RANDOM MATRICES: OVERCROWDING ESTIMATES FOR THE SPECTRUM

HOI H. NGUYEN

ABSTRACT. We address overcrowding estimates for the singular values of random iid matrices, as well as for the eigenvalues of random Wigner matrices. We show evidence of long range separation under arbitrary perturbation even in matrices of discrete entry distributions. In many cases our method yields nearly optimal bounds.

1. INTRODUCTION

1.1. Random iid matrices with subgaussian tails.

Consider a random matrix $M = (m_{ij})_{1 \le i, j \le n}$, where m_{ij} are iid copies of a random variable ξ of mean zero and variance one. Let $\sigma_n \le \cdots \le \sigma_1$ be the singular values of M.

An important problem with practical applications is to bound the condition number of M. As the asymptotic behavior of the largest singular value σ_1 is well understood under natural assumption on ξ, the main problem is to study the lower bound of the least singular value σ_n. This problem was first raised by Goldstine and von Neumann [10] well back in the 1940s, with connection to their investigation of the complexity of inverting a matrix.

To answer Goldstine and von Neumman's question, Edelman [7] computed the distribution of the least singular value of Ginibre matrix (where ξ is standard gaussian). He showed that for all fixed $\varepsilon > 0$

$$\mathbf{P}(\sigma_n \le \varepsilon n^{-1/2}) = \int_0^{\varepsilon^2} \frac{1 + \sqrt{x}}{2\sqrt{x}} e^{-(x/2 + \sqrt{x})} \, dx + o(1) = \varepsilon - \frac{1}{3}\varepsilon^3 + O(\varepsilon^4) + o(1).$$

Note that the same asymptotic continues to hold for any $\varepsilon > 0$ which can go to zero with n (see also [22])

$$\mathbf{P}(\sigma_n \le \varepsilon n^{-1/2}) \le \varepsilon. \tag{1}$$

For other singular values of Ginibre matrices, an elegant result by Szarek [24] shows that the σ_{n-k+1} are separated away from zero with an extremely fast rate.

Theorem 1.2. *Assume that ξ is standard gaussian, then there exist absolute constants C_1, C_2 such that for all $\varepsilon > 0$, and all $1 \le k \le n$*

$$(\frac{C_1}{k}\varepsilon)^{k^2} \le \mathbf{P}(\sigma_{n-k+1} \le \frac{\varepsilon}{\sqrt{n}}) \le (\frac{C_2}{k}\varepsilon)^{k^2}.$$

In what follows ε is always bounded by $O(1)$. Motivated by the universality phenomenon in random matrix theory, we expect similar repulsion bounds for general random matrix ensembles. More specifically, we will

---

assume ξ to have *mean zero, variance one*, and subgaussian distribution. In other words, there exists $B > 0$ such that

$$\mathbf{P}(|\xi| > t) \leq 2\exp(-t^2/B^2) \text{ for all } t. \tag{2}$$

The smallest of such B is called the subgaussian moment of ξ. A representative example of our study, on the opposite side of being gaussian, is the Bernoulli (Radamacher) random variable which takes value ± 1 with probability $1/2$.

When $k = 1$, Rudelson and Vershynin [18] extended (1) to more general iid random matrices.

Theorem 1.3. *Let $M = (m_{ij})_{1 \leq i,j \leq n}$ be a random matrix where m_{ij} are iid copies of a subgaussian random variable ξ as in (2). Then for any $\varepsilon > 0$ that might depend on n we have*

$$\mathbf{P}(\sigma_n \leq \frac{\varepsilon}{\sqrt{n}}) \leq C\varepsilon + \exp(-cn),$$

where C and c depend only on the subgaussian moment of ξ.

Furthermore, it was shown by Tao and Vu [27] that the statistics of $\sqrt{n}\sigma_n$ is universal. Thus our understanding in the case $k = 1$ is nearly complete.

In this note we address the overcrowding direction by investigating Theorem 1.2 under various settings.

Theorem 1.4 (the iid case). *Let $M = (m_{ij})_{1 \leq i,j \leq n}$ be a random matrix where m_{ij} are iid copies of a subgaussian random variable ξ as in (2). For any $k \geq 1$ there exist a constant C_k depending on k and a constant c depending only on the subgaussian moment of ξ such that for any $\varepsilon > 0$,*

$$\mathbf{P}(\sigma_{n-k+1} \leq \frac{\varepsilon}{\sqrt{n}}) \leq C_k \varepsilon^{k^2} + \exp(-cn). \tag{3}$$

Furthermore, for any $0 < \gamma < 1$, there exist C, c and γ_0 such that for $\gamma_0^{-1} < k < \gamma_0 n$

$$\mathbf{P}(\sigma_{n-k+1} \leq \frac{\varepsilon}{\sqrt{n}}) \leq (\frac{C\varepsilon}{k})^{(1-\gamma)k^2} + \exp(-cn). \tag{4}$$

Equivalently, with $k < \gamma_0 n$, for any $0 < \varepsilon < 1$ let I be the interval $[0, k\varepsilon/\sqrt{n}]$ and N_I be the number of singular values belonging to I. Then we obtain the following (overcrowding) Wegner-type estimate at the hard edge

$$\mathbf{P}(N_I \geq k) \leq (C\varepsilon)^{(1-\delta)k^2} + \exp(-cn). \tag{5}$$

Estimate (5) improves [4, Proposition 4.1] of Cacciapuoti, Maltsev and Schlein where they showed $\mathbf{P}(N_I \geq k) = O(\varepsilon^{Ck})$ with the assumption that ξ has bounded density and subgaussian tail. Under this assumption we can omit the additional terms $\exp(-cn)$ above, see Remark 3.11.

1.5. **Perturbation of random iid matrices.** In connection to Edelman's formula for Ginibre ensemble, and motivated by the study of smoothed analysis, Sankar, Spielman and Teng [22, 23] have found the following striking phenomenon.

Theorem 1.6. *Let M be a Ginibre ensemble, and let F be any deterministic matrix. Then for any $\varepsilon > 0$*

$$\mathbf{P}(\sigma_n(M + F) \leq \frac{\varepsilon}{\sqrt{n}}) = O(\varepsilon).$$

Thus, no matter deterministic matrix F we perturb, the gaussian randomness regularizes it out in such a way that $\sigma_n(M + F)$ behaves similar to $\sigma_n(M)$. Interestingly, this effect no longer holds if the entries of M are allowed to have discrete distributions, see for instance the construction by Tao and Vu in [26], or by Rudelson in [9]. In fact in the latter example, one can have iid M and deterministic $\|F\|_2 = N$ for any N such that $\mathbf{P}(\sigma_n(M + F) \leq \sqrt{n}/N) \geq 1/2$.

Although one can still get useful bounds on $\sigma_n(M + F)$ when $\|M\|_2 = n^{O(1)}$ (see for instance [26] by Tao and Vu), these examples just demonstrate that there is no universal F-independent asymptotic behavior of $\sigma_n(X + F)$ in terms of randomness.

However, we might still ask:

What about the local interaction of the eigenvalues (singular values)?

Our results below, Theorem 1.7 and Corollary 1.8, support the phenomenon that a major part of the eigenvalues (singular values) stays well-separated under arbitrary perturbation even in random matrices having discrete entry distributions.

Theorem 1.7 (the perturbed case for iid matrices). *Let M be a random matrix where the entries are iid copies of a random variable ξ of variance one. Let F be any deterministic matrix. There exists an absolute constant C such that for any $\varepsilon > 0$ and for any $k \geq 1$ we have*

$$\mathbf{P}\left(\sigma_{n-k+1}(M + F) \leq \frac{\varepsilon}{\sqrt{n}}\right) \leq n^{k-1}(Ckp)^{(k-1)^2} \tag{6}$$

where $p = p(\varepsilon) = \sup_{x \in \mathbf{R}} \mathbf{P}(|\xi - x| \leq \varepsilon)$.

Furthermore, for given $0 < \gamma < 1$, there exists a constant $C = C(\gamma)$ such that for any k we have

$$\mathbf{P}\left(\sigma_{n-k+1}(M + F) \leq \frac{\varepsilon k}{\sqrt{n}}\right) \leq n^{\lfloor(1-\gamma/2)k\rfloor}(Cp)^{(1-\gamma)k^2}. \tag{7}$$

Equivalently, with I being the interval $[0, \frac{\varepsilon k}{\sqrt{n}}]$ and N_I be the number of singular values of $M + F$ in I

$$\mathbf{P}(N_I \geq k) \leq n^{\lfloor(1-\gamma/2)k\rfloor}(Cp)^{(1-\gamma)k^2}. \tag{8}$$

Finally, if $|\xi| = O(1)$ with probability one then there exist constants K and c_1, c_2 depending on ξ such that for any $k > K$ we have

$$\mathbf{P}\left(\sigma_{n-k+1}(M + F) \leq \frac{c_1 k}{\sqrt{n}}\right) \leq n^k e^{-c_2 k^2}. \tag{9}$$

It is not clear what would be the optimal exponents on the multiplicative factors $n^{O(\cdot)}$ above (and also in Theorem 1.16 below). With our current method, the more (delocalization) information we know about the singular vectors (the eigenvectors), the better exponents we could obtain. We next deduce two consequences.

Corollary 1.8. *Let M be a random matrix where the entries are iid copies of a random variable ξ of variance one. Let F be any deterministic matrix.*

- *If the common distribution ξ has a density function bounded by K, then for given $0 < \gamma < 1$ there exists a constant $C = C(\gamma)$ such that for any $\varepsilon > 0$ and $k \geq 1$, with $I = [0, \frac{\varepsilon k}{\sqrt{n}}]$*

$$\mathbf{P}(N_I \geq k) \leq n^k (CK\varepsilon)^{(1-\gamma)k^2}. \tag{10}$$

- *If ξ has discrete distribution, then there exist constants C, c_1, c_2 depending on ξ such that for any $k \geq C \log n$, with $I = [0, \frac{c_1 k}{\sqrt{n}}]$*

$$\mathbf{P}(N_I \geq k) \leq e^{-c_2 k^2}. \tag{11}$$

The quadratic rate in (11) is consistent with Theorem 1.2. Furthermore, the bound is valid for any non-degenerate discrete distribution and any F.

1.9. **Symmetric Wigner matrices with sub-gaussian tails.** A symmetric Wigner matrix X of size n with sub-gaussian tail is a random symmetric matrix whose strictly upper triangular entries are iid copies of a real-valued random variable ξ as in (2), and whose diagonal entries are independent sub-gaussian random variables with mean zero and variances bounded by $n^{1-o(1)}$, with the diagonal entries independent of the strictly upper triangular entries.

Similarly to Theorem 1.3, the following result was shown by Vershynin [31].

Theorem 1.10. *With $X = (x_{ij})_{1 \leq i,j \leq n}$ a Wigner matrix of subgaussian entries as in (2), there exists an absolute constant $c < 1$ such that for any fixed real number z and any $\varepsilon > 0$ we have*

$$\mathbf{P}(\min_k |\lambda_k(X) - z| \leq \frac{\varepsilon}{\sqrt{n}}) = O(\varepsilon^{1/9} + e^{-n^c}).$$

This bound does not seem to be optimal, in fact the RHS is conjectured to be similar to that of Theorem 1.3 [31]. Note that it follows from a result by Bourgade, Erdős, Yau and Yin [2] that the distribution of $\min_k \sqrt{n}|\lambda_k(X)|$ is universal.

Under some strong smoothness and decay hypotheses on the entries of a symmetric Wigner matrix X, but without the assumption that the entries have the same variance, a near optimal Wegner-type estimate has been shown in [2, Theorem B1] (see also [8]).

Theorem 1.11. *Let $X = (x_{ij})_{1 \leq i,j \leq n}$ be a symmetric Wigner matrix with entries of finite p-moment for some sufficiently large p, and G be a GOE matrix. For any $t > 0$ we denote $\lambda_1(t) \leq \cdots \leq \lambda_n(t)$ the eigenvalues of $\sqrt{1-t}X + \sqrt{t}G$. Define the set*

$$\mathcal{G}_\delta := \left\{ |\lambda_i - \gamma_i| \leq n^{-2/3+\delta} (\min(i, n+1-i))^{-1/3} \text{ for all } i \in [n] \right\},$$

where γ_i denotes the classical location of the i-th eigenvalue under the semicircle law ordered in increasing order. For any fixed κ there exists $C_1 > 0$ such that for any $k \geq 1, \tau, \delta > 0$ there exists $C_2 > 0$ such that the following holds for any $z \in (-2+\kappa, 2-\kappa)/n^{1/2}, t \in [n^{-\tau}, 1]$ and $\varepsilon > 0$

$$\mathbf{P}\left(\mathcal{G}_\delta \wedge z - \frac{\varepsilon}{\sqrt{n}} \leq \lambda_i \leq \lambda_{i+k-1} \leq z + \frac{\varepsilon}{\sqrt{n}} \text{ for some } i \right) \leq C_2 n^{2k\delta + C_1 k^2 \tau} \varepsilon^{k(k+1)/2}. \tag{12}$$

Note that aside from the correcting factor $n^{2k\delta + C_1 k^2 \tau}$, the bound $\varepsilon^{k(k+1)/2}$ in (12) above is optimal for any fixed k.

Here we show a variant of Theorem 1.11 only under the subgaussian decay of the entries. We will also address the case that k might vary together with n, which seems to be new even for smooth ensembles.

Theorem 1.12 (the symmetric case). *Let $X = (x_{ij})_{1 \leq i,j \leq n}$ be a Wigner matrix of subgaussian entries as in (2). For any $k \geq 1$ there exist a constant C_k depending on k and a constant c depending only on the subgaussian moment of ξ such that for any $\varepsilon > 0$,*

$$\mathbf{P}\left(z - \frac{\varepsilon}{\sqrt{n}} \leq \lambda_i \leq \lambda_{i+k-1} \leq z + \frac{\varepsilon}{\sqrt{n}} \text{ for some } i\right) \leq C_k \varepsilon^{k(k-1)/2} + \exp(-n^c). \tag{13}$$

Furthermore, for given $0 < \gamma < 1$, there exist C, c and γ_0 such that for $\gamma_0^{-1} < k < \gamma_0 n$

$$\mathbf{P}\left(z - \frac{\varepsilon}{\sqrt{n}} \leq \lambda_i \leq \lambda_{i+k-1} \leq z + \frac{\varepsilon}{\sqrt{n}} \text{ for some } i\right) \leq (\frac{C\varepsilon}{k})^{(1-\gamma)k^2/2} + \exp(-n^c). \tag{14}$$

Equivalently, let I be the interval $[z - \frac{\varepsilon k}{\sqrt{n}}, z + \frac{\varepsilon k}{\sqrt{n}}]$ and N_I be the number of eigenvalues belonging to I. Then we have the following Wegner-type estimate

$$\mathbf{P}(N_I \geq k) \leq (C\varepsilon)^{(1-\gamma)k^2/2} + \exp(-n^c). \tag{15}$$

If we work with Hermitian Wigner matrices instead (where the real and imaginary parts of the off diagonal entries are iid subgaussian of zero mean and variance $1/2$) then the exponents of the bounds above are doubled. Furthermore, the additive term $\exp(-n^c)$ can be omitted if the subgaussian random variable has bounded density function (Remark 5.6).

Notice also that unlike in the iid case, our repulsion result is valid over any interval. By taking union bound over all $z = z_i = i\varepsilon/\sqrt{n}, |i| = O(n\varepsilon^{-1})$, we obtain the following bound on all gaps of given range (say for small k).

Corollary 1.13. *With the same assumption as in Theorem 1.12,*

$$\mathbf{P}\Big(\text{There exists } 1 \leq i \leq n - k + 1 \text{ such that } \lambda_{i+k-1} - \lambda_i \leq \frac{\varepsilon}{\sqrt{n}}\Big) = O_k(n\varepsilon^{k(k-1)/2-1}) + \exp(-n^c). \tag{16}$$

This bound is vacuous for $k = 1, 2$ but becomes nearly optimal for large k. It is also comparable to [14, Theorem 2.4] where the gaps $\lambda_{i+k-1} - \lambda_i$ were considered for given i.

1.14. Perturbation of symmetric Wigner matrices.
For symmetric Wigner matrices, Theorem 1.6 was extended to GOE (and GUE) recently by Aizenman, Peled, Schenker, Shamis and S. Sodin in [1] (see also a similar bound by Bourgain in [3]). Furthermore, Farell and Vershynin [9] showed an *F-independent* bound for $\min_i |\lambda_i(X + F)|$ whenever the upper diagonal entries of X have bounded density.

Here again, it is possible to modify the constructions for iid matrices to show that the phenomenon no longer holds when the entries of M have discrete distributions (see for instance [1, p. 19]). Our question here, similarly to the iid case is whether eigenvalue repulsion sustains perturbations.

To proceed further, we introduce a Wegner-type estimate by Aizenman, Peled, Schenker, Shamis and Sodin in [1, (1.9)].

Theorem 1.15. *Let G be a GOE matrix, and F be any deterministic symmetric matrix of size n. Then for any $\varepsilon > 0$, for any interval $I \subset \mathbf{R}$ of length $\varepsilon/\sqrt{n}$, the number of eigenvalues of $G + F$ in I satisfies the*

following F-independent bound for any $k \geq 1$

$$\mathbf{P}(N_I \geq k) \leq (\frac{C\varepsilon}{k})^k,$$

where C is an absolute constant.

It is remarked that although Aizenman et. al. considered Theorem 1.15, their primary focus was on the simultaneous distribution of eigenvalues from possibly different intervals. However, it seems that their method was designed only for GOE and GUE. Also, the repulsion rate should be quadratic in k (as also remarked in [1, p. 18]). Here we show the following.

Theorem 1.16 (the perturbed case for symmetric matrices). *Let X be a random symmetric Wigner matrix where the upper diagonal entries are iid copies of a random variable ξ of variance one. Let F be a deterministic symmetric matrix. There exists an absolute constant C such that for any $\varepsilon > 0$, for any $z \in \mathbf{R}$, and for any $k \geq 1$ we have*

$$\mathbf{P}\left(z - \frac{\varepsilon}{\sqrt{n}} \leq \lambda_i \leq \lambda_{i+k-1} \leq z + \frac{\varepsilon}{\sqrt{n}} \text{ for some } i\right) \leq n^{k-1}(Ckp)^{k(k-1)/2} \tag{17}$$

where $p = p(\varepsilon) = \sup_{x \in \mathbf{R}} \mathbf{P}(|\xi - x| \leq \varepsilon)$.

Furthermore, for given $0 < \gamma < 1$, there exists a constant $C = C(\gamma)$ such that for any k we have

$$\mathbf{P}\left(z - \frac{\varepsilon k}{\sqrt{n}} \leq \lambda_i \leq \lambda_{i+k-1} \leq z + \frac{\varepsilon k}{\sqrt{n}} \text{ for some } i\right) \leq n^{\lfloor(1-\gamma/2)k\rfloor}(Cp)^{(1-\gamma)k^2/2}. \tag{18}$$

Equivalently, with I being the interval $[z - \frac{\varepsilon k}{\sqrt{n}}, z + \frac{\varepsilon k}{\sqrt{n}}]$

$$\mathbf{P}(N_I \geq k) \leq n^{\lfloor(1-\gamma/2)k\rfloor}(Cp)^{(1-\gamma)k^2/2}. \tag{19}$$

Finally, if $|\xi| = O(1)$ with probability one then there exist constants K and c_1, c_2 depending on ξ such that for any $k > K$ we have

$$\mathbf{P}\left(z - \frac{c_1 k}{\sqrt{n}} \leq \lambda_i \leq \lambda_{i+k-1} \leq z + \frac{c_1 k}{\sqrt{n}} \text{ for some } i\right) \leq n^k e^{-c_2 k^2}. \tag{20}$$

Our result automatically extends to the case of Hermitian Wigner matrices where the real and imaginary parts of the off diagonal entries are iid subgaussian of zero mean and non-zero variance. (In fact it seems plaussible to improve the bounds with a double exponent for this model, but we will not pursue this here.) Similarly to Corollary 1.8, we present two corollaries.

Corollary 1.17. *Let X be a random symmetric Wigner matrix where the upper diagonal entries are iid copies of a random variable ξ of variance one. Let F be a deterministic symmetric matrix.*

- *If the common distribution ξ has a density function bounded by K, then for given $0 < \gamma < 1$, there exists a constant $C = C(\gamma)$ such that for any $\varepsilon > 0$ and $k \geq 1$, with I being the interval $[z - \frac{\varepsilon k}{\sqrt{n}}, z + \frac{\varepsilon k}{\sqrt{n}}]$*

$$\mathbf{P}(N_I \geq k) \leq n^k(CK\varepsilon)^{(1-\gamma)k^2/2}. \tag{21}$$

- *If ξ has discrete distribution, then there exists a constant C, c_1, c_2 depending on ξ such that for any $k \geq C \log n$, with I being the interval $[z - \frac{c_1 k}{\sqrt{n}}, z + \frac{c_1 k}{\sqrt{n}}]$*

$$\mathbf{P}(N_I \geq k) \leq e^{-c_2 k^2}. \tag{22}$$

1.18. **Proof method and presentation.** Our method is simple. Roughly speaking it translates the estimates under consideration to the events of having multiple independent columns of small distances to another independent subspace. One then uses the given distance estimates to show that these events are unlikely. As such, the starting point in each proof will be quite similar, however the later steps will evolve differently depending on the models.

The rest of the note is organized as follows. We will introduce the necessary ingredients in Section 2. The proof of Theorem 1.4 is broken down into two parts, Subsection 3.1 and Subsection 3.5 are devoted to proving (3) and (4) respectively. The proof of Theorem 1.12 is carried out in Subsections 5.1 and 5.4 in a similar fashion. The proof of Theorem 1.7 and Theorem 1.16 will be completed in Section 4 and Section 6 respectively. Finally, in the appendix sections we recast the proof of Theorem 1.2, of Theorem 1.11 and of Theorem 1.15 by following the given references.

1.19. **Notation.** We write $X = O(Y)$, $X \ll Y$, or $Y \gg X$ to denote the claim that $|X| \le CY$ for some fixed C; this fixed quantity C is allowed to depend on other fixed quantities such as the sub-gaussian parameter of ξ. We also write $X = \Theta(Y)$ to denote the claim that $X \ll Y$ and $Y \ll X$.

For a square matrix A and a number λ, for short we will write $A - \lambda$ instead of $A - \lambda I_n$. All the norms in this note, if not specified, will be the usual ℓ_2-norm. We use the notation $\mathbf{r}_i(A)$ and $\mathbf{c}_j(A)$ to denote its i-th row and j-th column respectively.

For notational convenience, identical constants will be reused in various contexts; these are usually different if not specified otherwise.

2. Some supporting lemmas

First of all, for the unperturbed models considered in Theorem 1.4 and Theorem 1.12 we will condition on the event $\mathcal{E}_{bound}$ that

$$\|M\|_2 = O(\sqrt{n}) \text{ and } \|X\|_2 = O(\sqrt{n}). \tag{23}$$

These hold with probability $1 - \exp(-\Theta(n))$. We now introduce the definition of compressible and incompressible vectors from [18] by Rudelson and Vershynin.

Definition 2.1. Let $c_0, c_1 \in (0, 1)$ be two numbers (chosen depending on the sub-gaussian moment of ξ.) A vector $x \in \mathbf{R}^n$ is called *sparse* if $|\mathbf{supp}(x)| \le c_0 n$. A vector $x \in S^{n-1}$ is called *compressible* if x is within Euclidean distance c_1 from the set of all sparse vectors. A vector $x \in S^{n-1}$ is called *incompressible* if it is not compressible.

The sets of compressible and incompressible vectors in S^{n-1} will be denoted by $\mathbf{Comp}(c_0, c_1)$ and $\mathbf{Incomp}(c_0, c_1)$ respectively.

Regarding the behavior of $M\mathbf{x}$ and $(X - z)\mathbf{x}$ for compressible vectors, the following was proved in [18] and in [31] (see also [14]) for iid matrices and symmetric Wigner matrices respectively.

Lemma 2.2. *There exist positive constants c_0, c_1, c, α depending on the subgaussian moment of ξ such that the following holds*

- *(iid matrices) For any fixed* $\mathbf{u} \in \mathbf{R}^n$,

$$\mathbf{P}(\inf_{\mathbf{x} \in \mathbf{Comp}(c_0, c_1)} \|M\mathbf{x} - \mathbf{u}\|_2 \leq c\sqrt{n}) \leq \exp(-\alpha n);$$

- *(symmetric matrices) For any fixed* z,

$$\mathbf{P}(\inf_{\mathbf{x} \in \mathbf{Comp}(c_0, c_1)} \|(X - z)\mathbf{x}\|_2 \leq c\sqrt{n}) \leq \exp(-\alpha n).$$

Our next ingredients are results about generic behavior of subspaces spanned by the matrix columns. We first cite a result by Rudelson and Vershynin [19, Theorem 4.1] for the iid model.

Theorem 2.3. *Let* H *be a random subspace in* $\mathbf{R}^n$ *spanned by any* $n - k$ *column vectors of* M *from Theorem 1.4, with* $1 \leq k < \gamma_0 n$ *for some constant* γ_0 *depending on the subgaussian moment of* ξ. *There exists a deterministic subset* $\mathcal{H}$ *of the Grassmanian of* $\mathbf{R}^n$ *where*

$$\mathbf{P}(H^\perp \in \mathcal{H}) \geq 1 - \exp(-c'n)$$

and such that for any subspace H *with* $H^\perp \in \mathcal{H}$ *we have*

$$\mathbf{P}_{\mathbf{c}}(\mathrm{dist}(\mathbf{c}, H) \leq \delta) \leq (\frac{C_0 \delta}{\sqrt{k}})^k + \exp(-c_0 n), \tag{24}$$

where c', C_0, c_0 *are positive constants (depending on the subgaussian moment of* ξ*) and where the entries of* $\mathbf{c}$ *are iid copies of* ξ.

A similar statement for random symmetric Wigner matrices with subexponential bounds, can be found in [15, Theorem 6.1] and [21, Section 8].

Theorem 2.4. *Let* H *be a random subspace in* $\mathbf{R}^n$ *spanned by any* $n - k$ *column vectors of* X *from Theorem 1.12, with* $0 < k < \gamma_0 n$ *for some constant* γ_0 *depending on the subgaussian moment of* ξ. *There exists a deterministic subset* $\mathcal{H}$ *of the Grassmanian of* $\mathbf{R}^n$ *where*

$$\mathbf{P}(H^\perp \in \mathcal{H}) \geq 1 - \exp(-n^{c'})$$

and such that for any subspace H *with* $H^\perp \in \mathcal{H}$ *we have*

$$\mathbf{P}_{\mathbf{c}}(\mathrm{dist}(\mathbf{c}, H) \leq \delta) \leq (\frac{C_0 \delta}{\sqrt{k}})^k + \exp(-n^{c_0}), \tag{25}$$

where c', c_0, C_0 *are positive constants (depending on the subgaussian moment of* ξ*) and where the entries of* $\mathbf{c}$ *are iid copies of* ξ.

We refer the reader to [19, 15, 21] for detailed constructions of $\mathcal{H}$ basing on certain non-Diophantine properties of random structures. To continue the discussion on distances, for perturbed models considered in Theorem 1.7 and Theorem 1.16 we will extract the randomness on the column vector $\mathbf{c}$ only by using the following result from [20, Theorem 1, Corollary 1.3].

Theorem 2.5. *Let* $P_{H^\perp}$ *be a (deterministic) orthogonal projection in* $\mathbf{R}^n$ *onto a subspace of dimension* $k \geq 1$. *Let* $\mathbf{c} = (x_1, \dots, x_n)$ *be a random vector where* x_i *are iid copies of a random vector* ξ *of mean zero, variance one. Then there exist an absolute constant* C_0 *such that*

$$\sup_{\mathbf{u} \in \mathbf{R}^n} \mathbf{P}\Big(\|P_{H^\perp}\mathbf{c} - \mathbf{u}\|_2 \leq t\sqrt{k}\Big) \leq (C_0 p)^k,$$

where $p = p_t$ *is a parameter such that* $\sup_x \mathbf{P}(|\xi - x| \leq t) \leq p$.

In particular, if ξ has density function bounded by K then

$$\sup_{\mathbf{u}\in\mathbf{R}^n} \mathbf{P}\left(\|P_{H^\perp}\mathbf{c} - \mathbf{u}\|_2 \leq t\sqrt{k}\right) \leq (C_0 K t)^k.$$

Notice that the above result does not include arbitrary discrete distribution. In this case, we will replace it by the following result (see also [29] by Tao and Vu).

Theorem 2.6. *Let $P_{H^\perp}$ as in Theorem 2.5. Let $\mathbf{c} = (x_1, \ldots, x_n)$ be a random vector where x_i are iid copies of a random vector ξ of mean zero, variance one and $|\xi| \leq K$ with probability one. Then there exist constants C, c_1, c_2 depending on ξ such that for any $0 < t < 1/2$*

$$\sup_{\mathbf{u}\in\mathbf{R}^n} \mathbf{P}\left(\|P_{H^\perp}\mathbf{c} - \mathbf{u}\|_2 \leq t\sqrt{k} - c_1 K\right) \leq C\exp(-c_2 t^2 k). \tag{26}$$

A proof of Theorem 2.6 is given in Appendix A.

The aforementioned ingredients are sufficient for our results with bounds implicitly depending on k. To work out the explicit dependences (especially for Theorem 1.4 and Theorem 1.12) we will be relying on several additional tools to be introduced below.

The first such ingredient is a non-gap delocalization result by Rudelson and Vershynin [21, Theorem 1.5] for both (unperturbed) iid and random Wigner matrices [1]. For a vector $\mathbf{x} = (x_1, \ldots, x_n)$, and for an index set $I \subset [n]$, we denote $\mathbf{x}_I$ by the projected vector $(x_{i_1}, \ldots, x_{i_{|I|}}), i_1 < \cdots < i_{|I|} \in I$.

Theorem 2.7. *Let $\gamma > 0$ be given. Then there exist positive constants $\gamma_1, \gamma_2, \gamma_3$ depending on γ such that with probability at least $1 - \exp(-\gamma_1 n)$ the following holds.*

(1) *Let M be a random iid matrix as in Theorem 1.4. Then or any $\mathbf{x} \in S^{n-1}$ such that $\|M\mathbf{x}\|_2 \leq \gamma_2 \sqrt{n}$, for any $I \subset [n]$ with $|I| \geq \gamma n$ we have*

$$\|\mathbf{x}_I\|_2 \geq \gamma_3.$$

(2) *Let X be a random Wigner matrix as in Theorem 1.12. Then for any $\mathbf{x} \in S^{n-1}$ such that $\|(X - z)\mathbf{x}\|_2 \leq \gamma_2 \sqrt{n}$, for any $I \subset [n]$ with $|I| \geq \gamma n$ we have*

$$\|\mathbf{x}_I\|_2 \geq \gamma_3.$$

We will also work with non-random matrices. Roughly speaking we would like to obtain many well-conditioned almost square minors from a well-conditioned rectangular matrix. For this type of restricted invertibility phenomenon, we will take into account two powerful such results. The first ingredient is the main result from [17] by Rudelson and Vershynin.

Theorem 2.8. *Let X be an $k \times n$ matrix with $r = \|X\|_{HS}^2/\|X\|_2^2$. Let $\varepsilon, \delta \in (0, 1)$ and let $d \leq n$ be an integer such that*

$$d \geq C\frac{r}{\varepsilon^4\delta}\log\frac{r}{\varepsilon^4\delta}.$$

---

[1]Note that although [21, Theorem 1.5] was stated for eigenvectors, it also extends to approximate eigenvectors as in Theorem 2.7. See [21, Section 4] for the reduction or [16, Remark 2.1.8] for a short discussion on this.

Consider the $k \times d$ matrix $\tilde{X}$ which consists of d unit-normalized columns of X picked independently with replacement, with probabilities propositional to the squares of their Euclidean lengths. Then with probability at least $1 - 2\exp(-c/\delta)$ the following holds

$$\|XX^T - \frac{\|X\|_{HS}^2}{d}\tilde{X}\tilde{X}^T\|_2 \leq \varepsilon\|X\|_2.$$

The second ingredient is a more recent paper [12, Theorem 6] by Naor and Youssef.

Theorem 2.9. *Suppose that X is a full rank matrix of size $k \times d$ with $k \leq d$. Then for $1 \leq l \leq k-1$, there exists l different indices $i_1, \ldots, i_l$ such that the matrix $X_{(i_1,\ldots,i_l)}$ with columns $\mathbf{c}_{i_1}(X), \ldots, \mathbf{c}_{i_l}(X)$ has the smallest non-zero singular value σ_l satisfying*

$$\sigma_l^{-1} \leq K_0 \min_{r \in \{l+1,\ldots,k\}} \sqrt{\frac{dr}{(r-l)\sum_{i=r}^{k}\sigma_i(X)^2}},$$

where K_0 is an absolute constant.

3. RANDOM IID MATRICES: PROOF OF THEOREM 1.4

3.1. **Proof of** (3) **of Theorem 1.4.** In this section we will condition on the event $\mathcal{E}_{bound}$ of (23) and the event $\mathcal{E}_{incomp}$ of Lemma 2.2 that for all unit vectors $\mathbf{x}$ such that $\|M\mathbf{x}\|_2 \leq c\sqrt{n}$, we must have $\mathbf{x} \in$ **Incomp**(c_0, c_1).

By the min-max principle,

$$\sigma_{n-k+1}(M) = \min_{\dim(H)=k} \max_{\mathbf{x} \in H, \|\mathbf{x}\|_2=1} \|M\mathbf{x}\|_2.$$

Thus if $\sigma_{n-k+1}(M) \leq \varepsilon/\sqrt{n}$ then there are k orthogonal unit vectors $\mathbf{z}_1, \ldots, \mathbf{z}_k$ such that

$$\|M\mathbf{z}_i\|_2 \leq \varepsilon/\sqrt{n}, 1 \leq i \leq k. \tag{27}$$

Assume that $\mathbf{z}_i = (z_{i1}, \ldots, z_{in})^T$, and let $\mathbf{c}_1, \ldots, \mathbf{c}_n$ be the column vectors of M. The condition (27) can be read as

$$\|\sum_{j=1}^{n} z_{ij}\mathbf{c}_j\|_2 \leq \varepsilon/\sqrt{n}, \text{ for all } 1 \leq i \leq k.$$

For short, let Z be the $k \times n$ matrix spanned by the row vectors $\mathbf{z}_1^T, \ldots, \mathbf{z}_k^T$, and let $\mathbf{y}_1, \ldots, \mathbf{y}_n \in \mathbf{R}^k$ be the columns of Z. For any (multi)-set $J \subset [n]$, Z_J denotes the matrix generated by $\mathbf{y}_j, j \in J$. By definition,

$$\sum_{1 \leq i \leq n} \|\mathbf{y}_i\|_2^2 = k.$$

We gather here a simple fact that will be useful.

Lemma 3.2. *The set of unit vectors in the subspace $H_{\mathbf{z}_1,\ldots,\mathbf{z}_k}$ spanned by $\mathbf{z}_1, \ldots, \mathbf{z}_k$ belongs to **Incomp**(c_0, c_1). Consequently, for any set $J \subset [n]$ of size at least $(1-c_0)n$, the least singular value of Z_J is at least c_1.*

Proof. (of Lemma 3.2) Note that as $\mathbf{z}_1, \ldots, \mathbf{z}_k$ are orthogonal, for any unit vector $\mathbf{z} = \sum_i \alpha_i \mathbf{z}_i \in H_{\mathbf{z}_1,\ldots,\mathbf{z}_k}$, with $\alpha = (\alpha_1, \ldots, \alpha_k) \in S^k$, by the triangle inequality we have

$$\|M\mathbf{z}\|_2 \leq \sum_i |\alpha_i|\|M\mathbf{z}_i\|_2 \leq \varepsilon\sqrt{k}/\sqrt{n}. \tag{28}$$

As this bound is of order $o(\sqrt{n})$, and as we are under $\mathcal{E}_{incomp}(c_0, c_1)$, this implies that $\mathbf{z} \in \mathcal{E}_{incomp}(c_0, c_1)$, and by definition $\|\mathbf{z}_J\|_2 \geq c_1$ for any set $J \subset [n]$ of size at least $(1 - c_0)n$. In other words, for any unit vector α we have $\|\alpha Z_J\|_2 \geq c_1$, as desired. $\square$

Basing on this elementary fact, we show that Z contains many well-conditioned minors.

Lemma 3.3. *We can choose* $\Theta_k(n^k)$ *tuples* $(j_1, \ldots, j_k) \in [n]^k$ *such that*

$$\|Z_{(j_1,\ldots,j_k)}^{-1}\|_2 = O_k(\sqrt{n}). \tag{29}$$

Proof. (of Lemma 3.3) We say an index i is *bad* if $\|\mathbf{y}_i\|_2 \geq c_0^{-1/2}\sqrt{\frac{k}{n}}$. As $\sum_i \|\mathbf{y}_i\|^2 = k$, the set $W^\sharp$ of bad at most $c_0 n$. Set W to be the complement of $W^\sharp$. Then by Lemma 3.2 the matrix Z_W has the least singular value at least c_1. So trivially, $\det(Z_W Z_W^T) \geq c_1^{2k}$. By the Cauchy-Binet expansion identity,

$$\sum_{j_1,\ldots,j_k \in W} |\det(X_{j_1,\ldots,j_k})|^2 \geq c_1^{2k}.$$

Note also that as $\|\mathbf{y}_i\|_2 \leq c_0^{-1/2}\sqrt{\frac{k}{n}}$. We trivially have $|\det(Z_{(j_1,\ldots,j_k)})|^2 \leq \prod_{i=1}^k \|\mathbf{y}_{j_i}\|_2^2 \leq (c_0^{-1}k/n)^k$. So there are $\Theta_k(n^k)$ tuples $(j_1, \ldots, j_k)$ tuples such that $|\det(Z_{(j_1,\ldots,j_k)})|^2 = \Theta(n^{-k})$. The obtained matrices $Z_{(j_1,\ldots,j_k)}$ clearly satisfy the conclusion of Lemma 3.3 because $|\det(Z_{(j_1,\ldots,j_{k-1}),l})|^2 \leq \prod_{i=1}^{k-1} \|\mathbf{y}_{j_i}\|_2^2 \leq (c_0^{-1}k/n)^{k-1}$, where $Z_{(j_1,\ldots,j_{k-1}),l}$ is the $(k-1) \times (k-1)$ matrix obtained from $Z_{(j_1,\ldots,j_{k-1})}$ by eliminating its l-th row. $\square$

Now assume that $Z_{(j_1,\ldots,j_k)}$ is as in Lemma 3.3. Let A be the $k \times k$ matrix $Z_{(j_1,\ldots,j_k)}^T$. Recall that the event $\mathcal{E}$ in (27) can be written as

$$B = (\mathbf{c}_{j_1}(M), \ldots, \mathbf{c}_{j_k}(M))A + (\mathbf{c}_{j_{k+1}}(M), \ldots, \mathbf{c}_{j_n}(M))A',$$

where $A' = Z_{(j_{k+1},\ldots,j_n)}^T$, and where by definition each of the k column vectors of B has norm at most $\varepsilon/\sqrt{n}$. Let A^{-1} be the inverse of A, then by (29), $\|A^{-1}\|_2 = O(\sqrt{n})$. We have

$$(\mathbf{c}_{j_1}, \ldots, \mathbf{c}_{j_k}) + (\mathbf{c}_{j_{k+1}}, \ldots, \mathbf{c}_{j_n})A'A^{-1} = BA^{-1}. \tag{30}$$

Notice that

$$\|BA^{-1}\|_{HS} \leq \|A^{-1}\|_2 \|B\|_{HS} = O(\varepsilon).$$

Let H be the subspace generated by $\mathbf{c}_{j_{k+1}}, \ldots, \mathbf{c}_{j_n}$. Project each of the identity from (30) onto the orthogonal complement of H, after taking the norm square, we obtain

$$\sum_{i=1}^k \text{dist}(\mathbf{c}_{j_i}, H)^2 \leq \|BA^{-1}\|_{HS}^2.$$

It then follows that

$$\text{dist}(\mathbf{c}_{j_1}, H) = O_k(\varepsilon) \wedge \cdots \wedge \text{dist}(\mathbf{c}_{j_k}, H) = O_k(\varepsilon).$$

By Theorem 2.3 (first conditioning on $H^\perp \in \mathcal{H}$, and then unfolding), this event $\mathcal{E}_{(j_1,\ldots,j_k)}$ is bounded by

$$\mathbf{P}(\mathcal{E}_{(j_1,\ldots,j_k)}) = \left(O_k(\varepsilon^k) + \exp(-\Theta(n)) \right)^k + \exp(-\Theta(n) \leq C_k \varepsilon^{k^2} + \exp(-\Theta(n)).$$

To complete (3) of Theorem 1.4, it remains to combine Lemma 3.3 with the following elementary claim.

Claim 3.4. *Let $\{\mathcal{E}_{(j_1,\ldots,j_k)}, (j_1,\ldots,j_k) \in [n^k]\}$ be a collection of n^k events with $\mathbf{P}(\mathcal{E}_{(j_1,\ldots,j_k)}) \leq p$ for all $(j_1,\ldots,j_k)$. Assume that if $\mathcal{E}$ holds then some cn^k events $\mathcal{E}_{(j_1,\ldots,j_k)}$ of the collection also hold. Then*

$$\mathbf{P}(\mathcal{E}) \leq c^{-1}p.$$

Proof. (of Claim 3.4) Let $I_{(j_1,\ldots,j_k)}$ be the indicator of $\mathcal{E}_{(j_1,\ldots,j_k)}$. Then by definition

$$\sum_{(j_1,\ldots,j_k)} I_{(j_1,\ldots,j_k)} \geq cn^k I_{\mathcal{E}}.$$

Taking expectation,

$$\mathbf{P}(\mathcal{E}) \leq c^{-1}n^{-k} \sum_{j_1,\ldots,j_k} \mathbf{P}(\mathcal{E}_{(j_1,\ldots,j_k)}) \leq c^{-1}p.$$

$\square$

3.5. **Proof of** (4) **of Theorem 1.4.** While (3) gives the optimal rate in terms of ε, the dependence on k is quite poor. Here we will try to improve on Lemma 3.3 by relying on Theorem 2.8 and Theorem 2.9. We will actually prove the following result.

Theorem 3.6. *For any fixed $0 < \gamma < 1$ there exist positive constants C, c and γ_0 depending only on the subgaussian moment of ξ and on γ such that for $1 \leq l \leq k - 1 < \gamma_0 n$ and for any $\varepsilon > 0$*

$$\mathbf{P}(\sigma_{n-k+1} \leq \frac{\varepsilon}{\sqrt{n}}) \leq \left(\frac{C\varepsilon k}{l^{1-\gamma}(k-l)}\right)^{l^2} + \exp(-cn). \tag{31}$$

It is clear that (4) follows from this result by taking l close to k. To prove Theorem 3.6, it suffices to prove for k of order $o(n/\log n)$ because the first error term would be significantly smaller than $\exp(-cn)$ if k becomes larger than this. The extra ingredient to be used is the non-gap delocalization result from Theorem 2.7 for random iid matrix M.

Together with the events $\mathcal{E}_{bound}$ and $\mathcal{E}_{incomp}$ in the previous section, we will also condition on the delocalization event $\mathcal{E}_{deloc}$ of Theorem 2.7 for appropriate choice of γ_0.

With Z as in Section 3.1, we will extract from it many almost square well-conditioned minors.

Lemma 3.7. *There exist constants $\delta, c, C > 0$ such that the following holds for $d = \lfloor Ck \log k \rfloor \leq n$. There exists $c^d n^d$ ordered tuples $(j_1,\ldots,j_d) \in [n]^d$ such that the matrix $Z_{(j_1,\ldots,j_d)}$ satisfies*

$$\sigma_k^{-1}(Z_{(j_1,\ldots,j_d)}) \leq \delta^{-1}\sqrt{\frac{n}{d}}. \tag{32}$$

Proof. (of Lemma 3.7) Similarly to the previous section, we call in index i *bad* if either $\|\mathbf{y}_i\|_2 \geq c_0^{-1/2}\sqrt{\frac{k}{n}}$ or $\|\mathbf{y}_i\|_2 \leq c_1^2\sqrt{\frac{k}{n}}$. As $\sum_i \|\mathbf{y}_i\|^2 = k$, the set $W^{\sharp}$ of indices of the first type is at most $c_0 n$. So by the incompressibility property, $\sum_{i \notin W^{\sharp}} \|\mathbf{y}_i\|^2 = \sum_{j=1}^{k} \sum_{i \notin W^{\sharp}} \mathbf{x}_{ji}^2 \geq c_1^2 k$. Furthermore, for the set $W^{\flat}$ of indices of second-type

$$\sum_{i \in W^{\flat}} \|\mathbf{y}_i\|^2 \leq \sum_{i \in W^{\flat}} c_1^2 \frac{k}{n} \leq c_1^4 k.$$

Thus, with $W = [n] \backslash (W^\sharp \cup W^\flat)$,

$$\sum_{i \in W} \|\mathbf{y}_i\|_2^2 \geq (c_1^2 - c_1^4)k \geq c_1^2 k/2.$$

So,

$$|W| \geq c_0 c_1^2 n/2.$$

Consider the matrix Z' of size $k \times |W|$ generated by the columns $\mathbf{y}_i, i \in W$. Recall that $c_1^2 \sqrt{\frac{k}{n}} \leq \|\mathbf{y}_i\|_2 \leq c_0^{-1/2} \sqrt{\frac{k}{n}}$. Theorem 2.7 applied to $\gamma = c_0 c_1^2/2$ (where we note that the condition $\|Mz\|_2 \leq \gamma_0 \sqrt{n}$ is fulfilled by (28) if n is sufficiently large) implies that

$$\sigma_k(Z') \geq \gamma_3.$$

Now if C is sufficiently large, Theorem 2.8 applied to Z' with $\varepsilon = \gamma_3^2/2$ yields $c^d n^d$ matrices $Z_{(j_1,\ldots,j_d)}$ such that its normalized matrix $\tilde{Z}$ satisfies $\|Z'Z'^T - \frac{\|Z'\|_{HS}^2}{d}\tilde{Z}\tilde{Z}^T\|_2 \leq \gamma_3^2 \|Z'\|_2/2 \leq \gamma_3^2/2$, and so the least singular values are also comparable,

$$\frac{\|Z'\|_{HS}^2}{d}\sigma_k(\tilde{Z}\tilde{Z}^T) \geq \sigma_k(Z'Z'^T) - \gamma_3^2/2 \geq \gamma_3^2 - \gamma_3^2/2 = \gamma_3^2/2.$$

Note that by the process of Theorem 2.8, the $j_1, \ldots, j_d$ are not necessary distinct, but the set $\{j_1, \ldots, j_d\}$ has cardinality at least k in any case. Rescaling back to $Z_{(j_1,\ldots,j_d)}$, we obtain

$$\sigma_k(Z_{(j_1,\ldots,j_d)}) \geq c_1^2 \sqrt{\frac{k}{n}}\sigma_k(\tilde{Z}) > c_1^2 \sqrt{\frac{k}{n}} \frac{\gamma_3 \sqrt{d}}{2\|Z'\|_{HS}} \geq \frac{c_1^2 \gamma_3}{2}\sqrt{\frac{d}{n}}.$$

$\square$

Consider a matrix $Z_{(j_1,\ldots,j_d)}$ that satisfies the conclusion of Lemma 3.7. Let $1 \leq l \leq k-1$. Theorem 2.9 applied to this matrix yields l distinct indices $i_1, \ldots, i_l \in \{j_1, \ldots, j_d\}$ such that the matrix $Z_{(i_1,\ldots,i_l)}$ has the smallest non-zero singular value σ_l satisfying

$$\sigma_l^{-1} \leq K_0 \min_{r \in \{l+1,\ldots,k\}} \sqrt{\frac{\delta^{-2}rn}{(r-l)(k-r+1)}} \leq 4K_0 \delta^{-1}\sqrt{\frac{kn}{(k-l)^2}}, \tag{33}$$

where we chose $r = \lceil \frac{k+l}{2} \rceil$.

As each such l-tuple $(i_1, \ldots, i_l)$ may appear in at most n^{d-l} d-tuples $(j_1, \ldots, j_d)$ obtained from Lemma 3.7, we thus have

Corollary 3.8. *There exists a constant $c > 0$ and $c^d n^l$ l-tuples $(i_1, \ldots, i_l) \in [n]^l$ such that the matrix $Z_{(i_1,\ldots,i_l)}$ satisfies* (33).

We next proceed as in the previous section. Let A be the $l \times k$ matrix $Z_{\{i_1,\ldots,i_l\}}^T$. Recall that the event $\mathcal{E}$ in (27) can be written as

$$B = (\mathbf{c}_{i_1}(M), \ldots, \mathbf{c}_{i_l}(M))A + (\mathbf{c}_{i_{l+1}}(M), \ldots, \mathbf{c}_{i_n}(M))A',$$

where $A' = Z_{(i_{l+1},\ldots,i_n)}^T$, and where by definition each of the k column vectors of B has norm at most $\varepsilon/\sqrt{n}$. Let $\bar{A}$ be the $k \times l$ matrix so that $A\bar{A} = I_l$. Then by Corollary 3.8, $\|\bar{A}\|_2 \leq 4K_0\delta^{-1}\sqrt{\frac{kn}{(k-l)^2}}$. We have

$$(\mathbf{c}_{i_1}, \ldots, \mathbf{c}_{i_l}) + (\mathbf{c}_{i_{l+1}}, \ldots, \mathbf{c}_{i_n})A'\bar{A} = B\bar{A}. \tag{34}$$

Notice that

$$\|B\bar{A}\|_{HS} \le \|\bar{A}\|_2 \|B\|_{HS} \le 4C_0\delta^{-1}\sqrt{\frac{kn}{(k-l)^2}}\sqrt{\frac{k\varepsilon^2}{n}} \le \frac{4K_0\delta^{-1}k}{k-l}\varepsilon.$$

Let H be the subspace generated by $\mathbf{c}_{i_{l+1}}, \ldots, \mathbf{c}_{i_n}$. Similarly to the previous section, (34) implies the event $\mathcal{E}_{(i_1,\ldots,i_l)}$ that

$$\text{dist}(\mathbf{c}_{i_1}, H)^2 + \cdots + \text{dist}(\mathbf{c}_{i_l}, H)^2 \le (\frac{4K_0\delta^{-1}k}{k-l})^2\varepsilon^2.$$

By Claim 3.4 and by Corollary 3.8 (and also because $d = O(k\log k)$, the factor c^d does not affect the bounds), it suffices to show that for any fixed tuple $(i_1, \ldots, i_l)$

$$\mathbf{P}(\mathcal{E}_{(i_1,\ldots,i_l)}) \ll \Big(\frac{C\varepsilon k}{l^{1-\gamma}(k-l)}\Big)^{l^2/2} + \exp(-cn),$$

for some absolute constants C and c.

Lemma 3.9. *Let $\mathcal{E}_t$ be the event* $\text{dist}(\mathbf{c}_{i_1}, H)^2 + \cdots + \text{dist}(\mathbf{c}_{i_l}, H)^2 \le t^2$. *Then there exist constants C and c (which also depend on γ) such that*

$$\mathbf{P}(\mathcal{E}_t) \ll (Ct/l^{1-\gamma})^{l^2} + \exp(-cn).$$

Observe that Lemma 3.9 follows from the following claim where the assumption is satisfied by Theorem 2.3.

Claim 3.10. *Let $t_0 > 0$ be any given number. Assume that ξ_i are independent random variables such that $\mathbf{P}(|\xi_i| < t) \le (Ct/\sqrt{l})^l + \exp(-cn)$ for all $t \ge t_0$. Then for all $0 < \gamma < 1$ there exists a constant C_γ such that for all $t \ge t_0$*

$$\mathbf{P}(\xi_1^2 + \cdots + \xi_l^2 < t^2) \ll \Big(\frac{C_\gamma t}{l^{1-\gamma}}\Big)^{l^2} + \exp(-cn).$$

Proof. (of Claim 3.10) As $\sum_i \xi_i^2 \le t^2$, there are at most $\lceil \gamma l \rceil$ indices such that $|\xi_i| \ge \gamma^{-1/2}t/\sqrt{l}$. For these events we use the simple bound $\mathbf{P}(|\xi_i| \le t)$, while for other events we use the bounds $\mathbf{P}(|\xi_i| \le \gamma^{-1/2}t/\sqrt{l})$. By taking union bound over at most 2^l possibilities, we obtain

$$\mathbf{P}(\xi_1^2 + \cdots + \xi_l^2 < t^2) \le 2^l\Big(\big(\frac{C\gamma^{-1/2}t}{l}\big)^l + \exp(-cn)\Big)^{l-\lceil\gamma l\rceil}\Big(\big(\frac{Ct}{\sqrt{l}}\big)^l + \exp(-cn)\Big)^{\lceil\gamma l\rceil}$$

$$\ll \Big(\frac{C_\gamma t}{l^{1-\gamma}}\Big)^{l^2} + \exp(-cn).$$

$\square$

Remark 3.11. *When ξ has density function bounded by K, then one applies Theorem 2.5 instead of Theorem 2.3 to bound the events $\text{dist}(\mathbf{c}_{i_j}, H) \le t$, conditioning on any realization of H. As a consequence, we obtain a bound $(Kt/\sqrt{l})^l$ without additional terms, and so (3) and (4) hold without $\exp(-cn)$.*

4. Perturbed iid matrices: proof of Theorem 1.7

Similarly to the starting point of Subsection 3.1, by the min-max principle, if $\sigma_{n-k+1}(M+F) \le \varepsilon/\sqrt{n}$ then there are k orthogonal unit vectors $\mathbf{z}_1, \ldots, \mathbf{z}_k$ such that

$$\|(M+F)\mathbf{z}_i\|_2 \le \varepsilon/\sqrt{n}, 1 \le i \le k. \tag{35}$$

Assume that $\mathbf{z}_i = (z_{i1}, \ldots, z_{in})^T$, and let $\mathbf{c}_1, \ldots, \mathbf{c}_n$ be the column vectors of $M + F$. The condition (35) can be read as

$$\|\sum_{j=1}^n z_{ij}\mathbf{c}_j\|_2 \le \varepsilon/\sqrt{n}, \text{ for all } 1 \le i \le k.$$

As usual, let Z be the $k \times n$ matrix spanned by the row vectors $\mathbf{z}_1^T, \ldots, \mathbf{z}_k^T$, and let $\mathbf{y}_1, \ldots, \mathbf{y}_n \in \mathbf{R}^k$ be the columns of Z. For any subset $J \subset [n]$, Z_J denotes the matrix generated by $\mathbf{y}_j, j \in J$. By definition,

$$\sum_{1 \le i \le n} \|\mathbf{y}_i\|_2^2 = k.$$

We will extract from Z one well-conditioned minor.

Lemma 4.1. *Let* $1 \le l \le k - 1$. *There is at least one l-tuple* $(i_1, \ldots, i_l) \in [n]^l$ *such that the matrix* $Z_{(i_1, \ldots, i_l)}$ *of size* $l \times k$ *satisfies*

$$\sigma_l^{-1} \le 4K_0\sqrt{\frac{kn}{(k-l)^2}}, \tag{36}$$

Proof. Theorem 2.9 applied to Z yields l distinct indices $i_1, \ldots, i_l \in \{j_1, \ldots, j_d\}$ such that the matrix $Z_{(i_1, \ldots, i_l)}$ has the smallest non-zero singular value σ_l satisfying

$$\sigma_l^{-1} \le C_0 \min_{r \in \{l+1, \ldots, k\}} \sqrt{\frac{rn}{(r-l)(k-r+1)}} \le 4K_0\sqrt{\frac{kn}{(k-l)^2}},$$

where we chose $r = \lceil\frac{k+l}{2}\rceil$. $\square$

We next proceed as in Subsection 3.5. Let A be the $l \times k$ matrix $Z_{\{i_1, \ldots, i_l\}}^T$. Recall that the event $\mathcal{E}$ in (35) can be written as

$$B = (\mathbf{c}_{i_1}(M+F), \ldots, \mathbf{c}_{i_l}(M+F))A + (\mathbf{c}_{i_{l+1}}(M+F), \ldots, \mathbf{c}_{i_n}(M+F))A',$$

where $A' = Z_{(i_{l+1}, \ldots, i_n)}^T$, and where by definition each of the k column vectors of B has norm at most $\varepsilon/\sqrt{n}$. Let $\bar{A}$ be the $k \times l$ matrix so that $A\bar{A} = I_l$. Then by Lemma 4.1, $\|\bar{A}\|_2 \le 4\sqrt{\frac{kn}{(k-l)^2}}$. We have

$$(\mathbf{c}_{i_1}, \ldots, \mathbf{c}_{i_l}) + (\mathbf{c}_{i_{l+1}}, \ldots, \mathbf{c}_{i_n})A'\bar{A} = B\bar{A}. \tag{37}$$

Notice that

$$\|B\bar{A}\|_{HS} \le \|\bar{A}\|_2\|B\|_{HS} \le 4K_0\sqrt{\frac{kn}{(k-l)^2}}\sqrt{\frac{k\varepsilon^2}{n}} \le \frac{4C_0k}{k-l}\varepsilon.$$

Let H be the subspace generated by $\mathbf{c}_{i_{l+1}}(M+F), \ldots, \mathbf{c}_{i_n}(M+F)$. Equation (37) implies the event $\mathcal{E}_{(i_1, \ldots, i_l)}$ that

$$\text{dist}(\mathbf{c}_{i_1}, H)^2 + \cdots + \text{dist}(\mathbf{c}_{i_l}, H)^2 \le t^2 \tag{38}$$

where

$$t = \frac{4K_0k\varepsilon}{k-l}.$$

Proof of (6). Choose $l = k - 1$. Conditioned on H, the subspace $H^\perp$ has dimension at least l (notice that the more H becomes degenerate, the better bound we will get).

With less focus on the implicit constants, we just simply bound $p_t = \sup_x \mathbf{P}(|\xi - x| \le t) \le 4K_0 k \sup_x \mathbf{P}(|\xi - x| \le \varepsilon) = 4K_0 kp$. By Theorem 2.5 we have

$$\mathbf{P}(\text{dist}(\mathbf{c}_{i_j}, H) \le t) \le \mathbf{P}(\text{dist}(\mathbf{c}_{i_j}, H) \le t\sqrt{l}) \le (C_0 p_t)^l \le (4K_0 C_0 kp)^l.$$

Hence, conditioned on H,

$$\mathbf{P}\Big(\text{dist}(\mathbf{c}_{i_1}, H) \le t) \wedge \cdots \wedge \text{dist}(\mathbf{c}_{i_1}, H) \le t\Big) \le (4K_0 C_0 kp)^{l^2}.$$

The proof of (6) is then complete by unfolding the condition, and by taking the union bound over all possible n^l choices of $(i_1, \ldots, i_l)$.

Proof of (7). Here we use our deduction as above, but with ε replaced by $k\varepsilon$ (as we are working with $(M + F)\mathbf{z}_i \le \varepsilon k/\sqrt{n}$ now), and hence $t = 4K_0 k^2 \varepsilon/(k - l)$.

Assume that k is sufficiently large, we will choose $l = \lfloor (1 - \gamma/2)k \rfloor$. From (38), by averaging there are at least $l' = \lfloor (1 - \gamma/2)l \rfloor$ indices i such that $\mathbf{P}(\text{dist}(\mathbf{c}_{i_j}, H) \le \gamma^{-1/2}t/\sqrt{l})$. Thus, by taking union bound over at most 2^l possibilities, it boils down to estimate the event $\mathcal{E}_{i_1, \ldots, i_{l'}}$ that

$$\text{dist}(\mathbf{c}_{i_1}, H) = O(t/\sqrt{l}) = O(\sqrt{k}\varepsilon) \wedge \cdots \wedge \text{dist}(\mathbf{c}_{i_{l'}}, H) = O(t/\sqrt{l}) = O(\sqrt{k}\varepsilon). \tag{39}$$

With $p = \sup_{x \in \mathbf{R}} \mathbf{P}(|\xi - x| \le \varepsilon)$. Conditioned on H, by Theorem 2.5

$$\mathbf{P}(\mathcal{E}_{i_1, \ldots, i_{l'}}) \le (Cp)^{l'^2}.$$

We then unfold the condition and take union bound over the choices of $(i_1, \ldots, i_{l'})$.

Proof of (9). Assume that k is sufficiently large, we proceed as in the proof of (7) above with $\varepsilon = c_1$ with sufficiently small c_1, and $t = 4K_0 k^2 \varepsilon/(k - l)$, $l = \lfloor (1 - \gamma/2)k \rfloor$, as well as $l' = \lfloor (1 - \gamma/2)l \rfloor$. After obtaining (39), instead of Theorem 2.5 we apply Theorem 2.6, which yields that (conditioned on any realization of H),

$$\mathbf{P}(\mathcal{E}_{i_1, \ldots, i_{l'}}) \le C^{l'} e^{-c_2' c_1^2 l'^2},$$

completing the proof.

5. Random symmetric matrices: proof of Theorem 1.12

5.1. **Proof of** (13) **of Theorem 1.12.** Our starting point is similar to that of Subsection 3.1. We will condition on the event $\mathcal{E}_{bound}$ of (23) and the event $\mathcal{E}_{incomp}$ of Lemma 2.2 that for all unit vectors $\mathbf{x}$ such that $\|(X - z)\mathbf{x}\|_2 \le c\sqrt{n}$, we must have $\mathbf{x} \in \mathbf{Incomp}(c_0, c_1)$.

By definition, there are k orthogonal unit vectors (i.e. the corresponding eigenvectors) $\mathbf{z}_1, \ldots, \mathbf{z}_k$ such that

$$\|(X - z)\mathbf{z}_i\|_2 \le \varepsilon/\sqrt{n}, 1 \le i \le k. \tag{40}$$

Assume that $\mathbf{z}_i = (z_{i1}, \ldots, z_{in})^T$, and let $\mathbf{c}_1, \ldots, \mathbf{c}_n$ be the column vectors of M. The condition (40) can be read as

$$\|\sum_{j=1}^n z_{ij} \mathbf{c}_j\|_2 \le \varepsilon/\sqrt{n}, \text{ for all } 1 \le i \le k.$$

As we are in $\mathcal{E}_{incomp}$, by Lemma 3.2 the set of unit vectors in the subspace $H_{\mathbf{z}_1, \ldots, \mathbf{z}_k}$ spanned by $\mathbf{z}_1, \ldots, \mathbf{z}_k$ belongs to $\mathbf{Incomp}(c_0, c_1)$.

Let Z be the $k \times n$ matrix spanned by the row vectors $\mathbf{z}_1^T, \ldots, \mathbf{z}_k^T$, and let $\mathbf{y}_1, \ldots, \mathbf{y}_n \in \mathbf{R}^k$ be the columns of Z. By the incompressibility property (applied for symmetric matrices), for any set $J \subset [n]$ of size at least $(1 - c_0)n$, the least singular value of Z_J at least c_1, and so Lemma 3.3 applies, hence we can choose $\Theta_k(n^k)$ tuples $(j_1, \ldots, j_k) \in [n]^k$ such that

$$\|Z^{-1}_{(j_1,\ldots,j_k)}\|_2 = O_k(\sqrt{n}). \tag{41}$$

Now assume that $Z_{(j_1,\ldots,j_k)}$ is as in (41). Let A be the $k \times k$ matrix $Z^T_{(j_1,\ldots,j_k)}$. Then by (41), $\|A^{-1}\|_2 = O(\sqrt{n})$. We have

$$(\mathbf{c}_{j_1}, \ldots, \mathbf{c}_{j_k}) + (\mathbf{c}_{j_{k+1}}, \ldots, \mathbf{c}_{j_n})A'A^{-1} = BA^{-1}. \tag{42}$$

Notice that

$$\|BA^{-1}\|_{HS} \leq \|A^{-1}\|_2 \|B\|_{HS} = O(\varepsilon).$$

Let H be the subspace generated by $\mathbf{c}_{j_{k+1}}, \ldots, \mathbf{c}_{j_n}$. Project (42) onto the orthogonal complement of H and taking the squared norm, we obtain

$$\sum_{i=1}^k \operatorname{dist}(\mathbf{c}_{j_i}, H)^2 \leq \|BA^{-1}\|_{HS}^2.$$

It then follows that

$$\operatorname{dist}(\mathbf{c}_{j_1}, H) = O_k(\varepsilon) \wedge \cdots \wedge \operatorname{dist}(\mathbf{c}_{j_k}, H) = O_k(\varepsilon). \tag{43}$$

Thus we have translated the event (40) to the event of having many small distances. Now our treatment with this event $\mathcal{E}_{j_1,\ldots,j_k}$ will be different from the iid case as the distances are now correlated. Without loss of generality, assume that $j_i = i, 1 \leq i \leq k$, and H is the subspace generated by the last $n - k$ columns of X. We will remove correlations by relying on the following simple fact.

Fact 5.2. *For any $I \subset [n]$,*

$$\operatorname{dist}(\mathbf{c}_i, H) \geq \operatorname{dist}(\mathbf{c}_{i,I}, H_I),$$

where $\mathbf{c}_{i,I}$ and H_I are the projections of $\mathbf{c}_i$ and H onto the coordinates indexed by I respectively.

To exploit this fact, we observe that the event $\mathcal{E}_{1,\ldots,k}$ implies the event $\mathcal{F}_{1,\ldots,k}$ where

$$\mathcal{F}_{1,\ldots,k} := \left(\operatorname{dist}(\mathbf{c}_{1,\{2,\ldots,n\}}, H_{\{2,\ldots,n\}}) = O_k(\varepsilon)\right) \wedge \cdots \wedge \left(\operatorname{dist}(\mathbf{c}_{k-2,\{k-1,\ldots,n\}}, H_{\{k-1,\ldots,n\}}) = O_k(\varepsilon)\right)$$
$$\wedge \left(\operatorname{dist}(\mathbf{c}_{k-1,\{k,\ldots,n\}}, H_{\{k,\ldots,n\}}) = O_k(\varepsilon)\right). \tag{44}$$

We next use the multiplicative rule $\mathbf{P}(E_{k-1} \wedge \cdots \wedge E_1) = \mathbf{P}(E_1)\mathbf{P}(E_2|E_1) \ldots \mathbf{P}(E_{k-1}|E_{k-2} \wedge \cdots \wedge E_1)$. Let $\mathscr{A}_{\{i+1,\ldots,n\}}$ be the the sigma-algebra generated by the X-entries $x_{kl}, i + 1 \leq k, l \leq n$.

For $1 \leq i \leq k - 1$, we apply Theorem 2.4 to obtain

$$\mathbf{P}\left(\operatorname{dist}(\mathbf{c}_{i,\{i+1,\ldots,n\}}, H_{\{i+1,\ldots,n\}}) = O_k(\varepsilon)|\mathscr{A}_{\{i+1,\ldots,n\}}\right) = O_k(\varepsilon^{k-i}) + \exp(-n^{c'}).$$

Hence,

$$\mathbf{P}(\mathcal{E}_{1,\ldots,k}) \leq \mathbf{P}(\mathcal{F}_{1,\ldots,k}) \leq O_k(\varepsilon^{(k-1)+\cdots+1} + \exp(-n^{c'})). \tag{45}$$

To complete the proof of (13) of Theorem 1.12 we then just need to use Claim 3.4 again, taking into account all tuples $(j_1, \ldots, j_k)$ obtained from (41).

Remark 5.3. *It is the process of passing from $\mathcal{E}_{j_1,\ldots,j_k}$ to $\mathcal{F}_{j_1,\ldots,j_k}$ via Fact 5.2 that we lost a factor of ε in each iterative conditional substep leading to (45). To recover the cumulative loss of ε^k in the final bound one must not use Fact 5.2 but work directly with $\mathcal{E}_{j_1,\ldots,j_k}$. This requires to work with anti-concentration of quadratic forms [13, 31]. A more plausible goal is to improve the RHS of Theorem 1.10 to $O(\varepsilon + \exp(-n^c))$.*

5.4. **Proof of** (14) **of Theorem 1.12.** Similar to the proof of (4) of Theorem 1.4 in Subsection 3.5, we will condition on the event $\mathcal{E}_{bound}$ of (23), the event $\mathcal{E}_{incomp}$ of Lemma 2.2, and also on the event $\mathcal{E}_{deloc}$ for appropriate choice of γ_0 from the non-gap delocalization result of Theorem 2.7 applied to X.

Next, assume as in Section 5.1 that there are k orthogonal unit vectors (eigenvectors) $\mathbf{z}_1, \ldots, \mathbf{z}_k$ such that

$$\|(X - z)\mathbf{z}_i\|_2 \leq \varepsilon/\sqrt{n}, 1 \leq i \leq k. \tag{46}$$

Let Z be the $k \times n$ matrix spanned by the row vectors $\mathbf{z}_1^T, \ldots, \mathbf{z}_k^T$. With the input of Theorem 2.7, we obtain the following analog of Corollary 3.8

Corollary 5.5. *With $d = \lfloor Ck \log k \rfloor$ and $l = (1 - \gamma/2)k$, there are $2^{-d} c^d n^l$ l-tuples $(i_1, \ldots, i_l) \in [n]^d$ such that the matrix $Z_{(i_1,\ldots,i_l)}$ satisfies (33).*

We proceed as in Section 3.5. Let A be the $l \times k$ matrix $Z_{\{i_1,\ldots,i_l\}}^T$. Recall that the event $\mathcal{E}$ in (46) can be written as

$$B = (\mathbf{c}_{i_1}(X), \ldots, \mathbf{c}_{i_l}(X))A + (\mathbf{c}_{i_{l+1}}(X), \ldots, \mathbf{c}_{i_n}(X))A',$$

where $A' = Z_{(i_{l+1},\ldots,i_n)}^T$, and where by definition each of the k column vectors of B has norm at most $\varepsilon/\sqrt{n}$. Let $\bar{A}$ be the $k \times l$ matrix so that $A\bar{A} = I_l$. Then by Corollary 5.5, $\|\bar{A}\|_2 \leq 4K_0\delta^{-1}\sqrt{\frac{kn}{(k-l)^2}}$. We have

$$(\mathbf{c}_{i_1}, \ldots, \mathbf{c}_{i_l}) + (\mathbf{c}_{i_{l+1}}, \ldots, \mathbf{c}_{i_n})A'\bar{A} = B\bar{A}. \tag{47}$$

As such, with H being the subspace generated by $\mathbf{c}_{i_{l+1}}, \ldots, \mathbf{c}_{i_n}$, we are in the event $\mathcal{E}_{(i_1,\ldots,i_l)}$ that

$$\mathrm{dist}(\mathbf{c}_{i_1}, H)^2 + \cdots + \mathrm{dist}(\mathbf{c}_{i_l}, H)^2 \leq (\frac{4K_0\delta^{-1}k}{k-l})^2\varepsilon^2 \leq (8K_0\delta^{-1}\gamma^{-1}\varepsilon)^2.$$

By Claim 3.4 and by Corollary 3.8 (as bounds of type $(c/2)^d$ are absorbed by other factors), to prove (14) of Theorem 1.12 it suffices to show that there exist absolute constants C and c such that

$$\mathbf{P}(\mathcal{E}_{(1,\ldots,l)}) \leq (\frac{C\varepsilon}{k})^{(1-\gamma)k^2/2} + \exp(-n^c). \tag{48}$$

To this end, let $\mathcal{E}_t$ be the event $\mathrm{dist}(\mathbf{c}_1, H)^2 + \cdots + \mathrm{dist}(\mathbf{c}_l, H)^2 \leq t^2$, where $t = 8\delta^{-1}\gamma^{-1}\varepsilon$. By averaging, there are at most $\gamma l/2$ indices such that $\mathrm{dist}(\mathbf{c}_i, H)^2 \geq \gamma^{-1/2}t/\sqrt{l}$. So there are at least $l' = (1 - \gamma/2)l$ indices i such that $\mathbf{P}(\mathrm{dist}(\mathbf{c}_i, H)^2 \leq \gamma^{-1/2}t/\sqrt{l})$. Again, by taking union bound over at most 2^l possibilities, it boils down to estimate the event $\mathcal{E}_{i_1,\ldots,i_{l'}}$ that

$$\mathrm{dist}(\mathbf{c}_{i_1}, H) = O(t/\sqrt{l}) \wedge \cdots \wedge \mathrm{dist}(\mathbf{c}_{i_{l'}}, H) = O(t/\sqrt{l}). \tag{49}$$

Finally, by the argument leading to (45) basing on Theorem 2.4

$$\mathbf{P}(\mathcal{E}_{1,\ldots,k}) \le \prod_{i=1}^{l'-1} (\frac{Ct}{\sqrt{li}})^i + \exp(-n^c) = (\frac{C't}{\sqrt{ll'}})^{l'(l'-1)/2} + \exp(-n^c)$$

$$\le (\frac{C\varepsilon}{k})^{(1-\gamma)k^2/2} + \exp(-n^c),$$

completing the proof of (48).

Remark 5.6. *When ξ has density function bounded by K, then one applies Theorem 2.5 instead of Theorem 2.4 to bound the events $\mathrm{dist}(\mathbf{c}_{i_j}, H) \le t$ in Subsection 5.1 and Subsection 5.4 (conditioned on any realization of H). As a consequence, we obtain bounds of type $(Kt/\sqrt{l})^l$ without additional terms. Consequently, (13) and (14) hold without the additive terms $\exp(-n^c)$ in this case.*

6. Perturbed symmetric matrices: proof of Theorem 1.16

Our approach here is similar to that of Section 4 and Section 5, so we will be brief. Let z be the midpoint of I. Assume that $\lambda_i \in I$, and $(X + F)\mathbf{z}_i = \lambda_i \mathbf{z}_i$ with orthogonal eigenvectors $\mathbf{x}_i$ of norm one. Then $(X + F - z)\mathbf{z}_i = (\lambda_i - z)\mathbf{z}_i$, and so, with $X' := X + F - z$

$$\|X'\mathbf{z}_i\|_2 \le \varepsilon/\sqrt{n}, 1 \le i \le k. \tag{50}$$

Assume that $\mathbf{z}_i = (z_{i1}, \ldots, z_{in})^T$, and let $\mathbf{c}_1, \ldots, \mathbf{c}_n$ be the column vectors of $X' - z$. Let Z be the $k \times n$ matrix spanned by the row vectors $\mathbf{z}_1^T, \ldots, \mathbf{z}_k^T$, then by (50) we have the following analog of Lemma 4.1.

Lemma 6.1. *Let $1 \le l \le k-1$. There is at least one l-tuple $(i_1, \ldots, i_l) \in [n]^l$ such that the matrix $Z_{(i_1,\ldots,i_l)}$ of size $l \times k$ satisfies*

$$\sigma_l^{-1} \le 4K_0 \sqrt{\frac{kn}{(k-l)^2}}, \tag{51}$$

Next, as in Subsection 5.4, let A be the $l \times k$ matrix $Z_{\{i_1,\ldots,i_l\}}^T$. The event $\mathcal{E}$ in (50) can be written as $B = (\mathbf{c}_{i_1}(X'), \ldots, \mathbf{c}_{i_l}(X'))A + (\mathbf{c}_{i_{l+1}}(X'), \ldots, \mathbf{c}_{i_n}(X'))A'$, where $A' = Z_{(i_{l+1},\ldots,i_n)}^T$, and where each of the k column vectors of B has norm at most $\varepsilon/\sqrt{n}$. Let $\bar{A}$ be the $k \times l$ matrix so that $A\bar{A} = I_l$. Then by Lemma 6.1, $\|\bar{A}\|_2 \le 4K_0 \sqrt{\frac{kn}{(k-l)^2}}$ and we also have

$$(\mathbf{c}_{i_1}, \ldots, \mathbf{c}_{i_l}) + (\mathbf{c}_{i_{l+1}}, \ldots, \mathbf{c}_{i_n})A'\bar{A} = B\bar{A}. \tag{52}$$

A simple calculation shows that $\|B\bar{A}\|_{HS} \le \frac{4K_0 k}{k-l}\varepsilon$. Thus with H being the subspace generated by $\mathbf{c}_{i_{l+1}}, \ldots, \mathbf{c}_{i_n}$, (52) implies the event $\mathcal{E}_{(i_1,\ldots,i_l)}$ that

$$\mathrm{dist}(\mathbf{c}_{i_1}, H)^2 + \cdots + \mathrm{dist}(\mathbf{c}_{i_l}, H)^2 \le t^2, \text{ where } t = \frac{2K_0 k\varepsilon}{k-l}. \tag{53}$$

Without loss of generality, assume $(i_1, \ldots, i_l) = (1, \ldots, l)$. We then pass to distances as in Subsection 5.

Proof of (17). Choose $l = k-1$. In analogy with (44) we pass from $\mathcal{E}_{1,\ldots,l}$ to $\mathcal{F}_{1,\ldots,l-1}$

$$\mathcal{F}_{1,\ldots,l-1} := \left(\mathrm{dist}(\mathbf{c}_{1,\{2,\ldots,n\}}, H_{\{2,\ldots,n\}}) \le t\right) \wedge \cdots \wedge \left(\mathrm{dist}(\mathbf{c}_{l-1,\{l,\ldots,n\}}, H_{\{l,\ldots,n\}}) \le t\right).$$

Let $\mathscr{A}_{\{i+1,\dots,n\}}$ be the the the sigma-algebra generated by the X-entries $x_{rs}, i+1 \leq r, s \leq n$. For $1 \leq i \leq k-1$, we apply Theorem 2.5, with $p_t = \sup_x \mathbf{P}(|\xi - x| \leq t) \leq 2K_0 k \sup_x \mathbf{P}(|\xi - x| \leq \varepsilon) = 2K_0 kp$, we obtain

$$\mathbf{P}\Big(\text{dist}(\mathbf{c}_{i,\{i+1,\dots,n\}}, H_{\{i+1,\dots,n\}}) \leq t | \mathscr{A}_{\{i+1,\dots,n\}}\Big) \leq (2CK_0 kp)^{l-i}.$$

Hence by multiplicative rule $\mathbf{P}(\mathcal{F}_{1,\dots,l-1}) \leq (2CK_0 kp)^{(k-1)+\dots+1}$ as desired.

Proof of (18). Here we use the deduction as above but with ε replaced by $k\varepsilon$ and $t = 2K_0 k^2 \varepsilon/(k-l)$. Choose $l = \lfloor (1-\gamma/2)k \rfloor$. From (53), by averaging there are at least $l' = \lfloor (1-\gamma/2)l \rfloor$ indices i such that $\mathbf{P}(\text{dist}(\mathbf{c}_i, H)^2 \leq \gamma^{-1/2} t/\sqrt{l})$. Thus, by taking union bound over at most 2^l possibilities, it boils down to estimate the event $\mathcal{E}_{1,\dots,l'}$ that

$$\text{dist}(\mathbf{c}_1, H) = O(t/\sqrt{l}) = O(\sqrt{k}\varepsilon) \wedge \dots \wedge \text{dist}(\mathbf{c}_{l'}, H) = O(t/\sqrt{l}) = O(\sqrt{k}\varepsilon).$$

where H is the subspace generated by the last $n-k$ columns of X'. The rest is similar to the proof of (17) where we use Fact 5.2 to pass to $\mathcal{F}_{1,\dots,l'-1} = (\text{dist}(\mathbf{c}_{1,\{2,\dots,n\}}, H_{\{2,\dots,n\}}) = O(\sqrt{k}\varepsilon)) \wedge \dots \wedge (\text{dist}(\mathbf{c}_{l'-1,\{l',\dots,n\}}, H_{\{l',\dots,n\}}) = O(\sqrt{k}\varepsilon))$. We then apply Theorem 2.5 with $p = \sup_x \mathbf{P}(|\xi-x| \leq \varepsilon)$ to obtain $\mathbf{P}(\text{dist}(\mathbf{c}_{i,\{i+1,\dots,n\}}, H_{\{i+1,\dots,n\}}) \leq O(\sqrt{k}\varepsilon)| \mathscr{A}_{\{i+1,\dots,n\}}) \leq (Cp)^{k-i} \leq (Cp)^{l'-i}$, which will then yield the upper bound $\mathbf{P}(\mathcal{F}_{1,\dots,l-1}) \leq (Cp)^{l'(l'-1)/2}$.

Proof of (22). Assume that k is sufficiently large, we proceed as in the proof of (18) with $\varepsilon = c_1$ where c_1 is sufficiently small, and with $t = 2K_0 k^2 \varepsilon/(k-l), l = \lfloor (1-\gamma/2)k \rfloor$ and $l' = \lfloor (1-\gamma/2)l \rfloor$. After using Fact 5.2 to pass to $\mathcal{F}_{1,\dots,l'-1}$, we apply Theorem 2.6 (noting that $H_{\{i+1,\dots,n\}}$ has co-dimension at least $k-i > \gamma k/2$) to obtain $\mathbf{P}(\text{dist}(\mathbf{c}_{i,\{i+1,\dots,n\}}, H_{\{i+1,\dots,n\}}) \leq O(\sqrt{k}c_1)| \mathscr{A}_{\{i+1,\dots,n\}}) \leq Ce^{-c_2' c_1^2 (k-i)}$.

Acknowledgement.

APPENDIX A. PROOF OF THEOREM 2.5

Recall that

$$\mathbf{P}(|\xi| \leq K) = 1.$$

Notice that the case $\mathbf{u} = 0$ follows from Talagrand's concentration bound [25] as $\|P_{H^\perp}\mathbf{c}\|_2$ is convex and 1-Lipschitz (see also [29]).

Theorem A.1. *For any $t > 0$*

$$\mathbf{P}(|\|P_{H^\perp}\mathbf{c}\|_2 - \sqrt{k}| \geq t\sqrt{k}) \leq C\exp(-c_2 t^2 k),$$

for some constants C, c_2 that depend on K.

Now we consider two cases.

Case 1. When $\|\mathbf{u}\|_2 \leq (1-2t)\sqrt{k}$ or $\|\mathbf{u}\|_2 \geq (1+2t)\sqrt{k}$. Then if $\|P_{H^\perp}\mathbf{c} - \mathbf{u}\|_2 \leq t\sqrt{k}$, by the triangle inequality we must have either $\|P_{H^\perp}\mathbf{c}\|_2 \leq (1-t)\sqrt{k}$ or $\|P_{H^\perp}\mathbf{c}\|_2 \geq (1+t)\sqrt{k}$, and so (26) follows from Theorem A.1.

Case 2. When $(1-2t)\sqrt{k} \leq \|\mathbf{u}\|_2 \leq (1+2t)\sqrt{k}$. Consider the random variable $X = |\|P_{H^\perp}\mathbf{c} - \mathbf{u}\|_2$. This function of $\mathbf{c}$ is again convex (as $\|P_{H^\perp}(x\mathbf{c}_1 + (1-x)\mathbf{c}_2) - \mathbf{u}\|_2 = \|x(P_{H^\perp}\mathbf{c}_1 - \mathbf{u}) + (1-x)(P_{H^\perp}\mathbf{c}_2 - \mathbf{u})\|_2 \leq$

$x\|P_{H^\perp}\mathbf{c}_1-\mathbf{u}\|_2+(1-x)\|(P_{H^\perp}\mathbf{c}_2-\mathbf{u}\|_2)$ and 1-Lipschitz, by Talagrand's concentration result, with $M = M(X)$ being the median of X

$$\mathbf{P}(|X - M| \geq t\sqrt{k}) \leq C \exp(-c_2 t^2 k), \tag{54}$$

for some constants C, c_2 that depend on K.

We need to specify the median of X. Notice that

$$X^2 = \mathbf{c}^T P_{H^\perp} \mathbf{c} - 2\mathbf{c}^T P_{H^\perp} \mathbf{u} + \|\mathbf{u}\|_2^2.$$

So $\mathbf{E}X^2 = k + \|\mathbf{u}\|_2^2$. We write

$$Y := X^2 - \mathbf{E}X^2 = X^2 - (k + \|\mathbf{u}\|_2^2) = \sum_{i,j} p_{ij}(x_i x_j - \delta_{ij}) - 2(\sum_i x_i v_i),$$

where $P_{H^\perp} = (p_{ij})$ and $\mathbf{v} = (v_1, \ldots, v_n) = P_{H^\perp}\mathbf{u}$.

After squaring out and as x_i are iid and $\mathbf{E}\xi = 0, \mathbf{E}\xi^2 = 1$ as well as $|\xi| \leq K$, we have

$$\mathbf{E}Y^2 = \mathbf{E}(\sum_{i,j} p_{ij}(x_i x_j - \delta_{ij}))^2 + 4\mathbf{E}(\sum_i x_i v_i)^2 - 4\mathbf{E}\sum_i v_i p_{ii}(x_i^3 - x_i)$$

$$= O(\sum_{i,j} p_{ij}^2 K^2) + O(\sum_i v_i^2) + O(K\sqrt{(\sum_i v_i^2)(\sum_i p_{ii}^2)})$$

$$= O(K^2 k),$$

where we used the fact that $\sum_i v_i^2 \leq \|\mathbf{u}\|_2^2 \leq (1 + 2t)^2 k$ and $\sum_{ij} p_{ij}^2 = \sum_i p_{ii} = k$.

Thus by Markov's inequality the median of $|Y|$ is $O(K\sqrt{k})$, and so the median of X^2 is at least $(k + \|\mathbf{u}\|_2^2) - O(K\sqrt{k})$. Consequently, the median $M(X)$ of X is at least $\sqrt{k + \|\mathbf{u}\|_2^2} - O(K)$. Substitute into (54), with sufficiently small t, we obtain

$$\mathbf{P}\Big(X \leq t\sqrt{k} - O(K)\Big) \leq \mathbf{P}\Big(X \leq \sqrt{k + \|\mathbf{u}\|_2^2} - O(K) - t\sqrt{k}\Big) \leq \mathbf{P}\Big(|X - M| \geq t\sqrt{k}\Big) \leq C \exp(-c_2 t^2 k).$$

REFERENCES

[1] M. Aizenman, R. Peled, J. Schenker, M. Shamis, S. Sodin, *Matrix regularizing effects of Gaussian perturbations* , Commun. Contemp. Math. 19 (2017), no. 3, 1750028, 22 pp.

[2] P. Bourgade, L. Erdős, H.-T. Yau and J. Yin, *Fixed energy universality for generalized Wigner matrices*, Communications on Pure and Applied Mathematics, vol. 69, no. 10, 1815-1881, 2016.

[3] J. Bourgain, *On a Problem of Farrell and Vershynin in Random Matrix Theory*, Geometric Aspects of Functional Analysis pp 65-69, Lecture Notes in Mathematics book series (LNM, volume 2169).

[4] C. Cacciapuoti, A. Maltsev, B. Schlein, *Local Marchenko-Pastur law at the hard edge of sample covariance matrices*, J. Math. Phys. 54 (2013), no. 4, 043302, 13 pp.

[5] M. Carmeli, *Statistical theory and random matrices*, 1983, Dekker, New York.

[6] J. -M. Combes, F. Germinet and A. Klein, *Generalized eigenvalue-counting estimates for the Anderson model*, J. Stat. Phys. 135 (2009), no. 2, 201-216.

[7] A. Edelman, *Eigenvalues and condition numbers of random matrices*, SIAM J. Matrix Anal. Appl. 9 (1988), no. 4, 543-560.

[8] L. Erdős, B. Schlein and H.-T. Yau, *Wegner estimate and level repulsion for Wigner random matrices*, International Mathematics Research notices, 2010, no. 3, 436-479.

[9] B. Farrell and R. Vershynin, *Smoothed analysis of symmetric random matrices with continuous distributions*, Proceedings of the AMS 144 (2016), 2257-2261.

[10] H. Goldstine and J. von Neumann, *Numerical inverting of matrices of high order*, Bull. Amer. Math. Soc. 53 (1947), 1021-1099.

[11] A.Maltsev and B. Schlein, *Average density of states of Hermitian Wigner matrices*, Advances in Mathematics, 228 (2011), no. 5, 2797-2836.

[12] A. Naor and P Youssef, *Restricted invertibility revisited*, to appear in Journey Through Discrete Mathematics. A Tribute to Jiri Matousek.

[13] H. Nguyen, *On the least singular value of random symmetric matrices*, Electronic Journal of Probability 17 (2012), no. 53.

[14] H. Nguyen, T. Tao and V. Vu, *Random matrices: tail bounds for gaps between eigenvalues*, Probability Theory and Related Fields, (2017), Vol. 167, 3, 777-816.

[15] H. Nguyen, *Concentration of distances in Wigner matrices*, Linear Algebra and its Applications 536 (2018), 52-86.

[16] M. Rudelson, *Delocalization of eigenvectors of random matrices*, PCMI 2017 lecture notes, `arxiv.org/pdf/1707.08461.pdf`.

[17] M. Rudelson and R. Vershynin, *Sampling from large matrices: an approach through geometric functional analysis*, Journal of the ACM (2007), Art. 21, 19 pp.

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

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

[20] M. Rudelson and R. Vershynin, *Small ball probabilities for linear images of high dimensional distributions*, International Mathematics Research Notices 19 (2015), 9594-9617.

[21] M. Rudelson and R. Vershynin, *No-gaps delocalization for general random matrices*, Geom. and Funct. Analysis. 26 (2016), 1716-1776.

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

[23] A. Sankar, D. Spielman, and S.-H. Teng, *Smoothed analysis of the condition numbers and growth factors of matrices*, SIAM J. Matrix Anal. Appl. 28 (2006), 446-476.

[24] S. Szarek, *Condition numbers of random matrices*, J. Complexity 7 (1991), no. 2, 131-149.

[25] M. Talagrand, A new look at independence, Annals of Probability 24, 1-34 (1996).

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

[27] T. Tao and V. Vu, *Random matrices: the distribution of the smallest singular values*, Geometric and Functional Analysis, 20 (2010), no. 1, 260-297.

[28] T. Tao and V. Vu, *Inverse Littlewood-Offord theorems and the condition number of random matrices*, Annals of Mathematics (2), 169 (2009), no 2, 595-632.

[29] T. Tao, V. Vu and appendix by M. Krishnapur, Random matrices: universality of ESDs and the circular law, Annals of Probability 38 (2010), no. 5, 2023-2065.

[30] T. Tao and V. Vu, *Random matrices have simple spectrum*, Combinatorica 37 (2017), no. 3, 539-553.

[31] R. Vershynin, *Invertibility of symmetric random matrices*, Random Structures & Algorithms, 44 (2014), no. 2, 135-182.

[32] S. Wilks, *Mathematical Statistics*, 1963, Wiley, New York.

E-mail address: `nguyen.1261@math.osu.edu`

DEPARTMENT OF MATHEMATICS, THE OHIO STATE UNIVERSITY, 231 WEST 18TH AVENUE, COLUMBUS, OH 43210