UC Irvine
UC Irvine Previously Published Works

Title
Delocalization of eigenvectors of random matrices with independent entries

Permalink
https://escholarship.org/uc/item/8m77h9v9

Journal
Duke Mathematical Journal, 164(13)

ISSN
0012-7094

Authors
Rudelson, Mark
Vershynin, Roman

Publication Date
2015-10-01

DOI
10.1215/00127094-3129809

Copyright Information

Peer reviewed

DELOCALIZATION OF EIGENVECTORS OF RANDOM MATRICES WITH INDEPENDENT ENTRIES

MARK RUDELSON AND ROMAN VERSHYNIN

ABSTRACT. We prove that an $n \times n$ random matrix G with independent entries is completely delocalized. Suppose the entries of G have zero means, variances uniformly bounded below, and a uniform tail decay of exponential type. Then with high probability all unit eigenvectors of G have all coordinates of magnitude $O(n^{-1/2})$, modulo logarithmic corrections. This comes a consequence of a new, geometric, approach to delocalization for random matrices.

1. INTRODUCTION

This paper establishes a complete delocalization of random matrices with independent entries having variances of the same order of magnitude. For an $n \times n$ matrix G, complete delocalization refers to the situation where all unit eigenvectors v of G have all coordinates of the smallest possible magnitude $n^{-1/2}$, up to logarithmic corrections. For example, a random matrix G with independent standard normal entries is completely delocalized with high probability. Indeed, by rotation invariance the unit eigenvectors v are uniformly distributed on the sphere S^{n-1}, so with high probability one has $\|v\|_\infty = \max_{i \le n} |v_i| = O(\sqrt{\log(n)/n})$ for all v.

Rotation-invariant ensembles seem to be the only example where delocalization can be obtained easily. Only recently was it proved by L. Erdös *et al.* that general symmetric and Hermitian random matrices H with independent entries are completely delocalized [11, 12, 13]. These results were later extended by L. Erdös *et al.* [15] and by Tao and Vu [35], see also surveys [14, 5]. Very recently, the optimal bound $O(\sqrt{\log n/n})$ was obtained by Vu and Wang [39] for the "bulk" eigenvectors of Hermitian matrices. Delocalization properties with varying degrees of strength and generality were then established for several other symmetric and Hermitian ensembles – band matrices [6, 7, 10], sparse matrices (adjacency matrices of Erdös-Renyi graphs) [37, 8, 9], heavy-tailed matrices [3, 1], and sample covariance matrices [4].

Despite this recent progress, no delocalization results were available for *non-Hermitian random matrices* prior to the present work. Similar to the Hermitian case, non-Hermitian random matrices have been successful in describing various physical phenomena, see [18, 19, 22, 17, 36] and the references therein. The distribution of eigenvectors of non-Hermitian random matrices has been studied in physics literature, mostly focusing on correlations of certain eigenvector entries, see [26, 21].

All previous approaches to delocalization in random matrices were spectral. Delocalization was obtained as a byproduct of local limit laws, which determine eigenvalue distribution on microscopic scales. For example, delocalization for symmetric random matrices was deduced from a local version of Wigner's semicircle law which controls the number of eigenvalues of H falling in short intervals, even down to intervals where the average number of eigenvalues is logarithmic in the dimension [11, 12, 13, 15].

In this paper we develop a new approach to delocalization of random matrices, which is geometric rather than spectral. The only spectral properties we rely on are crude bounds on the extreme singular values of random matrices. As a result, the new approach can work smoothly in situations where limit spectral laws are unknown or even impossible. In particular, one does not need to require that the variances of all entries be the same, or even that the matrix of variances be doubly-stochastic (as e.g. [15]).

The main result can be stated for random variables ξ with tail decay of exponential type, thus satisfying $\mathbb{P}\{|\xi| \geq t\} \leq 2\exp(-ct^\alpha)$ for some $c, \alpha > 0$ and all $t > 0$. One can express this equivalently by the growth of moments $\mathbb{E}\,|\xi|^p = O(p)^{p/\alpha}$ as $p \to \infty$, which is quantitatively captured by the norm

$$\|\xi\|_{\psi_\alpha} := \sup_{p \geq 1} p^{-1/\alpha}(\mathbb{E}\,|\xi|^p)^{1/p} < \infty.$$

The case $\alpha = 2$ corresponds to *sub-gaussian* random variables[1]. It is convenient to state and prove the main result for sub-gaussian random variables, and then deduce a similar result for general $\alpha > 0$ using a standard truncation argument.

Theorem 1.1 (Delocalization, subgaussian)**.** *Let G be an $n \times n$ real random matrix whose entries G_{ij} are independent random variables satisfying $\mathbb{E}\,G_{ij} = 0$, $\mathbb{E}\,G_{ij}^2 \geq 1$ and $\|G_{ij}\|_{\psi_2} \leq K$. Let $t \geq 2$. Then, with probability at least $1 - n^{1-t}$, the matrix G is completely delocalized, meaning that all eigenvectors v of G satisfy*

$$\|v\|_\infty \leq \frac{Ct^{3/2}\log^{9/2}n}{\sqrt{n}}\,\|v\|_2.$$

Here C depends only on K.

Remark 1.2 (Complex matrices)*.* The same conclusion as in Theorem 1.1 holds for a complex matrix G. One just needs to require that both real and imaginary parts of all entries are independent and satisfy the three conditions in Theorem 1.1.

Remark 1.3 (Logarithmic losses)*.* The exponent $9/2$ of the logarithm in Theorem 1.1 is suboptimal, and there are several points in the proof that can be improved. We believe that by taking care of these points, it is possible to improve the exponent to the optimal value $1/2$. However, such improvements would come at the expense of simplicity of the argument, while in this paper we aim at presenting the most transparent proof. The exponents $3/2$ is probably suboptimal as well.

Remark 1.4 (Dependence on sub-gaussian norms $\|G_{ij}\|_{\psi_2}$)*.* The proof of Theorem 1.1 shows that C depends polynomially on K, i.e., $C \leq 2K^{C_0}$ for some absolute constant C_0. This observation allows one to extend Theorem 1.1 to the situation where the entries G_{ij} of G have uniformly bounded ψ_α-norms, for any fixed $\alpha > 0$.

---

[1]Standard properties of sub-gaussian random variables can be found in [38, Section 5.2.3].

Corollary 1.5 (Delocalization, general exponential tail decay)**.** *Let G be an $n \times n$ real random matrix whose entries G_{ij} are independent random variables satisfying $\mathbb{E}\, G_{ij} = 0$, $\mathbb{E}\, G_{ij}^2 \geq 1$ and $\|G_{ij}\|_{\psi_\alpha} \leq M$. Let $t \geq 2$. Then, with probability at least $1 - n^{1-t}$, all eigenvectors v of G satisfy*

$$\|v\|_\infty \leq \frac{Ct^\beta \log^\gamma n}{\sqrt{n}} \, \|v\|_2.$$

Here C, β, γ depend only on $\alpha > 0$ and M.

Due to its spectral nature, our argument automatically establishes a stronger form of delocalization than was stated in Theorem 1.1. Indeed, one can show that not only eigenvectors but also *approximate eigenvectors* of G are delocalized. Given $\varepsilon > 0$, we call $v \in \mathbb{R}^n$ an ε-approximate eigenvector if there exists $z \in \mathbb{C}$ (an ε-approximate eigenvalue) such that $\|Gv - zv\|_2 \leq \varepsilon \|v\|_2$. We will prove the following extension of Theorem 1.1.

Theorem 1.6 (Delocalization of approximate eigenvectors)**.** *Let G be a random matrix as in Theorem 1.1, and let $t \geq 2$ and $s \geq 0$. Then, with probability at least $1 - n^{1-t(s+1)}$, all $(s/\sqrt{n})$-approximate eigenvectors v of G satisfy*

$$\|v\|_\infty \leq \frac{Ct^{3/2}(s+1)^{3/2} \log^{9/2} n}{\sqrt{n}} \, \|v\|_2. \tag{1.1}$$

Here C depends only on K.

Remark 1.7 (Further extensions)*.* The results in of this paper could be extended in several other ways. For instance, it is possible to drop the assumption that *all* variances of the entries are of the same order and prove a similar theorem for sparse random matrices. One can establish the *isotropic delocalization* in the sense of [2, 23]. We did not pursue these directions since it would have made the presentation heavier. It is also possible that a version of Theorems 1.1 and 1.6 can be proved for *Hermitian matrices*. We leave this direction for the future.

1.1. **Outline of the argument.** Our approach to Theorem 1.6 is based on a dimension reduction argument. If the matrix G has a localized approximate eigenvector, it will be detected from an imbalance of a suitable projection of G. As we shall see, this argument yields a lower bound on $\|(G - zI)v\|_2$ that is uniform over all unit vectors v such that $\|v\|_\infty \gg 1/\sqrt{n}$.

Let us describe this strategy in loose terms. We fix z and consider the random matrix $A = G - zI$; denote its columns by A_j. Consider a projection P whose kernel contains all A_j except for $j \in \{j_0\} \cup J_0$, where $j_0 \in [n]$ and $J \subset [n] \setminus \{j_0\}$ are a random uniform index and subset respectively, with $|J_0| = l \sim \log^{10} n$. We call such P a *test projection*. By its definition, triangle inequality and Cachy-Schwarz inequality, we have for any vector v that

$$\|Av\|_2 \geq \|PAv\|_2 = \Big\| \sum_{j=1}^n v_j PA_j \Big\|_2 = \Big\| v_{j_0} PA_{j_0} + \sum_{j \in J_0} v_j PA_j \Big\|_2$$

$$\geq |v_{j_0}| \|PA_{j_0}\|_2 - \Big(\sum_{j \in J_0} |v_j|^2 \Big)^{1/2} \Big(\sum_{j \in J_0} \|PA_j\|_2^2 \Big)^{1/2}.$$

Suppose that v is localized, say $\|v\|_\infty > l^2/\sqrt{n}$. Using the randomness of j_0 and J, with non-negligible probability (around $1/n$) we have $|v_{j_0}| = \|v\|_\infty > l^2/\sqrt{n}$ and $(\sum_{j \in J_0} |v_j|^2)^{1/2} \lesssim \sqrt{l/n}$. On this event, we have shown that

$$\|Av\|_2 \gtrsim \frac{l^2}{\sqrt{n}} \|PA_{j_0}\|_2 - \frac{l}{\sqrt{n}} \max_{j \in J_0} \|PA_j\|_2. \tag{1.2}$$

Since the right hand side of (1.2) does not depend on v, we obtained a uniform lower estimate for all localized vectors v.

It remains to estimate the magnitudes of $\|PA_j\|_2$ for $j \in \{j_0\} \cup J_0$. What helps us is that the test projection P can be made independent of the random vectors A_j appearing in (1.2). Since $A = G - zI$, we can represent $A_j = G_j - ze_j$ where e_j denote the standard basis vectors. Assume first that z is very close to zero, so $A_j \approx G_j$. Then, using concentration of measure we can argue that $\|PA_j\|_2 \approx \|PG_j\|_2 \sim \sqrt{l}$ with high probability (and thus for all $j \in \{j_0\} \cup J_0$ simultaneously). Substituting this into (1.2) we conclude that the nice lower bound

$$\|Av\|_2 \gtrsim \frac{l^{3/2}}{\sqrt{n}}$$

holds for all localized approximate eigenvectors v corresponding to (approximate) eigenvalues z that are very close to zero.

The challenging part of our argument is for z not close to zero, namely when the diagonal parts Pe_j dominates in the representation $PA_j = PG_j - zPe_j$. Estimating the magnitudes of $\|Pe_j\|_2$ might be as difficult as the original delocalization problem. However, it turns out that using concentration, it is possible to compare the terms $\|Pe_j\|_2$ with each other without knowing their magnitudes. This will require a careful construction of a test projection P.

The rest of the paper is organized as follows. In Section 2, we recall some known linear algebraic and probabilistic facts. In Section 3, we rigorously develop the argument that was informally described above. It reduces the delocalization problem to finding a test projection P for which the norms of the columns Pe_j have similar magnitudes. In Section 4, we shall develop a helpful tool for estimating $\|Pe_j\|_2$, an estimate of the distance between anisotropic random vectors and subspaces. In Section 5, we shall express $\|Pe_j\|_2$ in terms of such distances, and thus will be able to compare these terms with each other. In Section 6 we deduce Theorem 1.6. Finally, Appendix contains auxiliary results on the smallest singular values of random matrices.

2. Notation and preliminaries

We shall work with random variables ξ which satisfy the following assumption.

Assumption 2.1. ξ *is either real valued and satisfies*

$$\mathbb{E}\,\xi = 0, \quad \mathbb{E}\,\xi^2 \geq 1, \quad \|\xi\|_{\psi_2} \leq K, \tag{2.1}$$

or ξ is complex valued, where $\operatorname{Re}\xi$ and $\operatorname{Im}\xi$ are independent random variables each satisfying the three conditions in (2.1).

We will establish the conclusion of Theorem 1.6 for random matrices G with independent entries that satisfy Assumption 2.1. Thus we will simultaneously treat the real case and the complex case discussed in Remark 1.2.

We will regard the parameter K in Assumption 2.1 as a constant, thus $C, C_1, c, c_1, \ldots$ will denote positive numbers that may depend on K only; their values may change from line to line.

Without loss of generality, we can assume that G, as well as various other matrices that we will encounter in the proof, have full rank. This can be achieved by a perturbation argument, where one adds to G an independent Gaussian random matrix G' whose all entries are independent $N(0, \sigma^2)$ random variables with sufficiently small $\sigma > 0$. Such perturbation will not affect the proof of Theorem 1.6 since any ε-approximate eigenvector of G will be a (2ε)-approximate eigenvector of G' whenever $\|G - G'\| < \varepsilon$.

By $\mathbb{E}_X$, $\mathbb{P}_X$ we denote the conditional expectation and probability with respect to a random variable X, conditioned on all other variables.

The orthogonal projection onto a subspace E of $\mathbb{C}^m$ is denoted P_E. The canonical basis of $\mathbb{C}^n$ is denoted $e_1, \ldots, e_n$.

Let A be an $m \times n$ matrix; $\|A\|$ and $\|A\|_{\mathrm{HS}}$ denote the operator norm and Hilbert-Schmidt (Frobenius) norm of A, respectively. The singular values $s_i(A)$ are the eigenvalues of $(A^*A)^{1/2}$ arranged in a non-increasing order; thus $s_1(A) \geq \cdots \geq s_r(A) \geq 0$ where $r = \min(m, n)$. The extreme singular values have special meaning, namely

$$s_1(A) = \|A\| = \max_{x \in S^{n-1}} \|Ax\|_2, \quad s_m(A) = \|A^\dagger\|^{-1} = \min_{x \in S^{n-1}} \|Ax\|_2 \quad (\text{if } m \geq n).$$

Here $A^\dagger$ denotes the Moore-Penrose pseudoinverse of A, see e.g. [20]. We will need a few elementary properties of singular values.

Lemma 2.2 (Smallest singular value). *Let A be an $m \times n$ matrix and $r = \mathrm{rank}(A)$.*

(i) *Let P denote the orthogonal projection in $\mathbb{R}^n$ onto $\mathrm{Im}(A^*)$. Then $\|Ax\|_2 \geq s_r(A)\|Px\|_2$ for all $x \in \mathbb{R}^n$.*

(ii) *Let $r = m$. Then for every $y \in \mathbb{R}^m$, the vector $x = A^\dagger y \in \mathbb{R}^n$ satisfies $y = Ax$ and $\|y\|_2 \geq s_m(A)\|x\|_2$.*

Appendix A contains estimates of the smallest singular values of random matrices.

Next, we state a concentration property of sub-gaussian random vectors.

Theorem 2.3 (Sub-gaussian concentration). *Let A be a fixed $m \times n$ matrix. Consider a random vector $X = (X_1, \ldots, X_n)$ with independent components X_i which satisfy Assumption 2.1.*

(i) (Concentration) *For any $t \geq 0$, we have*

$$\mathbb{P}\left\{\left|\|AX\|_2 - M\right| > t\right\} \leq 2\exp\left(-\frac{ct^2}{\|A\|^2}\right)$$

where $M = (\mathbb{E}\|AX\|_2^2)^{1/2}$ satisfies $\|A\|_{\mathrm{HS}} \leq M \leq K\|A\|_{\mathrm{HS}}$.

(ii) (Small ball probability) *For every $y \in \mathbb{R}^m$, we have*

$$\mathbb{P}\left\{\|AX - y\|_2 < \frac{1}{6}(\|A\|_{\mathrm{HS}} + \|y\|_2)\right\} \leq 2\exp\left(-\frac{c\|A\|_{\mathrm{HS}}^2}{\|A\|^2}\right).$$

In both parts, $c = c(K) > 0$ is polynomial in K.

This result can be deduced from Hanson-Wright inequality. For part (ii), this was done in [24]. A modern proof of Hanson-Wright inequality and deduction of both

parts of Theorem 2.3 are discussed in [30]. There X_i were assumed to have unit variances; the general case follows by a standard normalization step.

Sub-gaussian concentration paired with a standard covering argument yields the following result on norms of random matrices, see [30].

Theorem 2.4 (Products of random and deterministic matrices)**.** *Let B be a fixed $m \times N$ matrix, and G be an $N \times n$ random matrix with independent entries that satisfy $\mathbb{E}\, G_{ij} = 0$, $\mathbb{E}\, G_{ij}^2 \geq 1$ and $\|G_{ij}\|_{\psi_2} \leq K$. Then for any $s, t \geq 1$ we have*

$$\mathbb{P}\left\{\|BG\| > C(s\|B\|_{\mathrm{HS}} + t\sqrt{n}\|B\|)\right\} \leq 2\exp(-s^2 r - t^2 n).$$

Here $r = \|B\|_{\mathrm{HS}}^2/\|B\|_2^2$ is the stable rank of B, and $C = C(K)$ is polynomial in K.

Remark 2.5. A couple of special cases in Theorem 2.4 are worth mentioning. If $B = P$ is a projection in $\mathbb{R}^N$ of rank r, then

$$\mathbb{P}\left\{\|PG\| > C(s\sqrt{r} + t\sqrt{n})\right\} \leq 2\exp(-s^2 r - t^2 n).$$

The same holds if $B = P$ is an $r \times N$ matrix such that $PP^* = I_r$.

In particular, for $B = I_N$ we obtain

$$\mathbb{P}\left\{\|G\| > C(s\sqrt{N} + t\sqrt{n})\right\} \leq 2\exp(-s^2 N - t^2 n).$$

3. Reducing delocalization to the existence of a test projection

We begin to develop a geometric approach to delocalization of random matrices. The first step, which we discuss in this section, is presented for a general random matrix A. Later it will be used for $A = G - zI_n$ where G is the random matrix from Theorem 1.6 and $z \in \mathbb{C}$.

We will first try to bound the probability of the following *localization event* for a random matrix A and parameters $l, W, w > 0$:

$$\mathcal{L}_{W,w} = \left\{\exists v \in S^{n-1} : \|v\|_\infty > W\sqrt{\frac{l}{n}} \text{ and } \|Av\|_2 \leq \frac{w}{\sqrt{n}}\right\}. \tag{3.1}$$

We will show that $\mathcal{L}_{W,w}$ is unlikely for $l \sim \log^2 n$, $W \sim \log^{7/2} n$ and $w = \mathrm{const}$.

In this section, we reduce our task to the existence of a certain linear map P which reduces dimension from n to $\sim l$, and which we call a *test projection*.

To this end, given an $m \times n$ matrix B, we shall denote by B_j the j-th column of B, and for a subset $J \subseteq [n]$, we denote by B_J the submatrix of B formed by the columns indexed by J. Fix n and $l \leq n$, and define the set of pairs

$$\Lambda = \Lambda(n,l) = \left\{(j, J) : j \in [n], J \subseteq [n] \setminus \{j\}, |J| = l - 1\right\}.$$

We equip Λ with the uniform probability measure.

Proposition 3.1 (Delocalization from test projection)**.** *Let $l \leq n$. Consider an $n \times n$ random matrix A with an arbitrary distribution. Suppose that to each $(j_0, J_0) \in \Lambda$ corresponds a number $l' \leq n$ and an $l' \times n$ matrix $P = P(n, l, A, j_0, J_0)$ with the following properties:*

(i) $\|P\| \leq 1$;
(ii) $\ker(P) \supseteq \{A_j\}_{j \notin \{j_0\} \cup J_0}$.

Let $\alpha, \kappa > 0$. Let $w > 0$ and $W = \frac{w}{\kappa l} + \frac{\sqrt{2}}{\alpha}$. Then we can bound the probability of the localization event (3.1) as follows:

$$\mathbb{P}_A \left(\mathcal{L}_{W,w} \right) \leq 2n \cdot \mathbb{E}_{(j_0, J_0)} \mathbb{P}_A \left(\mathcal{B}_{\alpha,\kappa}^c \mid (j_0, J_0) \right)$$

where $\mathcal{B}_{\alpha,\kappa}$ denotes the following balancing event:

$$\mathcal{B}_{\alpha,\kappa} = \left\{ \|PA_{j_0}\|_2 \geq \alpha \|PA_{J_0}\| \ and \ \|PA_{j_0}\|_2 \geq \kappa \sqrt{l} \right\}. \tag{3.2}$$

Proposition 3.1 states that in order to establish delocalization (as encoded by the complement of the event $\mathcal{L}_{W,w}$), it is enough to find a test projection P which satisfies the balancing property $\mathcal{B}_{\alpha,\kappa}$.

Proof. Let $v \in S^{n-1}$, $(j_0, J_0) \in \Lambda$ let P be as in the statement. Using the properties (i) and (ii) of P, we have

$$\|Av\|_2 \geq \|PAv\|_2 = \left\| \sum_{j=1}^{n} v_j PA_j \right\|_2 = \left\| v_{j_0} PA_{j_0} + \sum_{j \in J_0} v_j PA_j \right\|_2$$

$$\geq |v_{j_0}| \|PA_{j_0}\|_2 - \left(\sum_{j \in J_0} |v_j|^2 \right)^{1/2} \|PA_{J_0}\|. \tag{3.3}$$

The event $\mathcal{B}_{\alpha,\kappa}$ will help us balance the norms $\|PA_{j_0}\|_2$ and $\|PA_{J_0}\|$, while the following elementary lemma will help us balance the coefficients v_i.

Lemma 3.2 (Balancing the coefficients of v). *For a given $v \in S^{n-1}$ and for random $(j_0, J_0) \in \Lambda$, define the event*

$$\mathcal{V}_v = \left\{ |v_{j_0}| = \|v\|_\infty \ and \ \sum_{j \in J_0} |v_j|^2 \leq \frac{2l}{n} \right\}.$$

Then

$$\mathbb{P}_{(j_0, J_0)} \left(\mathcal{V}_v \right) \geq \frac{1}{2n}.$$

Proof of Lemma 3.2. Let $k_0 \in [n]$ denote a coordinate for which $|v_{k_0}| = \|v\|_\infty$. Then

$$\mathbb{P}_{(j_0, J_0)} \left(\mathcal{V}_v \right) \geq \mathbb{P}_{(j_0, J_0)} \left(\mathcal{V}_v \mid j_0 = k_0 \right) \cdot \mathbb{P}_{(j_0, J_0)} \left\{ j_0 = k_0 \right\}. \tag{3.4}$$

Conditionally on $j_0 = k_0$, the distribution of J_0 is uniform in the set $\{ J \subseteq [n] \setminus \{k_0\}, |J| = l - 1 \}$. Thus using Chebyshev's inequality we obtain

$$\mathbb{P}_{(j_0, J_0)} \left(\mathcal{V}_v^c \mid j_0 = k_0 \right) = \mathbb{P}_{J_0} \left\{ \sum_{j \in J_0} |v_j|^2 > \frac{2l}{n} \ \middle| \ j_0 = k_0 \right\}$$

$$\leq \frac{n}{2l} \mathbb{E}_{J_0} \left[\sum_{j \in J_0} |v_j|^2 \ \middle| \ j_0 = k_0 \right]$$

$$= \frac{n}{2l} \cdot \frac{l-1}{n} (\|v\|_2^2 - |v_{k_0}|^2) \leq \frac{1}{2}.$$

Moreover, $\mathbb{P}_{(j_0, J_0)} \left\{ j_0 = k_0 \right\} = \frac{1}{n}$. Substituting into (3.4), we complete the proof. $\square$

Assume that a realization of the random matrix A satisfies

$$\mathbb{P}_{(j_0, J_0)}\left(\mathcal{B}_{\alpha,\kappa} \,|\, A\right) > 1 - \frac{1}{2n}. \tag{3.5}$$

(We will analyze when this event occurs later.) Combining with the conclusion of Lemma 3.2, we see that there exists $(j_0, J_0) \in \Lambda$ such that both events $\mathcal{V}_v$ and $\mathcal{B}_{\alpha,\kappa}$ hold. Then we can continue estimating $\|Av\|_2$ in (3.3) using $\mathcal{V}_v$ and $\mathcal{B}_{\alpha,\kappa}$ as follows:

$$\|Av\|_2 \geq \|v\|_\infty \|PA_{j_0}\|_2 - \sqrt{\frac{2l}{n}}\, \|PA_{J_0}\| \geq \left(\|v\|_\infty - \frac{1}{\alpha}\sqrt{\frac{2l}{n}}\right)\kappa\sqrt{l},$$

provided the right hand side is non-negative. In particular, if $\|v\|_\infty > W\sqrt{l/n}$ where $W = \frac{w}{\kappa l} + \frac{\sqrt{2}}{\alpha}$, then $\|Av\|_2 > w/\sqrt{n}$. Thus the localization event $\mathcal{L}_{W,w}$ must fail.

Let us summarize. We have shown that the localization event $\mathcal{L}_{W,w}$ implies the failure of the event (3.5). The probability of this failure can be estimated using Chebyshev's inequality and Fubini theorem as follows:

$$\begin{aligned}
\mathbb{P}_A\left(\mathcal{L}_{W,w}\right) &\leq \mathbb{P}_A\left(\mathbb{P}_{(j_0,J_0)}\left\{\mathcal{B}_{\alpha,\kappa}^c \,|\, A\right\} > \frac{1}{2n}\right) \\
&\leq 2n \cdot \mathbb{E}_A\,\mathbb{P}_{(j_0,J_0)}\left(\mathcal{B}_{\alpha,\kappa}^c \,|\, A\right) \\
&= 2n \cdot \mathbb{E}_{(j_0,J_0)}\,\mathbb{P}_A\left(\mathcal{B}_{\alpha,\kappa}^c \,|\, (j_0,J_0)\right).
\end{aligned}$$

This completes the proof of Proposition 3.1. $\qquad\qquad\qquad\qquad\qquad\qquad \square$

3.1. **Strategy of showing that the balancing event is likely.** Our goal now is to construct a test projection P as in Proposition 3.1 in such a way that the balancing event $\mathcal{B}_{\alpha,\kappa}$ is likely for the random matrix $A = G - zI_n$ and for fixed $(j_0, J_0) \in \Lambda$ and $z \in \mathbb{C}$. We will be able to do this for $\alpha \sim (l\log^{3/2} n)^{-1}$ and $\kappa = c$.

We might choose P to be the orthogonal projection with

$$\ker(P) = \{A_j\}_{j \notin \{j_0\} \cup J_0}.$$

In reality, P will be a bit more adapted to A. Let us see what it will take to prove the two inequalities defining the balancing event $\mathcal{B}_{\alpha,\kappa}$ in (3.2). The second inequality can be deduced from the small ball probability estimate, Theorem 2.3(ii). Turning to the first inequality, note that

$$\|PA_{J_0}\| \sim \max_{j \in J_0}\|PA_j\|_2$$

up to a polynomial factor in $|J_0| = l - 1$ (thus logarithmic in n). So we need to show that

$$\|PA_{j_0}\|_2 \gtrsim \|PA_j\|_2 \quad \text{for all } j \in J_0.$$

Since $A = G - zI_n$, the columns A_i of A can be expressed as $A_i = G_i - ze_i$. Thus, informally speaking, our task is to show that with high probability,

$$\|PG_{j_0}\|_2 \gtrsim \|PG_j\|_2, \quad \|Pe_{j_0}\|_2 \gtrsim \|Pe_j\|_2 \quad \text{for all } j \in J_0. \tag{3.6}$$

The first inequality can be deduced from sub-gaussian concentration, Theorem 2.3. The second inequality in (3.6) is challenging, and most of the remaining work is devoted to validating it. Instead of estimating $\|Pe_j\|_2$, we will compare these terms with each other.

Later, in Proposition 5.3, we will relate $\|Pe_j\|_2$ to distances between anisotropic random vectors and subspaces. We will now digress to develop a general bound on such distances, which may be interesting on its own.

4. Distances between anisotropic random vectors and subspaces

We will be interested in the distribution of the distance $d(X, E_k)$ between a random vector $X \in \mathbb{R}^n$ and a k-dimensional random subspace E_k spanned by k independent vectors $X_1, \ldots, X_k \in \mathbb{R}^n$. A number of arguments in random matrix theory that appeared in the recent years rely on controlling such distances, see e.g. [33, 34, 28, 29, 16].

Let us start with the *isotropic* case, where the random vectors in question have all independent coordinates. Here one can use Theorem 2.3 to control the distances.

Proposition 4.1 (Distances between isotropic random vectors and subspaces)**.** *Let $1 \leq k \leq n$. Consider independent random vectors $X, X_1, X_2, \ldots, X_k$ with independent coordinates satisfying Assumption 2.1. Consider the subspace $E_k = \operatorname{span}(X_i)_{i=1}^k$. Then*

$$\sqrt{n-k} \leq \left(\mathbb{E}\, d(X, E_k)^2\right)^{1/2} \leq K\sqrt{n-k}.$$

Furthermore, one has
   (i) $\mathbb{P}\left\{d(X, E_k) \leq \frac{1}{2}\sqrt{n-k}\right\} \leq 2\exp(-c(n-k));$
   (ii) $\mathbb{P}\left\{d(X, E_k) > 2K\sqrt{n-k}\right\} \leq 2k\exp(-c(n-k)).$
Here $c = c(K) > 0$.

Proof. By adding small independent Gaussian perturbations to the vectors X_j, we can assume that $\dim(E_k) = k$ almost surely. We can represent the distance as

$$d(X, E_k) = \|P_{E_k^\perp} X\|_2.$$

Since $\|P_{E_k^\perp}\| = 1$ and $\|P_{E_k^\perp}\|_{\mathrm{HS}} = \sqrt{\dim(E_k^\perp)} = \sqrt{n-k}$, the conclusion of the proposition follows from Theorem 2.3 upon choosing $t = \frac{1}{2}\sqrt{n-k}$ in part (i) and $t = K\sqrt{n-k}$ in part (ii). $\qquad\square$

In this paper, we will need to control the distances in the more difficult *anisotropic* case, where all random vectors are transformed by a fixed linear map D. In other words, we will be interested in distances of the form $d(DX, E_k)$ where E_k is the span of the vectors $DX_1, \ldots, DX_k$. An ideal estimate should look like

$$d(DX, E_k) \asymp \left(\sum_{i>k} s_i(D)^2\right)^{1/2} \quad \text{with high probability,} \tag{4.1}$$

where $s_i(D)$ are the singular values of D arranged in the non-increasing order. To see why such estimate would make sense, note that in the isotropic case where $D = I_n$ the distance is of order $\sqrt{n-k}$, while for D of rank k or lower, the distance is zero.

The following result, based again on Theorem 2.3, establishes a somewhat weaker form of (4.1) with exponentially high probability.

Theorem 4.2 (Distances between anisotropic random vectors and subspaces)**.** *Let D be an $n \times n$ matrix with singular values[2] $s_i = s_i(D)$, and define $\bar{S}_m^2 = \sum_{i>m} s_i^2$*

---

[2]As usual, we arrange the singular values in a non-increasing order.

for $m \geq 0$. Let $1 \leq k \leq n$. Consider independent random vectors $X, X_1, X_2, \ldots, X_k$ with independent coordinates satisfying Assumption 2.1. Consider the subspace $E_k = \text{span}(DX_i)_{i=1}^k$. Then for every $k/2 \leq k_0 < k$ and $k < k_1 \leq n$, one has

(i) $\mathbb{P}\left\{d(DX, E_k) \leq c\bar{S}_{k_1}\right\} \leq 2\exp(-c(k_1 - k))$;

(ii) $\mathbb{P}\left\{d(DX, E_k) > CM(\bar{S}_{k_0} + \sqrt{k}\, s_{k_0+1})\right\} \leq 2k\exp(-c(k - k_0))$.

Here $M = Ck\sqrt{k_0}/(k - k_0)$ and $C = C(K)$, $c = c(K) > 0$.

Remark 4.3. It is important that the probability bounds in Theorem 4.2 are exponential in $k_1 - k$ and $k - k_0$. We will later choose $k \sim l \sim \log^2 n$ and $k_0 \approx (1 - \delta)k$, $k_1 \approx (1 + \delta)k$, where $\delta \sim 1/\log n$. This will allow us to make the exceptional probabilities Theorem 4.2 smaller than, say, n^{-10}.

Remark 4.4. As will be clear from the proof, one can replace the distance $d(DX, E_k)$ in part (ii) of Theorem 4.2 by the following bigger quantity:

$$\inf\left\{\left\|DX - \sum_{i=1}^k a_i DX_i\right\|_2 : \ a = (a_1, \ldots, a_k), \|a\|_2 \leq M\right\}.$$

Proof of Theorem 4.2. (i) We can represent the distance as

$$d(DX, E_k) = \|BX\|_2 \quad \text{where} \quad B = P_{E_k^\perp} D.$$

We truncate the singular values of B by defining an $n \times n$ matrix $\bar{B}$ with the same left and right singular vectors as B, and with singular values

$$s_i(\bar{B}) = \min\{s_i(B), s_{k_1-k}(B)\}.$$

Since $s_i(\bar{B}) \leq s_i(B)$ for all i, we have $\bar{B}\bar{B}^* \preceq BB^*$ in the p.s.d. order, which implies

$$\|\bar{B}X\|_2 \leq \|BX\|_2 = d(DX, E_k). \tag{4.2}$$

It remains to bound $\|\bar{B}X\|_2$ below. This can be done using Theorem 2.3(ii):

$$\mathbb{P}\left\{\|\bar{B}X\|_2 < c\|\bar{B}\|_{\text{HS}}\right\} \leq 2\exp\left(-\frac{c\|\bar{B}\|_{\text{HS}}^2}{\|\bar{B}\|^2}\right). \tag{4.3}$$

For $i > k_1 - k$, Cauchy interlacing theorem yields $s_i(\bar{B}) = s_i(B) \geq s_{i+k}(D)$, thus

$$\|\bar{B}\|_{\text{HS}}^2 = \sum_{i=1}^n s_i(\bar{B})^2 \geq (k_1 - k)s_{k_1-k}(B)^2 + \sum_{i>k_1-k} s_{i+k}(D)^2$$

$$= (k_1 - k)s_{k_1-k}(B)^2 + \bar{S}_{k_1}^2.$$

Further, $\|\bar{B}\| = \max_i s_i(\bar{B}) = s_{k_1-k}(B)$. In particular, $\|\bar{B}\|_{\text{HS}}^2 \geq \bar{S}_{k_1}^2$ and $\|\bar{B}\|_{\text{HS}}^2/\|\bar{B}\|^2 \geq k_1 - k$. Putting this along with (4.2) into (4.3), we complete the proof of part (i).

(ii) We truncate the singular value decomposition $D = \sum_{i=1}^n s_i u_i v_i^*$ by defining

$$D_0 = \sum_{i=1}^{k_0} s_i u_i v_i^*, \quad \bar{D} = \sum_{i=k_0+1}^n s_i u_i v_i^*.$$

By the triangle inequality, we have

$$d(DX, E_k) \leq d(D_0 X, E_k) + \|\bar{D}X\|_2. \tag{4.4}$$

We will estimate these two terms separately.

The second term, $\|\bar{D}X\|_2$, can be bounded using sub-gaussian concentration, Theorem 2.3(i). Since $\|\bar{D}\| = s_{k_0+1}$ and $\|\bar{D}\|_{\mathrm{HS}} = \bar{S}_{k_0}$, it follows that

$$\mathbb{P}\left\{\|\bar{D}X\|_2 > C\bar{S}_{k_0} + t\right\} \leq 2\exp\left(-ct^2/s_{k_0+1}^2\right), \quad t \geq 0.$$

Using this for $t = \sqrt{k}s_{k_0+1}$, we obtain that with probability at least $1 - 2\exp(-ck)$,

$$\|\bar{D}X\|_2 \leq C(\bar{S}_{k_0} + \sqrt{k}\,s_{k_0+1}). \tag{4.5}$$

Next, we estimate the first term in (4.4), $d(D_0X, E_k)$. Our immediate goal is to represent D_0X as a linear combination

$$D_0X = \sum_{i=1}^{k} a_i D_0 X_i \tag{4.6}$$

with some control of the norm of the coefficient vector $a = (a_1, \ldots, a_k)$. To this end, let us consider the singular value decomposition

$$D_0 = U_0 \Sigma_0 V_0^*; \quad \text{denote } P_0 = V_0^*.$$

Thus P_0 is a $k_0 \times n$ matrix satisfying $P_0 P_0^* = I_{k_0}$. Let G denote the $n \times k$ with columns $X_1, \ldots, X_k$.

We apply Theorem A.3 for the $k_0 \times k$ matrix $P_0 G$. It states that with probability at least $1 - 2k\exp(-c(k - k_0))$, we have

$$s_{k_0}(P_0 G) \geq c\frac{k - k_0}{k} =: \sigma_0. \tag{4.7}$$

Using Lemma 2.2(ii) we can find a coefficient vector $a = (a_1, \ldots, a_k)$ such that

$$P_0 X = P_0 G a = \sum_{i=1}^{k} a_i P_0 X_i, \tag{4.8}$$

$$\|a\|_2 \leq s_{k_0}(P_0 G)^{-1}\|P_0 X\|_2 \leq \sigma_0^{-1}\|P_0 X\|_2. \tag{4.9}$$

Multiplying both sides of (4.8) by $U_0 \Sigma_0$ and recalling that $D_0 = U_0 \Sigma_0 V_0^* = U_0 \Sigma_0 P_0$, we obtain the desired identity (4.6).

To finalize estimating $\|a\|_2$ in (4.9), recall that $\|P_0\|_{\mathrm{HS}}^2 = \mathrm{tr}(P_0 P_0^*) = \mathrm{tr}(I_{k_0}) = k_0$ and $\|P_0\| = 1$. Then Theorem 2.3(i) yields that with probability at least $1 - 2\exp(-ck_0)$, one has $\|P_0 X\|_2 \leq C\sqrt{k_0}$. Intersecting with the event (4.9), we conclude that with probability at least $1 - 4k\exp(-c(k - k_0))$, one has

$$\|a\|_2 \leq C\sigma_0^{-1}\sqrt{k_0} =: M. \tag{4.10}$$

Now we have representation (4.6) with a good control of $\|a\|_2$. Then we can estimate the distance as follows:

$$d(D_0X, E_k) = \inf_{z \in E_k}\|D_0X - z\|_2 \leq \left\|\sum_{i=1}^{k} a_i D_0 X_i - \sum_{i=1}^{k} a_i D X_i\right\|_2$$

$$= \left\|\sum_{i=1}^{k} a_i \bar{D} X_i\right\|_2 \leq \|a\|_2\|\bar{D}G\|.$$

(Recall that G denotes the $n \times k$ with matrix columns $X_1, \ldots, X_k$.) Applying Theorem 2.4, we have with probability at least $1 - 2\exp(-k)$ that

$$\|\bar{D}G\| \leq C(\|\bar{D}\|_{\mathrm{HS}} + \sqrt{k}\|\bar{D}\|) = C(\bar{S}_{k_0} + \sqrt{k}\,s_{k_0+1}).$$

Intersecting this with the event (4.10), we obtain with probability at least $1 - 6k\exp(-c(k-k_0))$ that

$$d(D_0X, E_k) \leq CM(\bar{S}_{k_0} + \sqrt{k}\,s_{k_0+1}).$$

Finally, we combine this with the event (4.5) and put into the estimate (4.4). It follows that with probability at least $1 - 8k\exp(-c(k-k_0))$, one has

$$d(DX, E_k) \leq C(M+1)(\bar{S}_{k_0} + \sqrt{k}s_{k_0+1}).$$

Due to our choice of M (in (4.10) and (4.7)), the theorem is proved.[3] $\square$

5. Construction of a test projection

We are now ready to construct a test projection P, which will be used later in Proposition 3.1.

Theorem 5.1 (Test projection). *Let $1 \leq l \leq n/4$ and $z \in \mathbb{C}$, $|z| \leq K_1\sqrt{n}$ for some absolute constant K_1. Consider a random matrix G as in Theorem 1.6, and let $A = G - zI_n$. Let A_j denote the columns of A. Then one can find an integer $l' \in [l/2, l]$ and an $l' \times n$ matrix P in such a way that l' and P are determined by l, n and $\{A_j\}_{j>l}$, and so that the following properties hold:*

(i) $PP^* = I_{l'}$;
(ii) $\ker P \supseteq \{A_j\}_{j>l}$;
(iii) with probability at least $1 - 2n^2\exp(-cl/\log n)$, one has

$$\|Pe_i\|_2 \leq C\sqrt{l}\,\log^{3/2}n \cdot \|Pe_j\|_2 \quad \text{for } 1 \leq i, j \leq l';$$
$$\|Pe_i\|_2 = 0 \quad \text{for } l' < i \leq l.$$

Here $C = C(K, K_1)$, $c = c(K, K_1) > 0$.

In the rest of this section we prove Theorem 5.1.

5.1. **Selection of the spectral window l'.** Consider the $n \times n$ random matrix A with columns A_j. Let $\bar{A}$ denote the $(n-l) \times (n-l)$ minor of A obtained by removing the first l rows and columns. By known invertibility results for random matrices, we will see that most singular values of $\bar{A}$, and thus also of $\bar{A}^{-1}$, are within a factor $n^{O(1)}$ from each other. Then we will find a somewhat smaller interval (a "spectral window") in which the singular values of $\bar{A}^{-1}$ are within constant factor from each other. This is a consequence of the following elementary lemma.

Lemma 5.2 (Improving the regularity of decay). *Let $s_1 \geq s_2 \geq \cdots \geq s_n$, and define $\bar{S}_k^2 = \sum_{j>k} s_j^2$ for $k \geq 0$. Assume that for some $l \leq n$ and $R \geq 1$, one has*

$$\frac{s_{l/2}}{s_l} \leq R. \tag{5.1}$$

Set $\delta = c/\log R$. Then there exists $l' \in [l/2, l]$ such that

$$\frac{s_{(1-\delta)\,l'}^2}{s_{(1+\delta)\,l'}^2} \leq 2 \quad \text{and} \quad \frac{\bar{S}_{(1-\delta)\,l'}^2}{\bar{S}_{(1+\delta)\,l'}^2} \leq 5. \tag{5.2}$$

---

[3]The factor 8 in the probability estimate can be reduced to 2 by adjusting c. We will use the same step in later arguments.

Proof. Let us divide the interval $[l/2, l]$ into $1/(8\delta)$ intervals of length $4\delta l$. Then for at least one of these intervals, the sequence s_i^2 decreases over it by a factor at most 2. Indeed, if this were not true, the sequence would decrease by a factor at least $2^{1/(8\delta)} > R$ over $[l/2, l]$, which would contradict the assumption (5.1). Set l' to be the midpoint of the interval we just found, thus

$$\frac{s_{l'-2\delta l}^2}{s_{l'+2\delta l}^2} \leq 2. \tag{5.3}$$

By monotonicity of s_i^2, this implies the first part of the conclusion (5.2). To see this, note that since $l' \leq l$, we have $l' - 2\delta l \leq (1-\delta)l' \leq (1+\delta)l' \leq l' + 2\delta l$.

To deduce the second part of (5.2), note that by monotonicity we have

$$\bar{S}_{l'-\delta l}^2 = \sum_{l'-\delta l < i \leq l'+\delta l} s_i^2 + \bar{S}_{l'+\delta l}^2 \leq 2\delta l \cdot s_{l'-2\delta l}^2 + \bar{S}_{l'+\delta l}^2, \tag{5.4}$$

$$\bar{S}_{l'+\delta l}^2 \geq \sum_{l'+\delta l < i \leq l'+2\delta l} s_i^2 \geq \delta l \cdot s_{l'+2\delta l}^2 \geq \frac{1}{2}\delta l \cdot s_{l'-2\delta l}^2, \tag{5.5}$$

where the very last inequality follows from (5.3). Estimates (5.4) and (5.5) together imply that $\bar{S}_{l'-\delta l}^2 \leq 5\bar{S}_{l'+\delta l}^2$. Like in the first part, we finish by monotonicity. $\square$

We shall apply Lemma 5.2 to the singular values of $\bar{A}^{-1}$, i.e. for

$$s_j = s_j(\bar{A}^{-1}), \quad \bar{S}_k^2 = \sum_{j > k} s_j^2.$$

To verify the assumptions of the lemma, we can use known estimates of the extreme singular values of random matrices. By Theorem 2.4 (see Remark 2.5), with probability at least $1 - \exp(-n)$, we have $\|G\| \leq C\sqrt{n}$, and thus

$$s_l^{-1} \leq s_n^{-1} = s_1(\bar{A}) = \|\bar{A}\| \leq \|A\| = \|G - zI_n\| \leq \|G\| + |z| \leq (C + K_1)\sqrt{n}.$$

Further, by Theorem A.2, with probability at least $1 - 2l\exp(-c(n-2l))$, one has

$$s_{l/2}^{-1} = s_{n-l/2+1}(\bar{A}) \geq \frac{c}{\sqrt{n}}.$$

(Here we used that $l \leq n/4$.) Summarizing, with probability at least $1 - 2n\exp(-cl)$,

$$\frac{c_1}{\sqrt{n}} \leq s_l \leq s_{l/2} \leq C_1\sqrt{n}. \tag{5.6}$$

Let us condition on $\bar{A}$ for which event (5.6) holds. We apply Lemma 5.2 with $R = (C_1/c_1)n$ and thus for

$$\delta = c/\log n. \tag{5.7}$$

We find $l' \in [l/2, l]$ such that (5.2) holds. Note that the value of l' depends only on the minor $\bar{A}$, thus only on $\{A_j\}_{j>l}$, as claimed in Theorem 5.1. Since we have conditioned on $\bar{A}$, the value of the "spectral window" l' is now fixed.

5.2. **Construction of P.** We construct P in two steps. First we define a matrix Q of the same dimensions that satisfies (ii) of the Theorem, and then obtain P by orthogonalization of the rows of Q.

Thus we shall look for an $l' \times n$ matrix Q that consists of three blocks of columns:

$$Q = \begin{bmatrix} q_{11} & 0 & \cdots & 0 & 0 & \cdots & 0 & \bar{q}_1^\mathsf{T} \\ 0 & q_{22} & \cdots & 0 & 0 & \cdots & 0 & \bar{q}_2^\mathsf{T} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & q_{l'l'} & 0 & \cdots & 0 & \bar{q}_{l'}^\mathsf{T} \end{bmatrix}, \quad q_{ii} \in \mathbb{C}, \ \bar{q}_i \in \mathbb{C}^{n-l}.$$

We require that Q satisfy condition (ii) in Theorem 5.1, i.e. that

$$\ker Q \supseteq \{A_j\}_{j>l}. \tag{5.8}$$

We explore this requirement in Section 5.4; for now let us assume that it holds.

Choose P to be an $l' \times n$ matrix that satisfies the following two defining properties:

(a) P has orthonormal rows;
(b) the span of the rows of P is the same as the span of the rows of Q.

One can construct P by Gram-Schmidt orhtogonalization of the rows of Q.

Note that the construction of P along with (5.8) implies (i) and (ii) of Theorem 5.1. It remains to estimate $\|Pe_j\|_2$ thereby proving (iii) of Theorem 5.1.

5.3. **Reducing $\|Pe_i\|_2$ to distances between random vectors and subspaces.**

Proposition 5.3 (Norms of columns of P via distances). *Let q_i denote the rows of Q and q_{ij} denote the entries of Q. Then:*

(i) *The values of $\|Pe_i\|_2$, $i \leq n$, are determined by Q, and they do not depend on a particular choice of P satisfying its defining properties (a), (b).*

(ii) *For every $i \leq l'$,*

$$\|Pe_i\|_2 = \frac{|q_{ii}|}{d(q_i, E_i)}, \quad \text{where } E_i = \text{span}(q_j)_{j \leq l', j \neq i}. \tag{5.9}$$

(iii) *For every $l' < i \leq l$, $\|Pe_i\|_2 = 0$.*

Proof. (i) Any P, P' that satisfy the defining properties (a), (b) must satisfy $P' = UP$ for some $l' \times l'$ unitary matrix U. It follows that $\|P'e_i\|_2 = \|Pe_i\|_2$ for all i.

(ii) Let us assume that $i = 1$; the argument for general i is similar. By part (i), we can construct the rows of P by performing Gram-Schmidt procedure on the rows of Q in any order. We choose the following order: $q_{l'}, q_{l'-1}, \ldots, q_1$, and thus construct the rows $p_{l'}, p_{l'-1}, \ldots, p_1$ of P. This yields

$$p_1 = \frac{\tilde{p}_1}{\|\tilde{p}_1\|_2}, \quad \text{where } \tilde{p}_1 = q_1 - P_{E_1} q_1 \tag{5.10}$$

$$p_j \in \text{span}(q_k)_{k \geq j}, \quad j = 1, \ldots, l' \tag{5.11}$$

Recall that we would like to estimate

$$\|Pe_1\|_2^2 = |p_{11}|^2 + |p_{21}|^2 \cdots + |p_{l'1}|^2 \tag{5.12}$$

where p_{ij} denote the entries of P.

First observe that all vectors in $E_1 = \text{span}(q_k)_{k \geq 2}$ have their first coordinate equal zero, because the same holds for the vectors q_k, $k \geq 2$, by the construction of Q.

Since $P_{E_1}q_1 \in E_1$, this implies by (5.10) that $\tilde{p}_{11} = q_{11}$. Further, again by (5.10) we have $\|\tilde{p}_1\|_2 = d(q_1, E_1)$. Thus

$$p_{11} = \frac{\tilde{p}_{11}}{\|\tilde{p}_1\|_2} = \frac{q_{11}}{d(q_1, E_1)}.$$

Next, for each $2 \leq j \leq l'$, (5.11) implies that $p_j \in \text{span}(q_k)_{k \geq 2} = E_1$, and thus the first coordinate of p_j equal zero. Using this in (5.12), we conclude that

$$\|Pe_1\|_2 = |p_{11}| = \frac{|q_{11}|}{d(q_1, E_1)}.$$

This completes the proof of (ii).

(iii) is trivial since $Qe_i = 0$ for all $l' < i \leq l$ by the construction of Q, while the rows of P are the linear combination of the rows of Q. $\square$

5.4. **The kernel requirement** (5.8). In order to estimate the distances $d(q_i, E_i)$ defined by the rows of Q, let us explore the condition (5.8) for Q. To express this condition algebraically, let us consider the $n \times (n - l)$ matrix $A^{(l)}$ obtained by removing the first l columns from A. Then (5.8) can be written as

$$QA^{(l)} = 0. \tag{5.13}$$

Let us denote the first l rows of $A^{(l)}$ by B_i^{T}, thus

$$A^{(l)} = \begin{bmatrix} B_1^{\mathsf{T}} \\ \vdots \\ B_l^{\mathsf{T}} \\ \hline \bar{A} \end{bmatrix}, \quad B_i \in \mathbb{C}^{n-l}. \tag{5.14}$$

Then (5.13) can be written as

$$q_{ii}B_i^{\mathsf{T}} + \bar{q}_i^{\mathsf{T}}\bar{A} = 0, \quad i \leq l'.$$

Without loss of generality, we can assume that the matrix $\bar{A}$ is almost surely invertible (see Section 2 for a perturbation argument achieving this). Multiplying both sides of the previous equations by $\bar{A}^{-1}$, we further rewrite them as

$$\bar{q}_i = -q_{ii}DB_i, \quad i \leq l', \qquad \text{where } D := (\bar{A}^{-1})^{\mathsf{T}}. \tag{5.15}$$

Thus we can choose Q to satisfy the requirement (5.8) by choosing $q_{ii} > 0$ arbitrarily and defining $\bar{q}_i$ as in (5.15).

5.5. **Estimating the distances, and completion of proof of Theorem 5.1.** We shall now estimate $\|Pe_i\|_2$, $1 \leq i \leq l'$, using identities (5.9) and (5.15). By the construction of Q and (5.15) we have

$$q_i = (0 \cdots q_{ii} \cdots 0 \; \bar{q}_i^{\mathsf{T}}) = -q_{ii}r_i, \quad \text{where} \quad r_i = (0 \cdots -1 \cdots 0 \; (DB_i)^{\mathsf{T}}).$$

Let us estimate $\|Pe_1\|_2$; the argument for general $\|Pe_i\|_2$ is similar. By (5.9),

$$\|Pe_1\|_2 = \frac{|q_{11}|}{d(q_{11}r_1, \text{span}(q_{jj}r_j)_{2 \leq j \leq l'})} = \frac{1}{d(r_1, \text{span}(r_j)_{2 \leq j \leq l'})} =: \frac{1}{d_1}. \tag{5.16}$$

We will use Theorem 4.2 to obtain lower and upper bounds on d_1.

5.5.1. *Lower bound on d_1.* By the definition of r_j, we have

$$d_1 \geq \sqrt{1 + d(DB_1, \mathrm{span}(DB_j)_{2 \leq j \leq l'})^2}.$$

We apply Theorem 4.2 in dimension $n - l$ instead of n, and with

$$k = l' - 1, \quad k_0 = (1 - \delta)l', \quad k_1 = (1 + \delta)l'.$$

Recall here that in (5.7) we selected $\delta = c/\log n$. Note that by construction (5.14), the vectors B_i do not contain the diagonal elements of A, and so their entries have mean zero as required in Theorem 4.2. Applying part (i) of that theorem, we obtain with probability at least $1 - 2\exp(-c\delta l')$ that

$$d_1 \geq \sqrt{1 + c\bar{S}_{k_1}^2} \geq \frac{1}{2}(1 + c\bar{S}_{k_1}). \tag{5.17}$$

5.5.2. *Upper bound on d_1.* Now we apply part (ii) of Theorem 4.2. This time we shall use a sharper bound stated in Remark 4.4. It yields that with probability at least $1 - 2l'\exp(-c\delta l')$, the following holds. There exists $a = (a_2, \ldots, a_{l'})$ such that

$$\left\| DB_1 - \sum_{j=2}^{l'} a_j DB_j \right\|_2 \leq CM(\bar{S}_{k_0} + \sqrt{k}\, s_{k_0+1}), \tag{5.18}$$

$$\|a\|_2 \leq M, \quad \text{where } M = \frac{Ck\sqrt{k_0}}{k - k_0} \leq 2\sqrt{l'}/\delta. \tag{5.19}$$

We can simplify (5.18). Using (5.2) and monotonicity, we have

$$ks_{k_0+1}^2 \leq 2ks_{k_1}^2 = \frac{2k}{k_1 - k_0}(k_1 - k_0)s_{k_1}^2 \leq \frac{2k}{k_1 - k_0}\bar{S}_{k_0}^2 \leq \frac{2}{\delta}\bar{S}_{k_0}^2,$$

thus again using (5.2), we have

$$(\bar{S}_{k_0} + \sqrt{k}\, s_{k_0+1})^2 \leq 2(\bar{S}_{k_0}^2 + ks_{k_0+1}^2) \leq \frac{6}{\delta}\bar{S}_{k_0}^2 \leq \frac{30}{\delta}\bar{S}_{k_1}^2.$$

Hence (5.18) yields

$$\left\| DB_1 - \sum_{j=2}^{l'} a_j DB_j \right\|_2 \leq \frac{CM}{\sqrt{\delta}}\bar{S}_{k_1}.$$

Recall that this holds with probability at least $1 - 2l'\exp(-c\delta l')$. On this event, by the construction of r_i and using the bound on a in (5.19), we have

$$d_1 = d(r_1, \mathrm{span}(r_j)_{2 \leq j \leq l'}) \leq \left\| r_1 - \sum_{j=2}^{l'} a_j r_j \right\|_2$$

$$= 1 + \|a\|_2 + \left\| DB_1 - \sum_{j=2}^{l'} a_j DB_j \right\|_2 \leq 2M\left(1 + \frac{C}{\sqrt{\delta}}\bar{S}_{k_1}\right). \tag{5.20}$$

5.5.3. *Completion of the proof of Theorem 5.1.* Combining the events (5.20) and (5.17), we have shown the following. With probability at least $1 - 4l' \exp(-c\delta l')$, the following two-sided estimate holds:

$$\frac{1}{2}(1 + c\bar{S}_{k_1}) \le d_1 \le 2M\Big(1 + \frac{C}{\sqrt{\delta}}\bar{S}_{k_1}\Big).$$

A similar statement can be proved for general d_i, $1 \le i \le l'$. By intersecting these events, we obtain that with probability at least $1 - 4(l')^2 \exp(-c\delta l')$, all such bounds for d_i hold simultaneously. Suppose this indeed occurs. Then by (5.16), we have

$$\frac{\|Pe_i\|_2}{\|Pe_j\|_2} = \frac{d_j}{d_i} \le \frac{4M\big(1 + (C/\sqrt{\delta})\bar{S}_{k_1}\big)}{1 + c\bar{S}_{k_1}} \le \frac{C_1}{\sqrt{\delta}}M \quad \forall 1 \le i, j \le l'. \tag{5.21}$$

We have calculated the conditional probability of (5.21); recall that we conditioned on $\bar{A}$ which satisfies the event (5.6), which itself holds with probability $1 - 2n\exp(-cl)$. Thus the unconditional probability of the event (5.21) is at least $1 - 2n\exp(-cl) - C_1(l')^2 \exp(-c\delta l')$. Recalling that $l/2 \le l \le n/4$ and $\delta = c/\log n$, and simplifying this expression, we arrive at the probability bound claimed in Theorem 5.1. Since $M \le 2\sqrt{l}/\delta$ according to (5.19), the estimate (5.21) yields the first part of (iii) in Theorem 5.1. The second part, stating that $Pe_i = 0$ for $l' < i \le l$, was already noted in (iii) or Proposition 5.3. Thus Theorem 5.1 is proved. $\square$

6. PROOF OF THEOREM 1.6 AND COROLLARY 1.5

Let G be a random matrix from Theorem 1.6. We shall apply Proposition 3.1 for

$$A = G - zI_n, \quad |z| \le K_1\sqrt{n}, \tag{6.1}$$

where $z \in \mathbb{C}$ is a fixed number for now, and K_1 is a parameter to be chosen later. The power of Proposition 3.1 relies on the existence of a test projection P for which the balancing event $\mathcal{B}_{\alpha,\kappa}$ is likely. We are going to validate this condition using the test projection constructed in Theorem 5.1.

Proposition 6.1 (Balancing event is likely)**.** *Let* $\alpha = c/(l \log^{3/2} n)$ *and* $\kappa = c$. *Then, for every fixed* $(j_0, J_0) \in \Lambda$, *one can find a test projection as required in Proposition 3.1. Moreover,*

$$\mathbb{P}_A \{\mathcal{B}_{\alpha,\kappa}\} \ge 1 - 2n^2 \exp(-cl/\log n).$$

Here $c = c(K, K_1) > 0$.

Proof. Without loss of generality, we assume that $j_0 = 1$ and $J_0 = \{2, \ldots, n\}$. We apply Theorem 5.1, and choose $l' \in [l/2, l]$ and P determined by $\{A_j\}_{j>l}$ guaranteed by that theorem. The test projeciton P automatically satisfies the conditions of Proposition 3.1. Moreover, with probability at least $1 - 2n^2 \exp(-cl/\log n)$, one has

$$\|Pe_j\|_2 \le C\sqrt{l} \log^{3/2} n \cdot \|Pe_1\|_2 \quad \text{for } 2 \le j \le l. \tag{6.2}$$

Let us condition on $\{A_j\}_{j>l}$ for which the event (6.2) holds; this fixes l' and P but leaves $\{A_j\}_{j\le l}$ random as before.

The definition (3.2) of balancing event $\mathcal{B}_{\alpha,\kappa}$ requires us to estimate the norms of

$$PA_1 = PG_1 - zPe_1 \quad \text{and} \quad PA_{J_0} = PG_{J_0} - zP_{J_0}.$$

For PA_1, we use the small ball probability estimate, Theorem 2.3(ii). Recall that $\|P\|_{\mathrm{HS}}^2 = \mathrm{tr}(PP^*) = \mathrm{tr}(I_{l'}) = l' \geq l/2$ and $\|P\| = 1$. It follows that with probability at least $1 - 2\exp(-cl)$, we have

$$\|PA_1\|_2 \geq c(\sqrt{l} + |z|\|Pe_1\|_2). \tag{6.3}$$

Next, we estimate

$$\|PA_{J_0}\| \leq \|PG_{J_0}\| + |z|\|P_{J_0}\|. \tag{6.4}$$

For the $l' \times (l-1)$ matrix PG_{J_0}, Theorem 2.4 (see Remark 2.5) implies that with probability at least $1 - 2\exp(-l)$ one has $\|PG_{J_0}\| \leq C\sqrt{l}$. Further, (6.2) allows us to bound $\|P_{J_0}\| \leq \|P_{J_0}\|_{\mathrm{HS}} \leq \sqrt{l} \max_{2 \leq j \leq l} \|Pe_j\|_2 \leq Cl \log^{3/2} n \cdot \|Pe_1\|_2$. Thus (6.4) yields

$$\|PA_{J_0}\| \leq C\sqrt{l} + Cl \log^{3/2} n \cdot |z|\|Pe_1\|_2. \tag{6.5}$$

Hence, estimates (6.3) and (6.5) hold simultaneously with probability at least $1 - 4\exp(-cl)$. Recall that this concerns conditional probability, where we conditioned on the event (6.2), which itself holds with probability at least $1 - 2n^2\exp(-cl/\log n)$. Therefore, estimates (6.3) and (6.5) hold simultaneously with (unconditional) probability at least $1 - 4\exp(-cl) - 2n^2\exp(-cl/\log n) \geq 1 - 6n^2\exp(-cl/\log n)$. Together they yield

$$\|PA_1\|_2 \geq \alpha\|PA_{J_0}\| \quad \text{where } \alpha = c/(l \log^{3/2} n).$$

This is the first part of the event $\mathcal{B}_{\alpha,\kappa}$. Finally, (6.3) implies that $\|PA_1\|_2 \geq c\sqrt{l}$, which is the second part of the event $\mathcal{B}_{\alpha,\kappa}$ for $\kappa = c$. The proof is complete. $\qquad\square$

Substituting the conclusion of Proposition 6.1 into Proposition 3.1, we obtain:

Proposition 6.2. *Let $0 < w < l$ and $W = Cl \log^{3/2} n$. Then*

$$\mathbb{P}\left\{\mathcal{L}_{W,w}\right\} \leq 4n^3 \exp(-cl/\log n).$$

Here $C = C(K, K_1)$, $c = c(K, K_1) > 0$.

From this we can readily deduce a slightly stronger version of Theorem 1.6.

Corollary 6.3. *Consider a random matrix G as in Theorem 1.6. Let $0 \leq s \leq n$, $s + 2 \leq l \leq n/4$ and $W = Cl \log^{3/2} n$. Then the event*

$$\mathcal{L}_W := \left\{\exists \left(\frac{s}{\sqrt{n}}\right)\text{-approximate eigenvector } v \text{ of } G \text{ with } \|v\|_2 = 1, \ \|v\|_\infty > W\sqrt{\frac{l}{n}}\right\}$$

is unlikely:

$$\mathbb{P}\left\{\mathcal{L}_W\right\} \leq Cn^5 \exp(-cl/\log n).$$

Here $C = C(K)$, $c = c(K) > 0$.

Proof. Recall that G is nicely bounded with high probability. Indeed, Theorem 2.4 (see Remark 2.5) states that the event

$$\mathcal{E}_{\mathrm{norm}} := \left\{\|G\| \leq C_1\sqrt{n}\right\} \quad \text{is likely: } \quad \mathbb{P}\left\{\mathcal{E}_{\mathrm{norm}}\right\} \leq 1 - 2\exp(-cn). \tag{6.6}$$

Assume that $\mathcal{E}_{\mathrm{norm}}$ holds. Then all $(s/\sqrt{n})$-approximate eigenvalues of G are contained in the complex disc centered at the origin and with radius $\|G\| + s/\sqrt{n} \leq 2C_1\sqrt{n}$. Let $\{z_1, \ldots, z_N\}$ be a $(1/\sqrt{n})$-net of this disc such that $N \leq C_2 n^2$.

Assume $\mathcal{L}_W$ holds, so there exists an $(s/\sqrt{n})$-approximate eigenvector v of G such that $\|v\|_2 = 1$ and $\|v\|_\infty > W\sqrt{l/n}$. Choose a point z_i in the net closest to z, so $|z - z_i| \leq 1/\sqrt{n}$. Then

$$\|(G - z_i I_n)v\|_2 \leq \|(G - z I_n)v\|_2 + |z - z_i| \leq \frac{s+1}{\sqrt{n}}.$$

This argument shows that $\mathcal{L}_W \cap \mathcal{E}_{\text{norm}} \subseteq \bigcup_{i=1}^{N} \mathcal{L}_W^{(i)}$, where

$$\mathcal{L}_W^{(i)} = \left\{ \exists v \in S^{n-1} : \|v\|_\infty > W\sqrt{\frac{l}{n}} \text{ and } \|(G - z_i I_n)v\|_2 \leq \frac{s+1}{\sqrt{n}} \right\}.$$

Recall that the probability of $\mathcal{E}_{\text{norm}}$ is estimated in (6.6), and the probabilities of the events $\mathcal{L}_W^{(i)}$ can be bounded using Proposition 6.2 with $w = s + 1$. (Our assumption that $l \geq s+2$ enforces the bound $w < l$ that is needed in Proposition 6.2.) It follows that

$$\mathbb{P}\{\mathcal{L}_W\} \leq \mathbb{P}\{\mathcal{E}_{\text{norm}}^c\} + \sum_{i=1}^{N} \mathbb{P}\left\{\mathcal{L}_W^{(i)}\right\} \leq 2\exp(-cn) + C_2 n^2 \cdot 4n^3 \exp(-cl/\log n).$$

Simplifying this bound we complete the proof. $\qquad\square$

Proof of Theorem 1.6. We are going to apply Corollary 6.3 for $l = Ct(s+1)\log^2 n$. This is possible as long as $s \leq n$ and $t(s+1) < cn/\log^2 n$, since the latter restriction enforces the bound $l \leq n/4$. In this regime, the conclusion of Theorem 1.6 follows directly from Corollary 6.3.

In the remaining case, where either $s \geq n$ or $t(\varepsilon\sqrt{n} + 1) > cn/\log^2 n$, the right hand side of (1.1) is greater than $\|v\|_2$ for an appropriate choice of the constant C. Thus, in this case, the bound (1.1) holds trivially since one always has $\|v\|_\infty \leq \|v\|_2$. Theorem 1.6 is proved. $\qquad\square$

6.1. **Deduction of Corollary 1.5.**

Using a standard truncation argument, we will now deduce Corollary 1.5 for general exponential tail decay. We will first prove the following relaxation of Proposition 6.2.

Proposition 6.4. *Let G be an $n \times n$ real random matrix whose entries G_{ij} are independent random variables satisfying $\mathbb{E}\,G_{ij} = 0$, $\mathbb{E}\,G_{ij}^2 \geq 1$ and $\|G_{ij}\|_{\psi_\alpha} \leq M$. Let $z \in \mathbb{C}$, $0 < w < l - 1$, and $t \geq 2$. Set $W = Clt^\beta \log^\gamma n$, and consider the event $\mathcal{L}_{W,w}$ defined as in (3.1) for the matrix $A = G - z I_n$. Then*

$$\mathbb{P}\{\mathcal{L}_{W,w}\} \leq 4n^3 \exp(-cl/\log n) + n^{-t}.$$

Here $\beta, \gamma, C, c > 0$ depend only on α and M.

Proof (sketch). Set $K := (Ct \log n)^{1/\alpha}$, and let $\tilde{G}$ be the matrix with entries $\tilde{G}_{ij} = G_{ij} \mathbf{1}_{|G_{ij}| \leq K}$. Since $\mathbb{E}\,G_{ij} = 0$, the bound on $\|G_{ij}\|_{\psi_\alpha}$ yields $|\mathbb{E}\,\tilde{G}_{ij}| \leq \exp(-cK^\alpha)$. Hence

$$\|\mathbb{E}\,\tilde{G}\| \leq \|\mathbb{E}\,\tilde{G}\|_{\text{HS}} \leq n\exp(-cK^\alpha) \leq n^{-1/2}.$$

Then the event $\mathcal{L}_{W,w}$ for the matrix $A = G - z I_n$ implies the event $\mathcal{L}_{W,w+1}$ for the matrix $\tilde{A} := G - \mathbb{E}\,\tilde{G} - z I_n$. It remains to bound the probability of the latter event. If the constant C in the definition of K is sufficiently large, then with probability at least $1 - n^{-t}$ we have $\tilde{G} = G$ and thus $\tilde{A} = \tilde{G} - \mathbb{E}\,\tilde{G} - z I_n$. Conditioned on this

likely event, the entries $\tilde{G} - \mathbb{E}\tilde{G}$ are independent, bounded by K, have zero means and variances at least $1/2$. Therefore, we can apply Proposition 6.2 for the matrix $\tilde{A}$ and thus bound the probability of $\mathcal{L}_{W,w+1}$ for $\tilde{A}$, as required.   □

Corollary 1.5 follows from Proposition 6.4 in the same way as Corollary 6.3 followed from Proposition 6.2. The only minor difference is that one would put a coarser bound the norm of G. For example, one can use that $\|G\| \leq \|G\|_{\mathrm{HS}} \leq n \cdot \max_{i,j \leq n} |G_{ij}| \leq n \cdot Ms$ with probability at least $1 - 2n^2 \exp(-cs^\alpha)$, for any $s > 0$. This, however, would only affect the bound on the covering number N in Corollary 6.3, changing the estimate in this Corollary to

$$\mathbb{P}\{\mathcal{L}_W\} \leq C(Ms)^2 n^6 \exp(-cl/\log n).$$

We omit the details.   □

APPENDIX A. INVERTIBILITY OF RANDOM MATRICES

Our delocalization method relied on estimates of the smallest singular values of rectangular random matrices. The method works well provided one has access to estimates that are polynomial in the dimension of the matrix (which sometimes was of order n, and other times of order $l \sim \log^2 n$), and provided the probability of having these estimates is, say, at least $1 - n^{-10}$.

In the recent years, significantly sharper bounds were proved than those required in our delocalization method, see survey [27]. We chose to include weaker bounds in this appendix for two reasons. First, they hold in somewhat more generality than those recorded in the literature, and also their proofs are significantly simpler.

Theorem A.1 (Rectangular matrices). *Let $N \geq n$, and let $A = D + G$ where D is an arbitrary $N \times n$ fixed matrix and G is an $N \times n$ random matrix with independent entries satisfying Assumption 2.1. Then*

$$\mathbb{P}\left\{ s_n(A) < c\sqrt{\frac{N-n}{n}} \right\} \leq 2n \exp(-c(N-n)). \tag{A.1}$$

Here $c = c(K) > 0$.

Proof. Using the negative second moment identity (see [34] Lemma A.4]), we have

$$s_n(A)^{-2} \leq \sum_{i=1}^n s_i(A)^{-2} = \sum_{i=1}^n d(A_i, E_i)^{-2} \tag{A.2}$$

where $A_i = D_i + G_i$ denote the columns of A and $E_i = \mathrm{span}(A_j)_{j \leq n, j \neq i}$. For fixed i, note that $d(A_i, E_i) = \|P_{E_i^\perp} A_i\|_2$. Since A_i is independent of E_i, we can apply the small ball probability bound, Theorem 2.3(ii). Using that $\|P_{E_i^\perp}\|_{\mathrm{HS}}^2 = \dim(E_i^\perp) \geq N - n$ and $\|P_{E_i^\perp}\| = 1$, we obtain

$$\mathbb{P}\left\{ d(A, E_i) < c\sqrt{N-n} \right\} \leq 2\exp(-c(N-n)).$$

Union bound yields that with probability at least $1 - 2n \exp(-c(N-n))$, we have $d(A_i, E_i) \geq c\sqrt{N-n}$ for all $i \leq n$. Plugging this into (A.2), we conclude that with the same probability, $s_n(A)^{-2} \leq c^{-2}n/(N-n)$. This completes the proof.   □

Corollary A.2 (Intermediate singular values). *Let $A = D + G$ where D is an arbitrary $N \times M$ fixed matrix and G is an $N \times M$ random matrix with independent entries satisfying Assumption 2.1. Then all singular values $s_n(A)$ for $1 \leq n \leq \min(N, M)$ satisfy the estimate (A.1) with $c = c(K) > 0$.*

Proof. Recall that $s_n(A) \geq s_n(A_0)$ where A_0 is formed by the first n columns of A. The conclusion follows from Theorem A.1 applied to A_0. $\qquad\square$

Theorem A.3 (Products of random and deterministic matrices). *Let $k, m, n \in \mathbb{N}$, $m \leq \min(k, n)$. Let P be a fixed $m \times n$ matrix such that $PP^{\mathsf{T}} = I_m$, and G be an $n \times k$ random matrix with independent entries that satisfy Assumption 2.1. Then*

$$\mathbb{P}\left\{ s_m(PG) < c\,\frac{k - m}{k} \right\} \leq 2k \exp(-c(k - m)).$$

Here $c = c(K)$.

Let us explain the idea of the proof of Theorem A.3. We need a lower bound for

$$\|(PG)^* x\|_2^2 = \sum_{i=1}^{k} \langle PG_i, x \rangle^2,$$

where G_i denote the columns of G. The bound has to be uniform over $x \in S^{m-1}$. Let $m = (1 - \delta)k$ and set $m_0 = (1 - \rho)m$ for a suitably chosen $\rho \ll \delta$.

First, we claim that if $x \in \mathrm{span}(PG_i)_{i \leq m_0} =: E$ then $\sum_{i=1}^{m_0} \langle PG_i, x \rangle^2 \gtrsim \|x\|_2^2$. This is equivalent to controlling the smallest singular value of the $m \times m_0$ random matrix with independent columns PG_i, $i = 1, \ldots, m_0$. Since $m \geq m_0$, this can be achieved with a minor variant of Theorem A.1. The same argument works for general $x \in \mathbb{C}^m$ provided x is not almost orthogonal onto E.

The vectors x that lie near the subspace $E^{\perp}$, which has dimension $m - m_0 = \rho m$, can be controlled by the remaining $k - m_0$ vectors PG_i, since $k - m_0 \gg m - m_0$. Indeed, this is equivalent to controlling the smallest singular value of a $(m - m_0) \times (k - m_0)$ random matrix whose columns are QG_i, where Q projects onto $E^{\perp}$. This is a version of Theorem A.3 for very fat matrices, and it can be proved in a standard way by using ε-nets.

Now we proceed to the formal argument.

Lemma A.4 (Slightly fat matrices). *Let $m_0 \leq m$. Consider the $m \times m_0$ matrix T_0 formed by the first m_0 columns of matrix $T = PG$. Then*

$$\mathbb{P}\left\{ s_{m_0}(T_0) < c\sqrt{\frac{m - m_0}{m_0}} \right\} \leq 2m_0 \exp(-c(m - m_0)).$$

This is a minor variant of Theorem A.1; its proof is very similar and is omitted. $\quad\square$

Lemma A.5 (Very tall matrices). *There exist $C = C(K)$, $c = c(K) > 0$ such that the following holds. Consider the same situation as in Theorem A.3, except that we assume that $k \geq Cm$. Then*

$$\mathbb{P}\left\{ s_m(PG) < c\sqrt{k} \right\} \leq \exp(-ck).$$

Lemma A.5 is a minor variation of [38, Theorem 5.39] for $k \geq Cm$ independent sub-gaussian columns, and it can be proved in a similar way (using a standard concentration and covering argument). $\qquad\square$

Proof of Theorem A.3. Denote $T := PG$; our goal is to bound below the quantity

$$s_m(T) = s_m(T^*) = \inf_{x \in S^{m-1}} \|T^* x\|_2^2.$$

Let $\varepsilon, \rho \in (0, 1/2)$ be parameters, and set $m_0 = (1 - \rho)m$. We decompose

$$T = [T_0 \ \bar{T}]$$

where T_0 is the $m \times m_0$ matrix that consists of the first m_0 columns of T, and $\bar{T}$ is the $(k-m_0) \times m$ matrix that consists of the last $k-m_0$ columns of T. Let $x \in S^{m-1}$. Then

$$\|T^* x\|_2^2 = \|T_0^* x\|_2^2 + \|\bar{T}^* x\|_2^2.$$

Denote

$$E = \text{Im}(T_0) = \text{span}(PG_i)_{i \le m_0}.$$

Assume that $s_{m_0}(T_0) > 0$ (which will be seen to be a likely event), so $\dim(E) = m_0$.

The argument now splits according to the position of x relative to E. Assume first that $\|P_E x\|_2 \ge \varepsilon$. Since $\text{rank}(T_0) = m_0$, using Lemma 2.2(i) we have

$$\|T^* x\|_2 \ge \|T_0^* x\|_2 \ge s_{m_0}(T_0^*)\|P_E x\|_2 \ge s_{m_0}(T_0)\varepsilon.$$

We will later apply Lemma A.4 to bound $s_{m_0}(T_0)$ below.

Consider now the opposite case, where $\|P_E x\|_2 < \varepsilon$. There exists $y \in E^\perp$ such that $\|x - y\|_2 \le \varepsilon$, and in particular $\|y\|_2 \ge \|x\|_2 - \varepsilon \ge 1 - \varepsilon > 1/2$. Thus

$$\|T^* x\|_2 \ge \|\bar{T}^* x\|_2 \ge \|\bar{T}^* y\|_2 - \|\bar{T}^*\|\varepsilon. \tag{A.3}$$

We represent $\bar{T} = P\bar{G}$, where $\bar{G}$ is the $n \times (k - m_0)$ matrix that contains the last $k - m_0$ columns of G. Consider an $m \times (m - m_0)$ matrix Q^* which is an isometric embedding of $\ell_2^{m-m_0}$ into ℓ_2^m, and such that $\text{Im}(Q^*) = E^\perp$. Then there exists

$$z \in \mathbb{C}^{m-m_0} \text{ such that } y = Q^* z, \quad \|z\|_2 = \|y\|_2 \ge 1/2.$$

Therefore

$$\|\bar{T}^* y\|_2 = \|\bar{G}^* P^* Q^* z\|_2.$$

Since both $Q^* : \mathbb{C}^{m-m_0} \to \mathbb{C}^m$ and $P^* : \mathbb{C}^m \to \mathbb{C}^n$ are isometric embeddings, $R^* := P^* Q^* : \mathbb{C}^{m-m_0} \to \mathbb{C}^n$ is an isometric embedding, too. Thus R is a $(m-m_0) \times n$ matrix which satisfies $RR^* = I_{m-m_0}$. Hence

$$\|\bar{T}^* y\|_2 = \|\bar{B}^* z\|_2, \text{ where } \bar{B} := R\bar{G}$$

is an $(m-m_0) \times (k-m_0)$ matrix. Since $\|z\|_2 \ge 1/2$, we have $\|\bar{T}^* y\|_2 \ge \frac{1}{2} s_{m-m_0}(B)$, which together with (A.3) yields

$$\|T^* x\|_2 \ge \frac{1}{2} s_{m-m_0}(\bar{B}) - \|\bar{T}\|\varepsilon.$$

A bit later, we will use Lemma A.5 to bound $s_{m-m_0}(\bar{B})$ below.

Putting the two cases together, we have shown that

$$s_m(T) \ge \min_{x \in S^{m-1}} \|T^* x\|_2 \ge \min\left\{ s_{m_0}(T_0)\varepsilon, \frac{1}{2} s_{m-m_0}(\bar{B}) - \|\bar{T}\|\varepsilon \right\}. \tag{A.4}$$

It remains to estimate $s_{m_0}(T_0)$, $s_{m-m_0}(\bar{B})$ and $\|\bar{T}\|$.

Since $m_0 = (1 - \rho)m$ and $\rho \in (0, 1/2)$, Lemma A.4 yields that with probability at least $1 - 2m \exp(-c\rho m)$, we have

$$s_{m_0}(T_0) \ge c\sqrt{\rho}.$$

Next, we use Lemma A.5 for the $(m - m_0) \times (k - m_0)$ matrix $\bar{B} = R\bar{G}$. Let $\delta \in (0, 1)$ be such that $m = (1 - \delta)k$. Since $m_0 = (1 - \rho)m$, by choosing $\rho = c_0\delta$ with a suitable $c_0 > 0$ we can achieve that $k - m_0 \geq C(m - m_0)$ to satisfy the dimension requirement in Lemma A.5. Then, with probability at least $1 - 2\exp(-c\delta k)$ we have

$$s_{m-m_0}(\bar{B}) \geq c\sqrt{\delta k}.$$

Further, by Theorem 2.4, with probability at least $1 - 2\exp(-k)$ we have

$$\|\bar{T}\| \leq \|T\| \leq C\sqrt{k}.$$

Putting all these estimates in (A.4), we find that with probability at least $1 - 2m\exp(-c\rho m) - 2\exp(-c\delta k) - 2\exp(-k)$, one has

$$s_m(T) \geq \min\left\{c\sqrt{\rho}\,\varepsilon, \frac{1}{2}c\sqrt{\delta k} - C\sqrt{k}\,\varepsilon\right\}.$$

Now we choose $\varepsilon = c_1\sqrt{\delta}$ with a suitable $c_1 > 0$, and recall that we have chosen $\rho = c_0\delta$. We conclude that $s_m(T) \geq c\min\{\delta, \sqrt{\delta k}\} = c\delta$. Since $m = (1 - \delta)k$, the proof of Theorem A.3 is complete. $\qquad\square$

References

[1] F. Benaych-Georges, S. Péché, *Localization and delocalization for heavy tailed band matrices*, Annales Inst. H. Poincaré, to appear.

[2] A. Bloemendal, L Erd?s, A. Knowles, H.T. Yau, J. Yin, *Isotropic local laws for sample covariance and generalized Wigner matrices*, Electron. J. Probab. 19 (2014), no. 33, 53 pp.

[3] C. Bordenave, A. Guionnet, *Localization and delocalization of eigenvectors for heavy-tailed random matrices*, Probability Theory and Related Fields 157 (2013), 885–953.

[4] C. Cacciapuoti, A. Maltsev, B. Schlein, *Local Marchenko-Pastur law at the hard edge of sample covariance matrices*, Journal of Mathematical Physics, to appear.

[5] L. Erdös, *Universality for random matrices and log-gases*, In: Current Developments in Mathematics 2012, Ed. D. Jerison, M. Kisin, T. Mrowka. R. Stanley, H.-T. Yau, S.-T. Yau, International Press, 59–132 (2013).

[6] L. Erdös, A. Knowles, *Quantum diffusion and eigenfunction delocalization in a random band matrix model*, Commun. Math. Phys. 303 (2011), 509–554.

[7] L. Erdös, A. Knowles, *Quantum diffusion and delocalization for band matrices with general distribution*, Annales Inst. H. Poincaré 12 (2011), 1227–1319.

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

[9] L. Erdös, A. Knowles, H.-T. Yau, J. Yin, *Spectral statistics of Erdös-Rényi graphs II: eigenvalue spacing and the extreme eigenvalues*, Comm. Math. Phys. 314 (2012), no. 3, 587–640.

[10] L. Erdös, A. Knowles, H.-T. Yau, J. Yin, *Delocalization and diffusion profile for random band matrices*, Comm. Math. Phys., to appear.

[11] L. Erdös, B. Schlein, H.-T. Yau, *Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices*, Ann. Probab. 37 (2009), 815–852.

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

[13] L. Erdös, B. Schlein, H.-T. Yau, *Wegner estimate and level repulsion for Wigner random matrices*, IMRN 2010 (2009), 436–479.

[14] L. Erdös, H.-T. Yau, *Universality of local spectral statistics of random matrices*, Bulletin of the AMS 49 (2012), 377–414.

[15] L. Erdös, H.-T. Yau, J. Yin, *Rigidity of eigenvalues of generalized Wigner matrices*, Advances in Mathematics 229 (2012), 1435–1515.

[16] L. Erdös, H.-T. Yau, J. Yin, *Bulk universality for generalized Wigner matrices*, Probability Theory and Related Fields 154 (2012), 341–407

[17] J. Fischmann, W. Bruzda, B. A. Khoruzhenko, H-J. Sommers, K. Zyczkowsk,i *Induced Ginibre ensemble of random matrices and quantum operations*, J. Phys. A, 45:075203, 2012.

[18] P. J. Forrester, T. Nagao, *Eigenvalue statistics of the real Ginibre ensemble*, Phys. Rev. Lett. 99 (2007).

[19] Y. V. Fyodorov, H.-J. Sommers. *Random matrices close to Hermitian or unitary: overview of methods and results*, J. Phys. A, 36(12):33033347, 2003.

[20] G. H. Golub, C. F. Van Loan, *Matrix computations* (3rd ed.). Baltimore: Johns Hopkins. 1996.

[21] R. A. Janik, W. Norenberg, M. A. Nowak, G. Papp, I. Zahed, *Correlations of eigenvectors for non-Hermitian random-matrix models*, Physical Review E, 60(3):2699–2705, 1999.

[22] E. Kanzieper, G. Akemann, *Statistics of real eigenvalues in Ginibres ensemble of random real matrices*, Physical Review Letters, 95:230201, 2005.

[23] A. Knowles, J. Yin, *The isotropic semicircle law and deformation of Wigner matrices*, Comm. Pure Appl. Math. 66 (2013), no. 11, 1663111750.

[24] R. Latala, P. Mankiewicz, K. Oleszkiewicz, N. Tomczak-Jaegermann, *Banach-Mazur distances and projections on random subgaussian polytopes*, Discrete Comput. Geom. 38 (2007), 29–50.

[25] M. Ledoux, *The concentration of measure phenomenon.* Mathematical Surveys and Monographs, 89. Providence: American Mathematical Society, 2005.

[26] B. Mehlig and J. T. Chalker, *Statistical properties of eigenvectors in non-Hermitian Gaussian random matrix ensembles*, J. Math. Phys. 41, 3233 (2000).

[27] M. Rudelson, R. Vershynin, *Non-asymptotic theory of random matrices: extreme singular values.* Proceedings of the International Congress of Mathematicians. Volume III, 1576–1602, Hindustan Book Agency, New Delhi, 2010.

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

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

[30] M. Rudelson, R. Vershynin, *Hanson-Wright inequality and sub-gaussian concentration*, Electronic Communications in Probability 18 (2013), 1–9.

[31] M. Talagrand, *Concentration of measure and isoperimetric inequalities in product spaces*, IHES Publ. Math. No. 81 (1995), 73–205.

[32] T. Tao, *Topics in random matrix theory.* American Mathematical Society, 2012.

[33] T. Tao, V. Vu, *On random ±1 matrices: singularity and determinant*, Random Structures and Algorithms 28 (2006), 1–23.

[34] T. Tao, V. Vu, *Random matrices: universality of ESDs and the circular law*, with an appendix by M. Krishnapur. Ann. Probab. 38 (2010), 2023–2065.

[35] T. Tao, V. Vu, *Random matrices: Universal properties of eigenvectors*, Random Matrices Theory Appl. 1 (2012), 1150001, 27 pp.

[36] M. Timme, F. Wolf, and T. Geisel, *Topological Speed Limits to Network Synchronization*, Physical Review letters, 92 (2004), 74–101.

[37] L. V. Tran, V. Vu, K. Wang, *Sparse random graphs: eigenvalues and eigenvectors*, Random Structures Algorithms 42 (2013), 110–134.

[38] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices.* Compressed sensing, 210–268, Cambridge Univ. Press, Cambridge, 2012.

[39] V. Vu, K. Wang, *Random weighted projections, random quadratic forms and random eigenvectors*, Random Structures and Algorithms, to appear.

Department of Mathematics, University of Michigan, 530 Church St., Ann Arbor, MI 48109, U.S.A.

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