j^2$, the assumptions on $a_{i,j}$ imply $\mathbb{E}Y = 1$ and $\mathbb{E}Y^2 \leqslant C'$. Applying the Paley–Zygmund inequality with $t = 1/2$, we conclude that $\mathbb{P}\left\{|z_j| \geqslant 1/2\right\} \geqslant c$. Using Chernoff's inequality, we derive that

$$\begin{aligned}\mathbb{P}\left\{\|Ay\|_2 \leqslant \frac{1}{4}\sqrt{m}\right\} &= \mathbb{P}\left\{\sum_{j=1}^m |z_j|^2 \leqslant \frac{1}{16}m\right\} \\ &\leqslant \mathbb{P}\left\{|\{j : |z_j| \leqslant 1/2\}| \geqslant \frac{m}{2}\right\} \leqslant \exp(-c_2 m). \qquad (2.2.3)\end{aligned}$$

In combination with the union bound, this yields

$$\mathbb{P}\left\{\exists y \in \mathcal{N} \ \|Ay\|_2 \leqslant (1/4)\sqrt{m}\right\} \leqslant (1+2/\varepsilon)^k \exp(-c_2 m). \qquad (2.2.4)$$

Let Ω be the event that $\|Ay\|_2 > (1/4)\sqrt{m}$ for all $y \in \mathcal{N}$ intersected with $\mathcal{B}_{A,M}$. Assuming that Ω occurs, we will show that the matrix is invertible on the whole V. To this end, take any $x \in V$, and find $y \in \mathcal{N}$ such that $\|x-y\|_2 < \varepsilon$. Then

$$\|Ax\|_2 \geqslant \|Ay\|_2 - \|A\| \cdot \|x-y\|_2 \geqslant \frac{1}{4}\sqrt{m} - M\sqrt{n}\cdot\varepsilon \geqslant \frac{1}{8}\sqrt{m}$$

if we set

$$\varepsilon = \frac{1}{8M} \cdot \sqrt{\frac{m}{n}} \wedge 1.$$

It remains to estimate the probability that Ω does not occur. By (2.2.4),

$$\mathbb{P}\left\{\Omega^c\right\} \leqslant \exp(k\log(1+2/\varepsilon) - c_2 m) + \mathbb{P}\left\{\mathcal{B}_{A,M}^c\right\} \leqslant \exp\left(-\frac{c_2}{2}m\right) + \mathbb{P}\left\{\mathcal{B}_{A,M}^c\right\}$$

if we choose

$$k \leqslant c\frac{m}{\log(2+n/m)}. \qquad \square$$

Comparing the bound (2.1.3) needed to establish delocalization with the smallest singular value estimate of lemma 2.2.1, we see several obstacles preventing the direct use of the ε-net argument.

Lack of independence As we recall from Assumption 1.0.1, we are looking for ways to control symmetric and non-symmetric matrices simultaneously. This forces us to consider random matrices with dependent entries making Chernoff's inequality inapplicable.

Small exceptional probability required Lemma 2.2.1 provides the smallest singular value bound for rectangular matrices whose number of rows is significantly greater than the number of columns. If we are to apply it in combination with Proposition 2.1.2, we would have to assume in addition that $\varepsilon > 1 - \varepsilon_0$ for some small $\varepsilon_0 < 1$. Considering smaller values of ε would require a small ball probability bound better than (2.2.3) that we used in the proof. We will show that such bound is possible to obtain in the case when the entries have a bounded density. In the general case, however, such bound is unavailable. Indeed, if the entries of the matrix may take the value 0 with a positive probability, then $\mathbb{P}(Ae_1 = 0) = \exp(-cm)$, which shows that the bound (2.2.3) is, in general, optimal. Overcoming this problem for a general distribution would require a delicate stratification of the unit sphere according to the number-theoretic structure of the coordinates of a vector governing the small ball probability bound.

A closer look at Proposition 2.1.2 demonstrates that the demands for a small ball probability bound are even higher. We need that the delocalization result, and thus the invertibility bound (2.1.6), 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 in (2.1.3) to be at most $\varepsilon^{\varepsilon n}$. Such small exceptional probabilities (smaller than $e^{-\varepsilon n}$) are hard to achieve in the general case.

Complex entries Even if the original matrix is real, its eigenvalues may be complex. This observation forces us to work with complex random matrices. Extending the known invertibility results to complex matrices poses two additional challenges. First, in order to preserve the matrix-vector multiplication, we replace a complex $n \times m$ random matrix $B = R + iT$ by the real $2m \times 2n$ 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 dependencies of the entries. Besides that, we encounter a major problem while trying to apply the ε-net argument to prove the smallest singular value bound. Indeed, since we have to consider a real $2m \times 2n$ 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 m, so the small ball probability will be exponential in terms of m. If $m < 2n$, the union bound would not be applicable.

Each of these obstacles requires a set of rather advanced tools to deal with in general case, i.e. under Assumption 1.0.5. Fortunately, under Assumption 1.0.3, these problems can be addressed in a much easier way allowing a short and rather non-technical proof. For this reason, we are going to concentrate on the continuous density case below.

3. Small ball probability for the projections of random vectors

3.1. Density of a marginal of a random vector. The proof of the no-gaps delocalization theorem requires a result on the distribution of the marginals of a random

vector which is of an independent interest. To simplify the presentation, we will consider a vector with independent coordinates having a bounded density. Let $X = (X_1, \dots, X_n)$ be independent real valued random variables with densities $f_{X_1}, \dots, f_{X_n}$ satisfying

$$f_{X_j}(t) \leqslant K \quad \text{for all } j \in [n],\ t \in \mathbb{R}.$$

The independence implies that the density of the vector is the product of the densities of the coordinates, and so, $f_X(x) \leqslant K^n$ for all $x \in \mathbb{R}^n$. Obviously, we can extend the previous observation to the coordinate projections of X showing that $f_{P_J X}(y) \leqslant K^{|J|}$ for any set $J \subset [n]$ and any $y \in \mathbb{R}^J$ with P_J standing for the coordinate projection of $\mathbb{R}^n$ to $\mathbb{R}^J$. It seems plausible that the same property should hold for the densities of all orthogonal projections to subspaces $E \subset \mathbb{R}^n$ with the dimension of E playing the role of $|J|$. Yet, a simple example shows that this statement fails even in dimension 2. Let X_1, X_2 be random variables uniformly distributed on the interval $[-1/2, 1/2]$, and consider the projection on the subspace $E \subset \mathbb{R}^2$ spanned by the vector $(1, 1)$. Then $Y = P_E X$ is the normalized sum of the coordinates of X:

$$P_Y = \frac{\sqrt{2}}{2}(X_1 + X_2).$$

A direct calculation shows that $f_Y(0) = \sqrt{2} > 1$. A delicate result of Ball [2] shows that this is the worst case for the uniform distribution. More precisely, consider a vector $X \in \mathbb{R}^n$ with i.i.d. coordinates uniformly distributed in the interval $[-1/2, 1/2]$. Then the projection of X onto any one-dimensional subspace $E = \operatorname{span}(a)$ with $a = (a_1, \dots, a_n) \in S^{n-1}$ is a weighted linear combination of the coordinates: $P_E(X) = \sum_{j=1}^n a_j X_j$. The theorem of Ball asserts that the density of such linear combination does not exceed $\sqrt{2}$ making $a = (\sqrt{2}/2, \sqrt{2}/2, 0, \dots, 0)$ the worst sequence of weights. This result can be combined with a theorem of Rogozin claiming that the density of a linear combination of independent random variables increases the most if these variables are uniformly distributed. This shows that if the coordinates of X are independent absolutely continuous random variables having densities uniformly bounded by K, then the density of $Y = \sum_{j=1}^n a_j X_j$ does not exceed $\sqrt{2}K$ for any $a = (a_1, \dots, a_n) \in S^{n-1}$.

Instead of discussing the proofs of the theorems of Ball and Rogozin, we will present here a simpler argument due to Ball and Nazarov [4] showing that the density of Y is bounded by CK for some unspecified absolute constant C. Moreover, we will show that this fact allows a multidimensional extension which we formulate in the following theorem [23].

Theorem 3.1.1 (Densities of projections)**.** *Let $X = (X_1, \dots, X_n)$ where X_i are real-valued independent random variables. Assume that the densities of X_i are bounded by K almost everywhere. Let P be the orthogonal projection in $\mathbb{R}^n$ onto a d-dimensional subspace. Then the density of the random vector PX is bounded by $(CK)^d$ almost everywhere.*

To avoid ambiguity, let us mention that we consider the density in the range $\mathbb{PR}^n$, as the density in $\mathbb{R}^n$ does not exist.

This theorem shows that the density bound K^d for coordinate projections holds also for general ones if we include a multiplicative factor depending only on the dimension. Recently, Livshyts et al. [18] proved a multidimensional version of Rogozin's theorem. Combining it with the multidimensional version of Ball's theorem [3], one can show that the optimal value of the constant C is $\sqrt{2}$ as in the one-dimensional case.

Proof. We will start the proof from the one-dimensional case. The proof in this case is a nice illustration of the power of characteristic functions approach in deriving the small ball and density estimates. As before, we restate the one-dimensional version of the theorem as a statement about the density of a linear combination.

Step 1. Linear combination of independent random variables.

Fix $X_1, \dots, X_n$ be real-valued independent random variables with densities bounded by K almost everywhere and $a_1, \dots, a_n$ real numbers with

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

Then the density of $\sum_{j=1}^n a_j X_j$ is bounded by CK almost everywhere.

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

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

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

After these reductions, we proceed to bounding $f_S(0)$ in terms of the characteristic function

$$\phi_S(t) = \mathbb{E} e^{itS}.$$

We intend to use the Fourier inversion formula

$$f_S(0) = \frac{1}{2\pi} \int_{\mathbb{R}} \phi_S(x)\, dx.$$

This formula requires the assumption that $\phi_S \in L_1(\mathbb{R})$, while we only know that $\|\phi_S\|_\infty \leqslant 1$. This, however, is not a problem. We can add an independent $N(0, \sigma^2)$ random variable to each coordinate of X. In terms of the characteristic functions, this amounts to multiplying each $\phi_{X_j} \in L_\infty(\mathbb{R})$ by a scaled Gaussian density making it an L_1-function. The bound on the density we are going to obtain will not depend on σ which would allow taking $\sigma \to 0$.

By independence of the coordinates of X, $\phi_S(x) = \prod_j \phi_{X_j}(a_j t)$. Combining it with the Fourier inversion formula, we obtain

$$f_S(0) = \frac{1}{2\pi}\int_{\mathbb{R}} \prod_j \phi_{X_j}(a_j x)\,dx \leqslant \frac{1}{2\pi}\prod_j \Big(\int_{\mathbb{R}} |\phi_{X_j}(a_j x)|^{1/a_j^2}\,dx\Big)^{a_j^2}, \tag{3.1.2}$$

where we used Holder's inequality with exponents $1/a_j^2$ whose reciprocals sum up to 1.

We will estimate each integral appearing in the right hand side of (3.1.2) separately. Denote by λ the Lebesgue measure on $\mathbb{R}$. Using the Fubini theorem, we can rewrite each integral as

$$\frac{1}{a_j}\cdot\int_{\mathbb{R}} |\phi_{X_j}(x)|^{1/a_j^2}\,dx = \int_0^1 \frac{1}{a_j^3}\cdot t^{1/a_j^2-1}\lambda\{x : |\phi_{X_j}(x)| > t\}\,dt. \tag{3.1.3}$$

To estimate the last integral, we need a bound on the measure of points where the characteristic function is large. Such bound is provided in the lemma below.

Lemma 3.1.4 (Decay of a characteristic function)**.** *Let X be a random variable whose density is bounded by 1. Then the characteristic function of X satisfies*

$$\lambda\{x : |\phi_X(x)| > t\} \leqslant \begin{cases} \frac{2\pi}{t^2}, & t \in (0, 3/4) \\ C\sqrt{1-t^2}, & t \in [3/4, 1]. \end{cases}$$

The value $3/4$ in Lemma 3.1.4 was chosen arbitrarily. It can be replaced by any other number $t_0 \in (0,1)$ at the price of changing the constant C.

Let us postpone the proof of the lemma for a moment and finish the proof of the one-dimensional case of Theorem 3.1.1. Fix $j \in [n]$ and denote for shortness $p = 1/a_j^2 \geqslant 4$. Combining Lemma 3.1.4 and (3.1.3), we obtain

$$\begin{aligned}
&\frac{1}{a_j}\cdot\int_{\mathbb{R}} |\phi_{X_j}(x)|^{1/a_j^2}\,dx \\
&\leqslant p^{3/2}\cdot\left(\int_0^{3/4} t^{p-1}\cdot\frac{2\pi}{t^2}\,dt + \int_{3/4}^1 t^{p-1}\cdot C\sqrt{1-t^2}\,dt\right) \\
&\leqslant p^{3/2}\cdot\left(\frac{2\pi}{p-2}\cdot(3/4)^{p-2} + C\int_0^{\sqrt{7}/4}(1-s^2)^{(p-2)/2}\cdot s^2\,ds\right),
\end{aligned}$$

where we used the substitution $s^2 = 1-t^2$ in the second term. The function

$$u(p) = p^{3/2}\cdot\frac{2\pi}{p-2}\cdot(3/4)^{p-2}$$

is uniformly bounded for $p \in [4,\infty)$. To estimate the second term, we can use the inequality $1-s^2 \leqslant \exp(-s^2)$, which yields

$$p^{3/2}\int_0^{\sqrt{7}/4}(1-s^2)^{(p-2)/2}\cdot s^2\,ds \leqslant p^{3/2}\int_0^\infty \exp\left(-\frac{p-2}{2}s^2\right)s^2\,ds.$$

The last expression is also uniformly bounded for $p \in [4,\infty)$. This proves that

$$\frac{1}{a_j}\cdot\int_{\mathbb{R}} |\phi_{X_j}(x)|^{1/a_j^2}\,dx \leqslant C$$

for all j, where C is an absolute constant. Substituting this into (3.1.2) and using that $\sum_{j=1}^n a_j^2 = 1$ yields $f_s(0) \leqslant C'$ completing the proof of Step 1 modulo Lemma 3.1.4. □

Let us prove the lemma now.

Proof of Lemma 3.1.4. The first bound in the lemma follows from Markov's inequality

$$\lambda\{x : |\phi_X(x)| > t\} \leqslant \frac{\|\phi_X\|_2^2}{t^2}.$$

To estimate the L_2-norm, we apply the Plancherel identity:

$$\|\phi_X\|_2^2 = 2\pi \|f_X\|_2^2 \leqslant 2\pi \|f_X\|_\infty \cdot \|f_X\|_1 \leqslant 2\pi. \tag{3.1.5}$$

The estimate for $t \in [3/4, 1]$ will be based on a regularity argument going back to Halasz [15].

We will start with the symmetrization. Let X' denote an independent copy of X. Then

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

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

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

Denoting $s^2 = 1 - t^2$, we see that to prove that

$$\lambda\{x : \ |\phi_X(x)| > t\} \leqslant C\sqrt{1-t^2} \quad \text{for } t \in [3/4, 1],$$

it is enough to show that

$$\lambda\{\tau : \psi(\tau) \leqslant s^2\} \leqslant Cs, \quad \text{for } 0 < s \leqslant 1/2. \tag{3.1.6}$$

Observe that (3.1.6) holds for some fixed constant value of s. This follows from the identity $|\phi_X(\tau)|^2 = 1 - \psi(\tau)$ and inequality (3.1.5):

$$\lambda\big\{\tau : \psi(\tau) \leqslant \frac{1}{4}\big\} = \lambda\{\tau : |\phi_X(\tau)| \geqslant \sqrt{3/4}\} \leqslant 8\pi/3 \leqslant 9. \tag{3.1.7}$$

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

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

Therefore

$$\lambda\big\{\tau : \psi(\tau) \leqslant \frac{1}{4m^2}\big\} \leqslant \lambda\big\{\tau : \psi(m\tau) \leqslant \frac{1}{4}\big\} = \frac{1}{m}\,\lambda\big\{\tau : \psi(\tau) \leqslant \frac{1}{4}\big\} \leqslant \frac{9}{m}, \tag{3.1.8}$$

where in the last step we used (3.1.7). This establishes (3.1.6) for the discrete set of values $s = \frac{1}{2m}$, $m \in \mathbb{N}$. We can extend this to arbitrary $s > 0$ in a standard way, by applying (3.1.8) for $m \in \mathbb{N}$ such that $s \in (\frac{1}{4m}, \frac{1}{2m}]$. This proves (3.1.6) and completes the proof of Lemma 3.1.4. □

We now pass to the multidimensional case. As for one dimension, our strategy will depend on whether all vectors Pe_j are small or some Pe_j are large. In the first case, we proceed with a high-dimensional version of the argument from Step 1, where Hölder's inequality will be replaced by Brascamp-Lieb's inequality. In the second case, we will remove the large vectors Pe_j one by one, using the induction over the dimension.

Step 2. Let X be a random vector and P be a projection which satisfy the assumptions of Theorem 3.1.1. Assume that

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

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

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

Theorem 3.1.9 (Brascamp-Lieb [7], see also [3])**.** *Let $u_1, \dots, u_n \in \mathbb{R}^d$ be unit vectors and $c_1, \dots, c_n > 0$ be real numbers satisfying*

$$\sum_{i=1}^{n} c_j u_j u_j^\top = I_d.$$

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

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

A short and very elegant proof of the Brascamp-Lieb inequality based on the measure transportation ideas can be found in [5].

The singular value decomposition of P yields the existence of a $d \times n$ matrix R satisfying

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

It follows that $\|Px\|_2 = \|Rx\|_2$ for all $x \in \mathbb{R}^n$. This allows us to work with the matrix R instead of P. As before, replacing each X_j by KX_j, we may assume that $K = 1$. Finally, translating X if necessary we reduce the problem to bounding the density of RX at the origin.

As in the previous step, Fourier inversion formula associated with the Fourier transform in n dimensions yields that the density of RX at the origin can be reconstructed from its Fourier transform as

$$f_{RX}(0) = (2\pi)^{-d} \int_{\mathbb{R}^d} \phi_{RX}(x) \, dx \leqslant (2\pi)^{-d} \int_{\mathbb{R}^d} |\phi_{RX}(x)| \, dx, \tag{3.1.10}$$

where

$$\phi_{RX}(x) = \mathbb{E} \exp\big(i \langle x, RX \rangle \big) \tag{3.1.11}$$

is the characteristic function of RX. Therefore, to complete the proof, it suffices to bound the integral in the right hand side of (3.1.10) by C^d.

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

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

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

$$\sum_{j=1}^n a_j^2 u_j u_j^\top = I_d. \tag{3.1.12}$$

Moreover, we have $\langle x, RX\rangle = \sum_{i=1}^n a_j \langle x, u_j\rangle X_j$. Substituting this into (3.1.11) and using independence, we obtain

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

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

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

Recalling (3.1.12), we apply Brascamp-Lieb inequality for these functions and obtain

$$\begin{aligned}\int_{\mathbb{R}^d} |\phi_{RX}(x)|\,dx &= \int_{\mathbb{R}^d} \prod_{j=1}^n f_j\big(\langle x, u_j\rangle\big)^{a_j^2}\,dx \\ &\leqslant \prod_{j=1}^n \Big(\int_{\mathbb{R}} f_j(t)\,dt\Big)^{a_j^2} = \prod_{j=1}^n \Big(\int_{\mathbb{R}} \left|\phi_{X_j}(a_j t)\right|^{1/a_j^2}\,dt\Big)^{a_j^2}. \end{aligned} \tag{3.1.13}$$

We arrived at the same quantity as we encountered in one-dimensional argument in (3.1.2). Following that argument, which uses the assumption that all $a_j \leqslant 1/2$, we bound the product above by

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

Recalling that $a_j = \|Re_j\|_2$ and , we find that

$$\sum_{j=1}^n a_j^2 = \sum_{j=1}^n \|Re_j\|_2^2 = \mathrm{Tr}(RR^\top) = \mathrm{Tr}(I_d) = d.$$

Thus the right hand side of (3.1.13) is bounded by $(2C)^d$. The proof of Theorem 3.1.1 in the case where all $\|Pe_j\|_2$ are small is complete.

Step 3. Inductive argument.

We will prove Theorem 3.1.1 by induction on the rank of the projection. The case $\mathrm{rank}(P) = 1$ has been already established. We have also proved the Theorem when $\|Pe_j\|_2 < 1/2$ for all j. Assume that the theorem holds for all projections Q with $\mathrm{rank}(Q) = d-1$ and $\|Pe_1\|_2 \geqslant 1/2$.

The density function is not a convenient tool to run the inductive argument since the density of P_X does not usually splits into a product of densities related

to the individual coordinates. Let us consider the *Lévy concentration function* of a random vector which would replace the density in our argument.

Definition 3.1.14. Let $r > 0$. For a random vector $Y \in \mathbb{R}^n$, define its Lévy concentration function by

$$\mathcal{L}(Y, r) := \sup_{y \in \mathbb{R}^n} \mathbb{P}\left\{\|Y - y\|_2 \leqslant r\right\}.$$

Note that the condition that the density of Y is bounded is equivalent to

$$\mathcal{L}(Y, r\sqrt{n}) \leqslant (Cr)^n \quad \text{for any } r > 0.$$

One direction of this equivalence follows from the integration of the density function over the ball of radius $t\sqrt{n}$ with the center at any $y \in \mathbb{R}^n$; another one from the Lebesgue differentiation theorem.

In terms of the Lévy concentration function, the statement of the theorem is equivalent to the claim that for for any $y \in P\mathbb{R}^n$ and any $t > 0$,

$$\mathbb{P}\left\{\|PX - y\|_2 \leqslant t\sqrt{d}\right\} \leqslant (Mt)^d \tag{3.1.15}$$

for some absolute constant M and with $d = \operatorname{rank}(P)$. The induction assumption then reads: for all projections Q of rank $d-1$, $z \in Q\mathbb{R}^n$, and $t > 0$, we have

$$\mathbb{P}\left\{\|QX - z\|_2 \leqslant t\sqrt{d-1}\right\} \leqslant (Mt)^{d-1}. \tag{3.1.16}$$

Comparison of (3.1.16) and (3.1.15) immediately shows the difficulties we are facing: the change from $d-1$ to d in the left hand side of these inequalities indicates that we have to work accurately to preserve the constant M while deriving (3.1.15) from (3.1.16). This is achieved by a delicate tensorization argument. By considering an appropriate shift of X, we can assume without loss of generality that $y = 0$. Let us formulate the induction step as a separate proposition.

Proposition 3.1.17 (Removal of large Pe_i)**.** *Let X be a random vector satisfying the assumptions of Theorem 3.1.1 with $K = 1$, and let P be an orthogonal projection in $\mathbb{R}^n$ onto a d-dimensional subspace. Assume that*

$$\|Pe_1\|_2 \geqslant 1/2.$$

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

$$\ker(Q) = \operatorname{span}\{\ker(P), Pe_1\}.$$

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

$$\mathbb{P}\left\{\|QX\|_2 \leqslant t\sqrt{d-1}\right\} \leqslant (Mt)^{d-1} \quad \textit{for all } t \geqslant 0, \tag{3.1.18}$$

then

$$\mathbb{P}\left\{\|PX\|_2 \leqslant t\sqrt{d}\right\} \leqslant (Mt)^d \quad \textit{for all } t \geqslant 0.$$

Proof. Let us record a few basic properties of Q. It is straightforward to see that

$$P - Q \text{ is the orthogonal projection onto } \operatorname{span}(Pe_1). \tag{3.1.19}$$

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

$$Qe_1 = 0. \tag{3.1.20}$$

It follows from (3.1.19) that P has the form

$$Px = \Big(\sum_{j=1}^{n} a_j x_j\Big) Pe_1 + Qx \quad \text{for } x = (x_1, \dots, x_n) \in \mathbb{R}^n, \tag{3.1.21}$$

where a_j are fixed numbers (independent of x). Substituting $x = e_1$, we obtain using (3.1.20) that $Pe_1 = a_1 Pe_1 + Qe_1 = a_1 Pe_1$. Thus

$$a_1 = 1. \tag{3.1.22}$$

Furthermore, we note that

$$Qx \text{ does not depend on } x_1 \tag{3.1.23}$$

since $Qx = Q(\sum_{i=1}^n x_j e_j) = \sum_{i=1}^n x_j Qe_j$ and $Qe_1 = 0$ by (3.1.20). Finally, since Pe_1 is orthogonal to the image of Q, the two vectors in the right side of (3.1.21) are orthogonal. Thus

$$\|Px\|_2^2 = \Big(\sum_{j=1}^{n} a_j x_j\Big)^2 \|Pe_1\|_2^2 + \|Qx\|_2^2. \tag{3.1.24}$$

Now let us estimate $\|PX\|_2$ for a random vector X. We express $\|PX\|_2^2$ using (3.1.24) and (3.1.22) as

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

Since by (3.1.23) Z_2 is determined by $X_2, \dots, X_n$ (and is independent of X_1), and $\|Pe_i\|_2 \geqslant 1/2$ by a hypothesis of the proposition, we have

$$\begin{aligned}\mathbb{P}\left\{Z_1 \leqslant t \mid Z_2\right\} &\leqslant \max_{X_2,\dots,X_n} \mathbb{P}\left\{\Big|X_1 + \sum_{j=2}^{n} a_j X_j\Big| \leqslant t/\|Pe_1\|_2 \;\Big|\; X_2, \dots, X_n\right\} \\ &\leqslant \max_{u\in\mathbb{R}} \mathbb{P}\left\{|X_1 - u| \leqslant 2t\right\} \leqslant 2t.\end{aligned}$$

The proof of the inductive step thus reduces to a two-dimensional statement, which we formulate as a separate lemma.

Lemma 3.1.25 (Tensorization). *Let $Z_1, Z_2 \geqslant 0$ be random variables and $K_1, K_2 \geqslant 0$, $d > 1$ be real numbers. Assume that*

(1) $\mathbb{P}\left\{Z_1 \leqslant t \mid Z_2\right\} \leqslant 2t$ *almost surely in Z_2 for all $t \geqslant 0$;*
(2) $\mathbb{P}\left\{Z_2 \leqslant t\sqrt{d-1}\right\} \leqslant (Mt)^{d-1}$ *for all $t \geqslant 0$.*

for a sufficiently large absolute constant M. Then

$$\mathbb{P}\left\{\sqrt{Z_1^2 + Z_2^2} \leqslant t\sqrt{d}\right\} \leqslant (Mt)^d \quad \text{for all } t \geqslant 0.$$

The proof of the tensorization lemma requires an accurate though straightforward calculation. We write

$$\mathbb{P}\left\{\sqrt{Z_1^2+Z_2^2}\leqslant t\sqrt{d}\right\}=\int_0^{t^2d}\mathbb{P}\left\{Z_1\leqslant(t^2d-x)^{1/2}\mid Z_2^2=x\right\}\,dF_2(x)$$

where $F_2(x)=\mathbb{P}\left\{Z_2^2\leqslant x\right\}$ is the cumulative distribution function of Z_2^2. Using hypothesis (1) of the lemma, we can bound the right hand side by

$$2\int_0^{t^2d}(t^2d-x)^{1/2}\,dF_2(x)=\int_0^{t^2d}F_2(x)(t^2d-x)^{-1/2}\,dx,$$

where the last equation follows by integration by parts. Hypothesis (2) of the lemma states that

$$F_2(x)\leqslant M^{d-1}\left(\frac{x}{d-1}\right)^{(d-1)/2}.$$

Substituting this into the equality above and estimating the resulting integral explicitly, we obtain

$$\begin{aligned}\mathbb{P}\left\{\sqrt{Z_1^2+Z_2^2}\leqslant t\sqrt{d}\right\}&\leqslant\int_0^{t^2d}M^{d-1}\left(\frac{x}{d-1}\right)^{(d-1)/2}(t^2d-x)^{-1/2}\,dx\\&=t^d\cdot M^{d-1}\frac{d^{d/2}}{(d-1)^{(d-1)/2}}\int_0^1 y^{(d-1)/2}(1-y)^{-1/2}\,dy\leqslant t^d\cdot M^{d-1}\cdot C,\end{aligned}$$

where the last inequality follows with an absolute constant C from the known asymptotic of the beta-function. Alternatively, notice that

$$\frac{d^{d/2}}{(d-1)^{(d-1)/2}}\leqslant\sqrt{ed},$$

and

$$\begin{aligned}\int_0^1 y^{(d-1)/2}(1-y)^{-1/2}\,dy&\leqslant\int_0^{1-1/d}y^{(d-1)/2}\sqrt{d}\,dy+\int_{1-1/d}^1(1-y)^{-1/2}\,dy\\&\leqslant\frac{2}{\sqrt{ed}}+\frac{1}{2\sqrt{d}}.\end{aligned}$$

This completes the proof of the lemma if we assume that $M\geqslant C$. □

3.2. Small ball probability for the image of a vector. Let us derive an application of Theorem 3.1.1 which will be important for us in the proof of the no-gaps delocalization theorem. We will prove a small ball probability estimate for the image of a fixed vector under the action of a random matrix with independent entries of bounded density.

Lemma 3.2.1 (Lower bound for a fixed vector)**.** *Let G be an $l\times m$ matrix with independent complex random entries. Assume that the real parts of the entries have uniformly bounded densities, and the imaginary parts are fixed. For each $x\in S_{\mathbb{C}}^{m-1}$ and $\theta>0$, we have*

$$\mathbb{P}\left\{\|Gx\|_2\leqslant\theta\sqrt{l}\right\}\leqslant(C_0\theta)^l.$$

To prove this lemma, let us derive the small ball probability bound for a fixed coordinate of Gx first.

Lemma 3.2.2 (Lower bound for a fixed row and vector)**.** *Let G_j denote the j-th row of G. Then for each j, $z \in S_{\mathbb{C}}^{n-1}$, and $\theta \geqslant 0$, we have*

$$\mathbb{P}\left\{ |\langle G_j, z\rangle| \leqslant \theta \right\} \leqslant C_0 K\theta. \tag{3.2.3}$$

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 $\|x\|_2 \geqslant 1/2$. Dropping the imaginary part, we obtain

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

The imaginary part Y is fixed. Thus

$$\mathbb{P}\left\{ |\langle Z, z\rangle| \leqslant \theta \right\} \leqslant \mathcal{L}(\langle X, x\rangle, \theta). \tag{3.2.4}$$

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

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

Here X_k are independent random variables with densities bounded by K. Recalling that $\sum_{k=1}^{m} x_k^2 \geqslant 1/2$, we can apply Theorem 3.1.1 for a rank one projection. It yields

$$\mathcal{L}(\langle X, x\rangle, \theta) \leqslant CK\theta. \tag{3.2.5}$$

Substituting this into (3.2.4) completes the proof of Lemma 3.2.2. □

Now we can complete the proof of Lemma 3.2.1 We can represent $\|Gx\|_2^2$ as a sum of independent non-negative random variables $\sum_{j=1}^{l} |\langle G_j, x\rangle|^2$. Each of the terms $\langle G_j, x\rangle$ satisfies (3.2.3). Then the conclusion follows from the following Tensorization Lemma applied to $V_j = |\langle G_j, x\rangle|$.

Lemma 3.2.6. *Let $V_1, \dots, V_l$ be independent non-negative random variables satisfying*

$$\mathbb{P}\left\{ V_j < t \right\} \leqslant Ct$$

for any $t > 0$. Then

$$\mathbb{P}\left\{ \sum_{j=1}^{l} V_j^2 < t^2 l \right\} \leqslant (ct)^l.$$

Proof. Since the random variables $V_1^2, \dots, V_l^2$ are independent as well, the Laplace transform becomes a method of choice in handling this probability. By Markov's

inequality, we have

$$\mathbb{P}\left\{\sum_{j=1}^{l} V_j^2 < t^2 l\right\} = \mathbb{P}\left\{l - \frac{1}{t^2}\sum_{j=1}^{l} V_j^2 > 0\right\} \leqslant \mathbb{E}\exp\left(l - \frac{1}{t^2}\sum_{j=1}^{l} V_j^2\right)$$
$$= e^l \prod_{j=1}^{l} \mathbb{E}\exp(-V_j^2/t^2).$$

To bound the expectations in the right hand side, we use the Fubini theorem:

$$\mathbb{E}\exp(-V_j^2/t^2) = \int_0^\infty 2xe^{-x^2}\mathbb{P}\left\{V_j < tx\right\}\,dx \leqslant Ct,$$

where the last inequality follows from the assumption on the small ball probability of V_j. Combining the previous two inequalities completes the proof. □

4. No-gaps delocalization for matrices with absolutely continuous entries.

In this section, we prove Theorem 1.0.4. To this end, we combine all the tools we discussed above: the bound on the density of a projection of a random vector obtained in Theorem 3.1.1, the ε-net argument, and the small ball probability bound of Lemma 3.2.1.

4.1. Decomposition of the matrix Let us recall that in Proposition 2.1.2 we have reduced the claim of delocalization Theorem 1.0.4 to the following quantitative invertibility problem:

- Let A be an $n \times n$ matrix satisfying Assumptions 1.0.1 and 1.0.3. Let $\varepsilon > 0$, $t > 0$, $M > 1$, and let $\lambda \in \mathbb{C}$, $|\lambda| \leqslant M\sqrt{n}$. Let $I \subset [n]$ be a fixed set of cardinality $|I| = \varepsilon n$. Estimate
$$p_0 := \mathbb{P}\left(s_{\min}((A-\lambda)_{I^c}) < t\sqrt{n} \text{ and } \|A\| \leqslant M\sqrt{n}\right).$$

Since the set I is fixed, we can assume without loss of generality that I consists of the last εn coordinates.

Let us decompose $(A-\lambda)_{I^c}$ as follows:

$$(A-\lambda)_{I^c} = \begin{bmatrix} B \\ G \end{bmatrix}, \tag{4.1.1}$$

where B and G are rectangular matrices of respective sizes $(1-\varepsilon/2)n \times (1-\varepsilon)n$ and $(\varepsilon/2)n \times (1-\varepsilon)n$. By Assumption 1.0.1, the random matrices B and G are independent, and moreover all entries of G are independent. At the same time, the matrix B is still rectangular, and the ratio of its number of rows and columns is similar to that of the matrix $(A-\lambda)_{I^c}$. This would allow us to prove a weaker statement for the matrix B. Namely, instead of bounding the smallest singular value, which is the minimum of $\|Bx\|_2$ over all unit vectors x, we will obtain the desired lower bound for all vectors which are far away from a certain low-dimensional subspace depending on B. The independence of B and G would

make it possible to condition on B fixing this subspace and apply Lemma 2.2.1 to the matrix G restricted to this subspace to ensure that the matrix $(A-\lambda)_{I^c}$ is well invertible on this space as well.

Following this road map, 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 $0.01\varepsilon n$ singular values are bounded below by $\Omega(\sqrt{\varepsilon n})$. It follows that $\|Bx\|_2$ is nicely bounded below when x restricted to a subspace of codimension $0.01\varepsilon n$. (This subspace is formed by the corresponding singular vectors.) Next, we condition on B and we use G to control the remaining $0.01\varepsilon n$ dimensions. Therefore, either $\|Bx\|_2$ or $\|Gx\|_2$ is nicely bounded below on the entire space, and thus $\|(A-\lambda)_{I^c}x\|_2$ is nicely bounded below on the entire space as well.

We will now pass to the implementation of this plan. To simplify the notation, *assume that the maximal density of the entries is bounded by 1*.
The general case can be reduced to this by scaling the entries.

4.2. The negative second moment identity Let $k \geqslant m$. The Hilbert-Schmidt norm of a $k \times m$ matrix V is just the Euclidean norm of the km-dimensional vector consisting of its entries. Like the operator norm, the Hilbert-Schmidt norm is invariant under unitary or orthogonal transformations of the matrix V. This allows to rewrite it in two ways:

$$\|V\|_{HS}^2 = \sum_{j=1}^{m} \|V_j\|_2^2 = \sum_{j=1}^{m} s_j(V_j)^2,$$

where $V_1, \dots, V_m$ are the columns of V, and $s_1(V) \geqslant s_2(V) \geqslant \dots \geqslant s_m(V) \geqslant 0$ are its singular values. Applying this observation to the inverse of the linear operator defined by V considered as an operator from $V\mathbb{C}^m$ to $\mathbb{C}^m$, we obtain *the negative second moment identity*, see [26]:

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

Here B_j denote the columns of B, and $H_j = \operatorname{span}(B_l)_{l\neq j}$.

Returning to the matrix B, denote for shortness $m = (1-\varepsilon)n$ and $\varepsilon' = \frac{\varepsilon}{2(1-\varepsilon)}$. In this notation, B is a $(1+\varepsilon')m \times m$ matrix. To bound the sum above, we have to establish a lower bound on the distance between the random vector $B_j \in \mathbb{C}^{(1+\varepsilon')m}$ and random subspace $H_j \subseteq \mathbb{C}^{(1+\varepsilon')m}$ of complex dimension $m-1$.

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 1.0.1 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 $B_j' \in \mathbb{C}^{(1+\varepsilon')m-1}$ 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 1.0.1. Second,

$$\operatorname{dist}(B_j, H_j) \geqslant \operatorname{dist}(B_j', H_j'), \tag{4.2.1}$$

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

Summarizing, we have

$$\sum_{j=1}^{m} s_j(B)^{-2} \leqslant \sum_{j=1}^{m} \operatorname{dist}(B_j', H_j')^{-2}. \tag{4.2.2}$$

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 $(H_j')^{\perp}$:

$$\operatorname{dist}(B_j', H_j') = \|P_{E_j} B_j'\|_2, \quad \text{where} \quad E_j = (H_j')^{\perp}. \tag{4.2.3}$$

Recall that $B_j' \in \mathbb{C}^{(1+\varepsilon')m-1}$ is a random vector with independent entries whose real parts have densities bounded by 1 (by Assumptions 1.0.1 and 1.0.3); and H_j' is an independent subspace of $\mathbb{C}^{(1+\varepsilon')m-1}$ of complex dimension $m-1$. This puts us on a familiar ground as we have already proved Theorem 3.1.1. Now, the main strength of this result becomes clear. The bound of Theorem 3.1.1 is uniform over the possible subspaces E_j meaning that we do not need any information about the specific position of this subspace in $\mathbb{C}^{(1+\varepsilon')m-1}$. This is a major source of simplifications in the proof of Theorem 1.0.4 compare to Theorem 1.0.7. Under Assumption 1.0.5, a bound on the small ball probability for $\|P_{E_j} B_j'\|_2$ depends on the arithmetic structure of the vectors contained in the space E_j. Identifying subspaces of $\mathbb{C}^{(1+\varepsilon')m-1}$ containing vectors having exceptional arithmetic structure and showing that, with high probability, the space E_j avoids such positions, takes a lot of effort. Fortunately, under Assumption 1.0.3, this problem does not arise thanks to the uniformity mentioned above.

Transferring the problem from $\mathbb{C}$ to $\mathbb{R}$ If the real and the imaginary part of each entry of A are random variables of bounded density, one can apply Theorem 3.1.1 directly. However, this case does not cover many matrices satisfying Assumption 1.0.1, most importantly, the matrices with real entries and complex spectrum.

The general case, when only the real parts of the vector $B_j' \in \mathbb{C}^{(1+\varepsilon')m-1}$ are random, requires an additional symmetrization step. Indeed, if we transfer the problem from the complex vector space to a real one of the double dimension, only a half of the coordinates will be random. Such vector would not be absolutely continuous, so we cannot operate in terms of the densities. As in the previous section, the *Lévy concentration function* of a random vector would replace the density in our argument.

Let us formally transfer the problem from the complex to the real field. To this end, we define the operation $z \mapsto \operatorname{Real}(z)$ that makes complex vectors real in the

obvious way: for $z = x + iy \in \mathbb{C}^N$, set $\mathrm{Real}(z) := \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] \in \mathbb{R}^{2N}$. Similarly, to make a complex subspace $E \subset \mathbb{C}^N$ real, we set $\mathrm{Real}(E) := \{\mathrm{Real}(z) : z \in E\} \subset \mathbb{R}^{2N}$. Note that this operation doubles the dimension of E.

We begin by recording two properties of this operation that follow straight from this definition.

Lemma 4.2.4. *(Elementary properties of operation $x \mapsto \mathrm{Real}(x)$)*

(1) *For a complex subspace E and a vector z, one has*

$$\mathrm{Real}(P_E z) = P_{\mathrm{Real}(E)} \mathrm{Real}(z).$$

(2) *For a complex-valued random vector X and $r \geqslant 0$, one has*

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

The next symmetrization lemma allows randomizing all coordinates.

Lemma 4.2.5 (Randomizing all coordinates)**.** *Let $Z = X + iY \in \mathbb{C}^N$ be a random vector whose imaginary part $Y \in \mathbb{R}^N$ is fixed and then set $\widehat{Z} = \left[\begin{smallmatrix} X_1 \\ X_2 \end{smallmatrix}\right] \in \mathbb{R}^{2N}$ with X_1 and X_2 independent copies of X. Let E be a subspace of $\mathbb{C}^N$. Then*

$$\mathcal{L}(P_E Z, r) \leqslant \left(\mathcal{L}(P_{\mathrm{Real}(E)} \widehat{Z}, 2r)\right)^{1/2}, \quad r \geqslant 0.$$

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

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

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

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

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:

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

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

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

Further, using part 2 and then part 1 of Lemma 4.2.4, we see that

$$\begin{aligned} \mathcal{L}(P_E(X_1 + iX_2), 2r) &= \mathcal{L}(P_{\mathrm{Real}(E)}(\mathrm{Real}(X_1 + iX_2)), 2r) \\ &= \mathcal{L}(P_{\mathrm{Real}(E)} \widehat{Z}, 2r). \end{aligned}$$

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

Bounding the distances below We are ready to control the distances appearing in (4.2.3).

Lemma 4.2.6 (Distance between random vectors and subspaces)**.** *For every* $j \in [n]$ *and* $\tau > 0$*, we have*

$$\mathbb{P}\left\{\operatorname{dist}(B_j', H_j') < \tau\sqrt{\varepsilon' m}\right\} \leqslant (C\tau)^{\varepsilon' m}. \tag{4.2.7}$$

Proof. Representing these distances via the projections of B_j' onto the subspaces $E_j = (H_j')^\perp$ as in (4.2.3), and using the definition of the Lévy concentration function, we have

$$p_j := \mathbb{P}\left\{\operatorname{dist}(B_j', H_j') < \tau\sqrt{\varepsilon' m}\right\} \leqslant \mathcal{L}(P_{E_j} B_j', \tau\sqrt{\varepsilon' m}).$$

Recall that B_j' and E_j are independent, and let us condition on E_j. Lemma 4.2.5 implies that

$$p_j \leqslant \left(\mathcal{L}(P_{\mathrm{Real}(E_j)}\widehat{Z}, 2\tau\sqrt{\varepsilon' m})\right)^{1/2}$$

where $\widehat{Z}$ is a random vector with independent coordinates that have densities bounded by 1.

The space H_j' has codimension $\varepsilon' m$; thus E_j has dimension $\varepsilon' m$ and $\mathrm{Real}(E_j)$ has dimension $2\varepsilon' m$. By Theorem 3.1.1, the density of $P_{\mathrm{Real}(E_j)}\widehat{Z}$ is bounded by $C^{2\varepsilon' m}$. Integrating the density over a ball of radius $2\tau\sqrt{\varepsilon' m}$ in the subspace $\mathrm{Real}(E_j)$ which has volume $(C\tau)^{2\varepsilon' m}$, we conclude that

$$\mathcal{L}(P_{\mathrm{Real}(E_j)}\widehat{Z}, 2\tau\sqrt{\varepsilon n}) \leqslant (C\tau)^{2\varepsilon' m}.$$

It follows that

$$p_j \leqslant (C\tau)^{\varepsilon' m},$$

as claimed. The proof of Lemma 4.2.6 is complete. □

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

Using the second moment inequality Denote $p = \varepsilon' m/4$, and let

$$Y_j = \varepsilon' m \cdot \operatorname{dist}^{-2}(B_j', H_j') \quad \text{for } j \in [m].$$

By Lemma 4.2.6, for any $s > 0$,

$$\mathbb{P}\left\{Y_j > s\right\} \leqslant \left(\frac{C}{s}\right)^{2p}.$$

Using Fubini's theorem, we conclude that

$$\mathbb{E}Y_j^p \leqslant 1 + p\int_1^\infty s^{p-1} \cdot \mathbb{P}\left(Y_j > s\right) ds \leqslant 1 + \bar{C}^p,$$

so $\|Y_j\|_p = \left(\mathbb{E}Y_j^p\right)^{1/p} \leqslant C$. Here, once again the assumption of the bounded density of the entries simplifies the proof. For a general distribution of entries, the event $\operatorname{dist}(B_j', H_j') = 0$ may have a positive probability, so $\|Y_j\|_p$ may be infinite.

The bound on $\|Y_j\|_p$ yields $\left\|\sum_{j=1}^m Y_j\right\|_p \leqslant Cm$. By Markov's inequality, we get

$$\mathbb{P}\left(\sum_{j=1}^m \operatorname{dist}^{-2}(B_j', H_j') \geqslant \frac{1}{\varepsilon' t}\right) = \mathbb{P}\left(\sum_{j=1}^m Y_j \geqslant \frac{m}{t}\right) \leqslant \frac{\mathbb{E}(\sum_{j=1}^m Y_j)^p}{(m/t)^p} \leqslant (Ct)^p$$

for any $t > 0$.

This estimate for $t = \tau^2$ combined with inequality (4.2.2) shows that the event

$$\mathcal{E}_1 := \left\{\sum_{i=1}^m s_i(B)^{-2} \leqslant \frac{1}{\tau^2 \varepsilon'}\right\} \tag{4.3.1}$$

is likely: $\mathbb{P}\left((\mathcal{E}_1)^c\right) \leqslant (C'\tau)^{\varepsilon' m/2}$.

A large subspace E^+ on which B is bounded below Fix a parameter $\tau > 0$ for now, and assume that the event (4.3.1) occurs. By Markov's inequality, for any $\delta > 0$ we have

$$\left|\{i : s_i(B) \leqslant \delta\sqrt{m}\}\right| = \left|\{i : s_i(B)^{-2} \geqslant \frac{1}{\delta^2 m}\}\right| \leqslant \frac{\delta^2 m}{\tau^2 \varepsilon'}.$$

Setting $\delta = \tau\varepsilon'/10$, we have

$$\left|\{i : s_i(B) \leqslant \frac{\tau\varepsilon'}{10}\sqrt{m}\}\right| \leqslant \frac{\varepsilon' m}{100}. \tag{4.3.2}$$

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

$$E^- = \operatorname{span}\{v_i(B) : s_i(B) \leqslant \frac{\tau\varepsilon'}{10}\sqrt{m}\},$$
$$E^+ = \operatorname{span}\{v_i(B) : s_i(B) > \frac{\tau\varepsilon'}{10}\sqrt{m}\}.$$

Inequality (4.3.2) means that $\dim_{\mathbb{C}}(E^-) \leqslant \frac{\varepsilon' m}{100}$.

Let us summarize. Recall that $\varepsilon' m = \varepsilon n/2$ and set $\tau = (\varepsilon s)^2$ for some $s \in (0,1)$. We proved that the event

$$\mathcal{D}_{E^-} := \left\{\dim(E^-) \leqslant \frac{\varepsilon' m}{100}\right\}$$

satisfies

$$\mathbb{P}\left((\mathcal{D}_{E^-})^c\right) \leqslant (C_2\tau)^{\varepsilon' m} = (C_3\varepsilon s)^{\varepsilon n}, \tag{4.3.3}$$

so E^- is likely to be a small subspace and E^+ a large subspace. The choice of τ was made to create the factor $\varepsilon^{\varepsilon n}$ in the probability bound above ensuring that we can suppress the factor $\binom{n}{\varepsilon n}$ arising from the union bound. Moreover, by definition, B is nicely bounded below on $S_{E^+} = S^{n-1} \cap E^+$:

$$\inf_{x \in S_{E^+}} \|Bx\|_2 \geqslant \frac{\tau\varepsilon'}{10}\sqrt{m} \geqslant \frac{s^2\varepsilon^3}{80}\sqrt{n}. \tag{4.3.4}$$

4.4. G is bounded below on the small complementary subspace E_- The previous argument allowed us to handle the subspace E_+ whose dimension is only slightly lower than m. Yet, it provided no information about the behavior of the infimum of $\|Bx\|_2$ over the unit vectors from the complementary subspace E_-. To get such a lower bound, we will use the submatrix G we have put aside. Recall that although the space E_- is random, it depends only on B, and thus is independent of G. Conditioning on the matrix B, we can regard this space as fixed. Our task therefore, is to establish a lower bound on $\|Gx\|_2$ over the unit vectors from E_-. To this end, we can use the Lemma 2.2.1. However, this lemma establishes the desired bound with probability at least $1-\exp(-c'\varepsilon' m)$. This probability is insufficient for our purposes (remember, the probability for a fixed set $I \subset [n]$ is multiplied by $\binom{n}{\varepsilon n} \sim (e/\varepsilon)^{\varepsilon n}$), but is easy to improve in case of the bounded densities. Replacing the small ball probability estimate for a fixed vector used in the proof of Lemma 2.2.1 with Lemma 3.2.1, we derive the following lemma.

Lemma 4.4.1 (Lower bound on a subspace)**.** *Let $M \geqslant 1$ and $\mu \in (0,1)$. Let E be a fixed subspace of $\mathbb{C}^m$ of dimension at most $\varepsilon' m/100$. Then, for every $\rho > 0$, we have*

$$\mathbb{P}\left\{ \inf_{x \in S_E} \|Gx\|_2 < \rho\sqrt{\varepsilon' m} \text{ and } \mathcal{B}_{G,M} \right\} \leqslant \left(\frac{CM\rho^{0.98}}{\varepsilon'^{0.01}} \right)^{\varepsilon' m}. \tag{4.4.2}$$

The proof of this lemma follows the same lines as that of Lemma 2.2.1 and is left to a reader.

Lemma 4.4.1 provides the desired bound for the space E_-. Recall that $m = (1-\varepsilon)n$ and $\varepsilon' = \varepsilon/2(1-\varepsilon)$. Namely, if the events $\mathcal{B}_{G,M}$ and $\mathcal{D}_{E_-}$ occur, then the event

$$\mathcal{L}_{E_-} := \left\{ \inf_{x \in S^{m-1} \cap E_-} \|Gx\|_2 \geqslant \rho\sqrt{\varepsilon' m} \right\}$$

holds with probability at least

$$1 - \left(\frac{CM\rho^{0.98}}{\varepsilon'^{0.01}} \right)^{\varepsilon' m}.$$

This is already sufficient since choosing a sufficiently small ρ, say $\rho = (s\varepsilon')^3$ with any $s \in (0,1)$, we see that

$$\mathbb{P}\left(\mathcal{L}_{E_-}^c\right) \leqslant (CMs^3\varepsilon^{2.9})^{\varepsilon n/2},$$

so again we can suppress the factor $\binom{n}{\varepsilon n}$ arising from the union bound.

4.5. Extending invertibility from subspaces to the whole space. Assume that the events $\mathcal{D}_{E_-}$ and $\mathcal{L}_{E_-}$ occur. We know that if $\mathcal{B}_{A,M}$ occurs, then this is likely:

$$\mathbb{P}\left(\mathcal{B}_{A,M} \cap \mathcal{D}_{E_-} \cap \mathcal{L}_{E_-}\right) \geqslant \mathbb{P}\left(\mathcal{B}_{A,M}\right) - (Cs)^{\varepsilon n}.$$

Under this assumption, we have uniform lower bounds on $\|Ax\|_2$ on the unit spheres of both E_+ and E_-. The extension of these bounds to the whole unit sphere of $\mathbb{C}^m$ is now deterministic. It relies on the following lemma from linear algebra.

Lemma 4.5.1 (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 an orthogonal decomposition $\mathbb{C}^n = E^- \oplus E^+$ where E^- and E^+ are eigenspaces[1] *of B^*B. Denote*

$$\begin{aligned} s_A &= s_{\min}(A), \\ s_B &= s_{\min}(B|_{E^+}) = \min_{x \in S^{n-1} \cap E^+} \|Bx\|_2, \\ s_G &= s_{\min}(G|_{E^-}) = \min_{x \in S^{n-1} \cap E^-} \|Gx\|_2. \end{aligned}$$

Then

$$s_A \geqslant \frac{s_B s_G}{4\|A\|}. \tag{4.5.2}$$

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 $\theta \in (0, 1/2)$ and consider two cases.

Case 1: $\|x^+\|_2 \geqslant \theta$. Since Bx^+ and Bx^- are orthogonal,

$$\|Ax\|_2 \geqslant \|Bx\|_2 \geqslant \|Bx^+\|_2 \geqslant s_B \cdot \theta.$$

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

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

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

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

Optimizing the parameter θ, we conclude that

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

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

Combining Lemma 4.5.1 with the bounds (4.3.4) and (4.4.2), we complete the proof of Proposition 2.1.2, and thus, the no-gaps delocalization Theorem 1.0.4.

5. Applications of the no-gaps delocalization

5.1. Erdős-Rényi graphs and their adjacency matrices In this section we consider two applications of the no-gaps delocalization to the spectral properties of

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

the Erdős-Rényi random graphs. Let $p \in (0,1)$. Consider a graph $G = (V, E)$ with n vertices such that any pair of vertices is connected by an edge with probability p, and these events are independent for different edges. This model of a random graph is called an Erdős-Rènyi or $G(n,p)$ graph. Let A_G be the adjacency matrix of a graph G, i.e., the matrix of zeros and ones with 1 appearing on the spot (i,j) whenever the vertices i and j are connected. We will need several standard facts about the Erdős-Rényi graphs listed in the following proposition [10].

Proposition 5.1.1. *Let $p \geqslant C_0 \frac{\log n}{n}$ for some $C_0 > 1$. Let $G(V,E)$ be a $G(n,p)$ graph. Then G has the following properties with probability $1 - o(1)$.*

(1) *Let $R \subset V$ be an independent set, i.e., no two vertices from R are connected by an edge. Then*
$$|R| \leqslant C\frac{\log n}{p}.$$

(2) *Let $P, Q \subset V$ be disjoint sets of vertices with*
$$|P|, |Q| \geqslant C\frac{\log n}{p}.$$
Then there is an edge connecting a vertex from P and a vertex from Q.

(3) *The degree of any vertex $v \in V$ is close to its expectation:*
$$np - \log n \cdot \sqrt{np} \leqslant d_v \leqslant np + \log n \cdot \sqrt{np}$$

(4) *Define the normalized adjacency matrix of G to be $\hat{A} := D_G^{-1/2} A_G D_G^{-1/2}$ where D_G is the diagonal matrix $D_G = \operatorname{diag}(d_v,\ v \in V)$ and et $\hat{\lambda}_1 \geqslant, \dots, \geqslant \hat{\lambda}_n$ be eigenvalues of $\hat{A}$. Then*
$$\hat{\lambda}_1 = 1, \quad \text{and} \quad |\hat{\lambda}_j| \leqslant \frac{C}{\sqrt{np}} \text{ for } j \geqslant 1.$$

(5) *For every subset of vertices $J \subset V$, let Non-edges(J) be the set of all pairs of vertices $v, w \in J$ which are not connected by an edge. Then*
$$(1-p)\binom{|J|}{2} - n^{3/2} \leqslant |\text{Non-edges}(J)| \leqslant (1-p)\binom{|J|}{2} + n^{3/2}.$$

We leave the proof of these properties to a reader.

Considering the vector of all ones, we realize that $\|A_G\| = \Omega(np)$ with high probability. Hence, when p is fixed, and $n \to \infty$, this makes the event $\mathcal{B}_{A_G, M}$ unlikely. However, Remark 1.0.9 shows that we can replace this event by the event $\mathcal{B}_{A_G - p\mathbf{1}_n, M}$ which holds with probability close to 1. Indeed,
$$A_G - p\mathbf{1}_n = B - \Delta,$$
where B is a symmetric random matrix with centered Bernoulli(p) entries which are independent on and above the diagonal, and Δ is the diagonal matrix with i.i.d. Bernoulli(p) entries. Here, $\|\Delta\| \leqslant 1$ and $\|B\| \leqslant C\sqrt{np}$ with probability close to 1, by a simple ε-net argument.

This decomposition is reflected in the structure of the spectrum of A_G. Let us arrange the eigenvalues of A_G in the decreasing order: $\lambda_1(G) \geqslant \dots \geqslant \lambda_n(G)$. Then with high probability, $\lambda_1(G) = \Omega(np)$ and $|\lambda_j(G)| = O(\sqrt{np})$, where the last

equality follows from $\|A_G - p\mathbf{1}_n\| = O(\sqrt{np})$ and the interlacing property of the eigenvalues.

Remark 1.0.9 shows that no-gaps delocalization can be extended to the matrix A_G as well. We will use this result in combination with the ℓ_∞ delocalization which was established for the $G(n,p)$ graphs by Erdős et. al. [11]. They proved that with probability at least $1 - \exp(-c\log^2 n)$, any unit eigenvector x of A_G satisfies

$$\|x\|_\infty \leqslant \frac{\log^C n}{\sqrt{n}}. \tag{5.1.2}$$

5.2. Nodal domains of the eigenvectors of the adjacency matrix Let f be an eigenfunction of a self-adjoint linear operator. Define the (strong) nodal domains of f as connected components of the sets where f is positive or negative. Nodal domains of the Laplacian on a compact smooth manifold is a classical object in analysis. If the eigenvalues are arranged in the increasing order, the number of nodal domains of the eigenfunction corresponding to the k-th eigenvalue does not exceed k and tends to infinity as $k \to \infty$.

If we consider a finite-dimensional setup, the eigenfunctions of self-adjoint linear operators are replaced by the eigenvectors of symmetric matrices. In 2008, Dekel, Lee, and Linial [9] discovered that the nodal domains of the adjacency matrices of $G(n,p)$ graphs behave strikingly different from the eigenfunctions of the Laplacian on a manifold. Namely, they proved that with high probability, the number of nodal domains of any non-first eigenvector of a $G(n,p)$ graph is bounded by a constant depending only on p. Later, their result was improved by Arora and Bhaskara [1], who showed that with high probability, the number of nodal domains is 2 for all non-first eigenvectors. Also, Nguyen, Tao, and Vu [19] showed that the eigenvector of a $G(n,p)$ graph cannot have zero coordinates with probability close to 1. These two results in combination mean that for each non-first eigenvector, the set of vertices of a $G(n,p)$ graph splits into the set of positive and negative coordinates both of which are connected.

Let us derive Dekel-Lee-Linial-Arora-Bhaskara theorem from the delocalization properties of an eigenvector. Assume that p is fixed to make the presentation easier. Let $x \in S^{n-1}$ be a non-first eigenvector of A_G, and denote its coordinates by x_v, $v \in V$. Let P and N be the largest nodal domains of positive and negative coordinates. Since x is orthogonal to the first eigenvector having all positive coordinates, both P and N are non-empty. Denote $W = V \setminus (P \cup N)$. Our aim is to prove that with high probability, $W = \varnothing$. We start with proving a weaker statement that the cardinality of W is small.

Proposition 5.2.1.

$$|W| \leqslant C\frac{\log^2 n}{p^2}$$

with probability $1 - o(1)$.

Proof. Pick a vertex from each positive nodal domain. These vertices cannot be connected by edges as they belong to different connected components, so they form an independent set. Using Proposition 5.1.1 (1), we derive that, with high probability, the number of such domains does not exceed $C\frac{\log n}{p}$. The same bound holds for the number of negative nodal domains.

Consider a nodal domain $W_0 \subset W$ and assume that $|W_0| \geqslant C\frac{\log n}{p}$. If this domain is positive, $|P| \geqslant C\frac{\log n}{p}$ as well, since P is the largest nodal domain. This contradicts Proposition 5.1.1 (2) as two nodal domains of the same sign cannot be connected. Combining this with the previous argument, we complete the proof of the proposition. □

Now, we are ready to prove that $W = \varnothing$ with probability $1 - o(1)$. Assume to the contrary that there is a vertex $v \in W$, and assume that $x_v < 0$. Let $\Gamma(v)$ be the set of its neighbors in G. Then $\Gamma(v) \cap N = \varnothing$ as otherwise v would be an element of N. Since x is an eigenvector,

$$\lambda x_v = \sum_{u\in\Gamma(v)} x_u = \sum_{u\in\Gamma(v)\cap P} x_u + \sum_{u\in\Gamma(v)\cap W} x_u.$$

Here $|\lambda| \leqslant \sqrt{np}$ because λ is a non-first eigenvalue. Then

$$\begin{aligned}\left\|x|_{\Gamma(v)}\right\|_1 &\leqslant \sum_{u\in\Gamma(v)\cap P} x_u + \sum_{u\in\Gamma(v)\cap W} |x_u| \leqslant 2 \sum_{u\in\Gamma(v)\cap W} |x_u| + |\lambda|\cdot|x_v| \\ &\leqslant (2|\Gamma(v)\cap W| + |\lambda|)\cdot \|x\|_\infty .\end{aligned}$$

By Proposition 5.2.1 and (5.1.2), this quantity does not exceed $\log^C n$ (recall that we assumed that $p \in (0,1)$ is fixed). Applying (5.1.2) another time, we conclude that

$$\left\|x|_{\Gamma(v)}\right\|_2 \leqslant \sqrt{\left\|x|_{\Gamma(v)}\right\|_1 \cdot \|x\|_\infty} \leqslant n^{-1/4}\log^C n.$$

In combination with Proposition 5.1.1 (3), this shows that a large set $\Gamma(v)$ carries a small mass, which contradicts the no-gaps delocalization. This completes the proof of Dekel-Lee-Linial-Arora-Bhaskara theorem.

The same argument shows that, with high probability, any vertex of the positive nodal domain is connected to the negative domain and vice versa, answering positively an old question of Linial. More precisely, we have the following stronger statement.

Lemma 5.2.2. *Let $p \in (0,1)$. Let $x \in S^{n-1}$ be a non-first eigenvector of A_G. Let $V = P \cup N$ be the decomposition of V into the positive and negative nodal domains corresponding to x. Then with probability greater than $1 - \exp(-c'\log^2 n)$, any vertex in P has at least $\frac{n}{\log^C n}$ neighbors in N, and any vertex in N has at least $\frac{n}{\log^C n}$ neighbors in P.*

Proof. Since λ is a non-first eigenvalue, $|\lambda| \leqslant c\sqrt{n}$ with high probability. Assume that the vector x is delocalized in both ℓ_∞ and no-gaps sense. Let $w \in P$, and

assume that

$$|\Gamma(w) \cap N| \leqslant \frac{n}{\log^{4C} n},$$

where $\Gamma(w)$ denotes the set of neighbors of w. We have

$$\lambda x_w = \sum_{v \in \Gamma(w) \cap P} x_v + \sum_{v \in \Gamma(w) \cap N} x_v,$$

and as before,

$$\begin{aligned}\left\| x|_{\Gamma(w)} \right\|_1 &= \sum_{v \in \Gamma(w) \cap P} x_v + \sum_{v \in \Gamma \cap N} |x_v| \leqslant 2 \sum_{v \in \Gamma(w) \cap N} |x_v| + |\lambda| \cdot |x_w| \\ &\leqslant 2 \frac{n}{\log^{4C} n} \cdot \frac{\log^C n}{\sqrt{n}} + c\sqrt{n} \cdot \frac{\log^{4C} n}{\sqrt{n}}.\end{aligned}$$

Hence,

$$\left\| x|_{\Gamma(w)} \right\|_2 \leqslant \sqrt{\|x\|_\infty \cdot \left\| x|_{\Gamma(w)} \right\|_1} \leqslant \frac{\sqrt{2}}{\log^C n},$$

which contradicts the no-gaps delocalization, as $|\Gamma(w)| \geqslant cnp$ with high probability. The proof finishes by application of the union bound over w. □

5.3. Spectral gap of the normalized Laplacian and Braess's paradox In some cases, the addition of a new highway to an existing highway system may increase the traffic congestion. This phenomenon discovered in 1968 by Braess became known as Braess's paradox. Since its discovery, a number of mathematical models have been suggested to explain this paradox. We will consider one such model suggested by Chung et. al. [8].

We will model the highway system by an Erdős-Rènyi graph $G(n, p)$. The congestion of the graph will be measured in terms of its *normalized Laplacian* which we will define in a moment. Let A_G be the adjacency matrix of the graph G, and let $D_G = (d_v,\ v \in V)$ be $n \times n$ the diagonal matrix whose diagonal entries are the degrees of the vertices. The normalized Laplacian of G is defined as

$$\mathcal{L}_G := I_n - D_G^{-1/2} A_G D_G^{-1/2}.$$

The normalized Laplacian is a positive semidefinite matrix, so it has a real non-negative spectrum that we write in increasing order: $0 = \lambda_1(\mathcal{L}_G) \leqslant \ldots \leqslant \lambda_n(\mathcal{L}_G)$. The eigenvalue $\lambda_1(\mathcal{L}_G) = 0$ corresponds to the eigenvector Y, whose coordinates are $Y_v = d_v^{1/2}$, $v \in V$. The quantity $\lambda_2(\mathcal{L}_G)$ is called *the spectral gap* of G. The spectral gap appears in the Poincare inequality, so it is instrumental in establishing measure concentration properties of various functionals, see, e.g. [16]. Also, the reciprocal of the spectral gap defines the relaxation time for a random walk on a graph, [17]. In this quality, it can be used to measure the congestion of the graph considered as a traffic network: the smaller spectral gap corresponds to a bigger congestion.

For a graph G, and let $a_-(G)$ be the fraction of non-edges $(u, v) \notin E$ such that the addition of (u, v) to the set of edges decreases the spectral gap. Intuitively,

the addition of an edge should increase the spectral gap as it brings the graph closer to the complete one, for which the spectral gap is maximal. However, the numerical experiments showed that the addition of an edge to a random graph frequently yields an opposite effect. This numerical data led to the following conjecture, which is a variant of the original conjecture of Chung.

Conjecture 5.3.1. For $p \in (0,1)$ fixed, there exists a constant $c(p)$ such that

$$\lim_{n\to\infty} \mathbb{P}\left\{a_-(G) \geqslant c(p)\right\} = 1.$$

This conjecture has been proved by Eldan, Ràsz, and Shramm [10]. Their proof is based on the following deterministic condition on the eigenvectors which ensures that the spectral gap decreases after adding an edge.

Proposition 5.3.2. *Let* G *be a graph such that* $(1/2)np \leqslant d_v \leqslant (3/2)np$ *for all vertices* $v \in V$. *Let* $x \in S^{n-1}$ *be the eigenvector of* $\mathcal{L}_G$ *corresponding to* $\lambda_2(G)$. *If* $(u,w) \notin E$ *is a non-edge, and*

$$\frac{1}{\sqrt{np}}\left(x_u^2 + x_w^2\right) + c_1(np)^{-2} < c_2 x_u x_v,$$

then the addition of the edge (u,w) *to* G *decreases the spectral gap.*

The proof of proposition 5.3.2 requires a tedious, although a rather straightforward calculation. Denote by $y \in S^{n-1}$ the *first* eigenvector of the Laplacian of graph G_+ obtained from G by adding the edge (u,w), and let $Q : \mathbb{R}^n \to \mathbb{R}^n$ be the orthogonal projection on the space $y^\perp$. By the variational definition of the second eigenvalue,

$$\lambda_2(G_+) = \inf_{z \in y^\perp \setminus \{0\}} \frac{\langle z, \mathcal{L}_{G_+} z\rangle}{\|z\|_2^2} \leqslant \frac{\langle Qx, \mathcal{L}_{G_+} Qx\rangle}{\|Qx\|_2^2} = \frac{\langle x, \mathcal{L}_{G_+} x\rangle}{1 - \langle x, y\rangle^2},$$

where the last equality follows since $\mathcal{L}_{G_+} y = 0$. In the last formula, $y = \Delta / \|\Delta\|_2$, where Δ is the vector with coordinates $\Delta_v = \sqrt{d_v}$ for $v \notin \{u,w\}$ and $\Delta_v = \sqrt{d_v + 1}$ for $v \in \{u,w\}$. The matrix $\mathcal{L}_{G_+}$ can be represented in a similar way:

$$\mathcal{L}_{G_+} = I_n - D_{G_+}^{-1/2} A_{G_+} D_{G_+}^{-1/2},$$

where $A_G + (e_u e_w^T + e_w e_u^T)$ and D_{G_+} is defined as D_G above. The proposition follows by substituting these formulas in the previous estimate of $\lambda_2(G_+)$ and simplifying the resulting expression. A reader can find the detailed calculation in [10].

Proposition 5.3.2 allows us to lower bound $a_-(G)$. The main technical tool in obtaining such a bound is delocalization. We will need both the ℓ_∞ and no-gaps delocalization of the second eigenvector of $\mathcal{L}_G$. Both properties hold for the eigenvectors of A_G, so our task is to extend them to the normalized Laplacian.

To derive the ℓ_∞ delocalization, we need some information on the distribution of the eigenvalues of A_G. The classical Wigner semi-circular law states that as $n \to \infty$, the percentage of eigenvalues of $n^{-1/2} A_G$ lying in any fixed interval

$(b, b+\theta) \subset \mathbb{R}$ approaches

$$\int_b^{b+\theta} \phi_{sc}(x)\,dx, \quad \text{where } \phi_{sc}(x) = \frac{1}{2\pi}\sqrt{(4-x^2)_+}.$$

Moreover, the *local* semi-circular law asserts that the same phenomenon holds on a short scale, namely when $\theta = \theta(n) = \Omega\Big(\frac{\log^C n}{n}\Big)$ with some absolute constant $C > 0$. Since ϕ_{sc} is a bounded function, this implies that as $n \to \infty$, an interval $(b, b+\rho)$ with $\rho = \Omega(1)$ should contain $O(\rho\sqrt{n})$ eigenvalues. Since we work with a fixed n instead of $n \to \infty$, we need a non-asymptotic version of the local semi-circular law proved by Erdős et. al. ([11], Theorem 2.10). Their result implies the following upper estimate for the number of eigenvalues in a fixed interval.

Theorem 5.3.3. *Let $b \in \mathbb{R}$ and $\rho \geqslant 1$. With probability greater than $1-\exp(-c\log^2 n)$, the interval $[b, b+\rho]$ contains at most*

$$N(\rho) := c'\rho\sqrt{n}$$

eigenvalues of A_G. The constants $c, c' > 0$ are absolute.

Theorem 5.3.3 together with the ℓ_∞ delocalization for the eigenvectors of the adjacency matrix allows to prove a similar delocalization for the normalized Laplacian.

Lemma 5.3.4. *Let $p \in (0,1)$. Let $f \in S^{n-1}$ be the second eigenvector of $\mathcal{L}_G$. Then with probability at least $1-\exp(-c\log^2 n)$,*

$$\|f\|_\infty \leqslant n^{-1/4}\log^C n$$

and

$$\Big|\Big\{v \in V:\ |f_v| \leqslant n^{-5/8}\Big\}\Big| \leqslant c' n^{1-1/48}.$$

Here, C, c, c' are positive constants whose value may depend on p.

Proof. Let us start with the ℓ_∞ delocalization. Let $d = np$ be the expected degree of a vertex, and set

$$x = d^{1/2} D_G^{-1/2} f.$$

By Proposition 5.1.1 (3), $d^{1/2}D_G^{-1/2} = \mathrm{diag}(s_v,\ v \in V)$, where $s_v = 1 + o(1)$ for all $v \in V$, and $\|x\|_2 = 1 + o(1)$ with probability close to 1. Hence, it is enough to bound $\|x\|_\infty$. Let us check that x is an approximate eigenvector of A_G corresponding to the approximate eigenvalue $\hat{\lambda}_2 d$, where $\hat{\lambda}_2$ is the second eigenvalue of the normalized adjacency matrix $D_G^{-1/2} A_G D_G^{-1/2}$. By Proposition 5.1.1 (4), $\hat{\lambda}_2 \leqslant c/\sqrt{np}$ with high probability, hence

$$\begin{aligned}
\Big\|A_G D_G^{-1/2} f - \hat{\lambda}_2 d D_G^{-1/2} f\Big\|_2 &= |\hat{\lambda}_2| \cdot \Big\|D_G^{1/2} f - d D_G^{-1/2} f\Big\|_2 \\
&\leqslant \frac{c}{\sqrt{np}} \cdot \max_{v \in V} d_v^{-1/2} \cdot \max_{v \in V} |d_v - d| \\
&\leqslant \frac{c}{np} \cdot \max_{v \in V} |d_v - d| \leqslant \frac{C\log n}{\sqrt{np}},
\end{aligned}$$

and so

$$\left\|A_G x - \hat{\lambda}_2 d x\right\|_2 \leqslant C \log n. \tag{5.3.5}$$

Denote the eigenvalues of A_G by $\mu_1, \dots, \mu_n$ and the corresponding eigenvectors by $u_1, \dots, u_n \in S^{n-1}$, and let $\alpha_j = \langle x, u_j \rangle$. Set $\mu = \hat{\lambda}_2 d$ and let P_τ be the orthogonal projection on the span of the eigenvectors corresponding to the eigenvalues of A_G in the interval $[\mu - \tau, \mu + \tau]$. Then

$$\begin{aligned}\tau \|(I - P_\tau)x\|_2 = \tau \left(\sum_{|\mu_j - \mu| > \tau} \alpha_j^2 \right)^{1/2} &\leqslant \left(\sum_{|\mu_j - \mu| > \tau} (\mu_j - \mu)^2 \alpha_j^2 \right)^{1/2} \\ &\leqslant \|(A_G - \mu)x\|_2 \leqslant C \log n.\end{aligned}$$

and so,

$$\|(I - P_\tau)x\|_2 \leqslant \left(C \frac{\log n}{\tau} \wedge 1 \right). \tag{5.3.6}$$

For any $\tau \geqslant 0$ and any $\rho \geqslant 1$,

$$\begin{aligned}\|(P_{\tau+\rho} - P_\tau)x\|_\infty = \left\| \sum_{|\mu_j - \mu| \in [\tau, \tau+\rho]} \alpha_j u_j \right\|_\infty &= \max_{\nu \in V} \left| \sum_{|\mu_j - \mu| \in [\tau, \tau+\rho]} \alpha_j u_{j,\nu} \right| \\ &\leqslant \left(\sum_{|\mu_j - \mu| \in [\tau, \tau+\rho]} \alpha_j^2 \right)^{1/2} \cdot \max_{\nu \in V} \left(\sum_{|\mu_j - \mu| \in [\tau, \tau+\rho]} u_{j,\nu}^2 \right)^{1/2} \\ &\leqslant \|(P_{\tau+\rho} - P_\tau)x\|_2 \cdot N^{1/2}(\rho) \cdot \max_{j \in [n]} \|u_j\|_\infty \\ &\leqslant \|(I - P_\tau)x\|_2 \cdot \sqrt{\rho n^{1/2}} \cdot \frac{\log^C n}{\sqrt{n}}\end{aligned}$$

with probability greater than $1 - \exp(-c \log^2 n)$, where we used (5.1.2) and Theorem 5.3.3 in the last inequality. Combining this with (5.3.6), we get

$$\|(P_{\tau+\rho} - P_\tau)x\|_\infty \leqslant C \sqrt{\rho} \frac{\log^C n}{n^{1/4}} \cdot \left(\tau^{-1} \wedge 1 \right).$$

By the union bound, with probability greater than $1 - \exp(-c \log^2 n)$, the same inequality holds for all $\tau = \rho = 2^k$, such that $1 \leqslant 2^k \leqslant 2n$. Also, with probability at least $1 - \exp(-cn)$, $I = P_{2n}$, i.e., there are no eigenvalues outside of the interval $[-2n, 2n]$. Therefore,

$$\begin{aligned}\|x\|_\infty &\leqslant \|P_1 x\|_\infty + \sum_{k=0}^{\log_2 2n} \|(P_{2^{k+1}} - P_{2^k})x\|_\infty \\ &\leqslant C \frac{\log^C n}{n^{1/4}} + \sum_{k=0}^{\log_2 2n} C 2^{-k/2} \frac{\log^C n}{n^{1/4}} \leqslant C n^{-1/4} \log^C n\end{aligned}$$

with the required probability. By the discussion above, $\|f\|_\infty \leqslant 2\|x\|_\infty$ which finishes the proof of the first part of the lemma.

Now, let us prove the lower bound on the absolute values of most of the coordinates of f. As before, it is enough to prove a similar bound on the coordinates of x. Assume to the contrary that there is a set $U \subset V$ with $|U| > cn^{1-1/48}$ such that for any $v \in U$, $|x_v| \leqslant n^{-5/8}$. Then

$$\|x_U\|_2 \leqslant \sqrt{n} \cdot n^{-5/8} = n^{-1/8}.$$

Inequality (5.3.5) shows that x is an approximate eigenvector of A_G. Since, by Remarks 1.0.9 and 2.1.8, $n^{-1/8} \gg Cn^{-1/2}\log^C n$, we can apply Theorem 1.0.7 to x with s being an appropriately small constant and $\varepsilon = (1/s)n^{-1/48}$, so $(\varepsilon s)^6 = n^{-1/8}$. This theorem shows that such set U exists with probability at most $\exp(-\varepsilon n) \ll \exp(-c\log^2 n)$. The proof of the lemma is complete. □

Equipped with Proposition 5.3.2 and Lemma 5.3.4, we can prove a stronger form of the conjecture showing that $c \geqslant 1/2 - o(1)$. Let us formulate it as a theorem.

Theorem 5.3.7. *Let $p \in (0,1)$, and let G be a $G(n,p)$ graph. Then with probability $1-o(1)$,*

$$a_-(G) \geqslant \frac{1}{2} - O(n^{-c}).$$

Proof. Let $f \in S^{n-1}$ be the eigenvector of $\mathcal{L}_G$ corresponding to the second eigenvalue, and assume that the event described in Lemma 5.3.4 occurs. Let

$$W = \left\{ v \in V : \ |f_v| \geqslant n^{-5/8} \right\},$$

and set

$$W_+ = \{v \in W : \ f_v > 0\}, \quad \text{and} \quad W_- = \{v \in W : \ f_v < 0\}.$$

For any $v, w \in W_+$,

$$\frac{f_v^2 + f_w^2}{f_v f_w} \leqslant 2 \max_{v,w \in W_+} \frac{f_v}{f_w} \leqslant Cn^{3/8}\log^C n \ll \sqrt{n}.$$

Hence, if (v,w) is a non-edge, then Proposition 5.3.2 implies that adding it to G decreases the spectral gap. Similarly, we can show that adding any non-edge whose vertices belong to W_-, decreases the spectral gap as well. Let us count the number of the non-edges in W_+ and W_- and compare it to the total number of the non-edges. Using Property (5), and the bound $|W^c| \leqslant cn^{1-1/48}$, we obtain

$$\begin{aligned} a_-(G) &\geqslant \frac{|\text{Non-edges}(W_+)| + |\text{Non-edges}(W_-)|}{|\text{Non-edges}(V)|} \\ &\geqslant \frac{(1-p)\left[\binom{|W_+|}{2} + \binom{|W_-|}{2}\right] - 2n^{-3/2}}{(1-p)\binom{n}{2} + n^{3/2}} \\ &\geqslant \frac{(1-p)\left[\left(\frac{|W_+|+|W_-|}{2}\right)^2 - |W_+| - |W_-|\right] - 2n^{3/2}}{(1-p)\binom{n}{2} + n^{3/2}} \geqslant \frac{1}{2} - O(n^{-c}), \end{aligned}$$

as claimed. □

References

[1] S. Arora and A. Bhaskara, *Eigenvectors of random graphs: delocalization and nodal domains*, 2011. Manuscript, http://www.cs.princeton.edu/~bhaskara/files/deloc.pdf. ←30

[2] K. Ball, *Cube slicing in* $\mathbb{R}^n$, Proc. Amer. Math. Soc. **97** (1986), 465–473. ←11

[3] K. Ball, *Volumes of sections of cubes and related problems*, Geometric aspects of functional analysis (1987–88), 1989, pp. 251–260. ←12, 15

[4] K. Ball and F. Nazarov, *Little level theorem and zero-Khinchin inequality*, 1996. Manuscript, http://www.math.msu.edu/~fedja/prepr.html. ←11

[5] F. Barthe, *Inégalités de Brascamp-Lieb et convexité*, C. R. Acad. Sci. Paris Sér. I Math. **324** (1997), no. 8, 885–888. ←15

[6] P. Bourgade and H.-T. Yau, *The eigenvector moment flow and local quantum unique ergodicity*, Comm. Math. Phys. **350** (2017), no. 1, 231–278. ←2

[7] H. J. Brascamp and E. H. Lieb, *Best constants in Young's inequality, its converse, and its generalization to more than three functions*, Advances in Math. **20** (1976), 151–173. ←15

[8] F. Chung, S. Young, and W. Zhao, *Braess's paradox in expanders*, Random Structures and Algorithms **41** (2012), no. 4, 451–468. ←32

[9] Y. Dekel, J. R. Lee, and N. Linial, *Eigenvectors of random graphs: nodal domains*, Random Structures and Algorithms **39** (2011), no. 1, 39–58. ←30

[10] R. Eldan, M. Rász, and T. Schramm, *Braess's paradox for the spectral gap in random graphs and delocalization of eigenvectors*, Random Structures and Algorithms **50** (2017), no. 4, 584–611. ←29, 33

[11] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, *Spectral statistics of Erdős-Rényi graphs I: local semicircle law*, Annals of Probability **41** (2013), no. 3B, 2279–2375. ←30, 34

[12] L. Erdős, B. Schlein, and H.-T. Yau, *Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices*, Annals of Probability **37** (2009), 815–852. ←2

[13] L. Erdős, B. Schlein, and H.-T. Yau, *Local semicircle law and complete delocalization for Wigner random matrices*, Comm. Math. Phys. **287** (2009), 641–655. ←2

[14] F. Götze, A. Naumov, A. Tikhomirov, and D. Timushev, *On the local semicircular law for Wigner ensembles*, 2016. arXiv:1602.03073v2. ←2

[15] G. Halász, *Estimates for the concentration function of combinatorial number theory and probability*, Periodica Mathematica Hungarica **8** (1977), 197–211. ←14

[16] M. Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001. ←32

[17] Y. Peres D. A. Levin E. L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, Providence, RI, 2017. ←32

[18] G. Livshyts, G. Paouris, and P. Pivovarov, *On sharp bounds for marginal densities of product measures*, Israel J. Math. **216** (2016), no. 2, 877–889. ←12

[19] H. Nguyen, T. Tao, and V. Vu, *Random matrices: tail bounds for gaps between eigenvalues*, Probability Theory Related Fields **167** (2017), no. 3–4, 777–816. ←30

[20] S. O'Rourke, V. Vu, and K. Wang, *Eigenvectors of random matrices: a survey*, J. Combin. Theory Ser. A **144** (2016), 361–442. ←3

[21] M. Rudelson, *Recent developments in non-asymptotic theory of random matrices*, Modern aspects of random matrix theory, 2014, pp. 1576–1602. ←2

[22] M. Rudelson and R. Vershynin, *Non-asymptotic theory of random matrices: extreme singular values*, Proceedings of the International Congress of Mathematicians, 2010, pp. 83–120. ←2

[23] M. Rudelson and R. Vershynin, *Small ball probabilities for linear images of high dimensional distributions*, Int. Math. Res. Not. **19** (2015), 9594–9617. ←11

[24] M. Rudelson and R. Vershynin, *Delocalization of eigenvectors of random matrices with independent entries*, Duke Math. J. **164** (2015), no. 13, 2507–2538. ←2

[25] M. Rudelson and R. Vershynin, *No-gaps delocalization for general random matrices*, Geom. Funct. Anal. **26** (2016), no. 6, 1716–1776. ←3, 4, 5

[26] T. Tao, *Topics in random matrix theory*, Graduate Studies in Mathematics, vol. 132, American Mathematical Society, Providence, RI, 2012. ←3, 22

[27] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices*, Compressed sensing, 2012, pp. 210–268. ←2

[28] V. Vu and K. Wang, *Random weighted projections, random quadratic forms and random eigenvectors*, Random Structures and Algorithms **47** (2015), no. 4, 792–821. ←2

Department of Mathematics, University of Michigan.
Email address: rudelson@umich.edu