ON THE SPIELMAN-TENG CONJECTURE

ASHWIN SAH, JULIAN SAHASRABUDHE, AND MEHTAAB SAWHNEY

ABSTRACT. Let M be an $n \times n$ matrix with iid subgaussian entries with mean 0 and variance 1 and let $\sigma_n(M)$ denote the least singular value of M. We prove that

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2}\big) = (1 + o(1))\varepsilon + e^{-\Omega(n)}$$

for all $0 \leqslant \varepsilon \ll 1$. This resolves, up to a $1 + o(1)$ factor, a seminal conjecture of Spielman and Teng.

1. INTRODUCTION

For an $n \times n$ matrix A, the *least singular value* of A is defined to be $\sigma_n(A) = \min_{x \in \mathbb{S}^{n-1}} \|Ax\|_2$. This fundamental quantity has been extensively studied in the context of random matrices and goes back, at least, to the seminal work of von Neumann [33], in the 1960s, on approximate solutions to linear systems of equations. More recently, it has played a central role in breakthroughs on the limiting spectral laws of random matrices, for example in the proof of the famous "circular law" of Tao and Vu [28], and on the "smoothed analysis" of algorithms [23].

In this paper we consider large matrices with iid entries; let M be an $n \times n$ iid random matrix with entries that have mean 0 and variance 1. Here we expect that $\sigma_n(M) \approx n^{-1/2}$ and one is naturally led, in theory and applications, to the study of the quantity

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2}\big). \tag{1}$$

Aside from the several prominent applications mentioned above, this research direction has been stimulated by the work of Spielman and Teng, who formulated a precise conjecture concerning (1) in the case of *Rademacher* random matrices, that is, when the entries of M are iid uniform in $\{-1, 1\}$. Let B be such an $n \times n$ Rademacher matrix; they conjectured that

$$\mathbb{P}\big(\sigma_n(B) \leqslant \varepsilon n^{-1/2}\big) \leqslant \varepsilon + e^{-\Omega(n)}, \tag{2}$$

for all $\varepsilon \geqslant 0$. This conjecture, put forward in their 2002 ICM survey [25], has since become known as the *Spielman–Teng conjecture* and has stimulated a great deal of work on the least singular value of iid matrices in the past 20 years. Most notably Rudelson and Vershynin [20] proved (2) up to a constant factor, and Tao and Vu [27] proved (2) in the regime when $\varepsilon \geqslant n^{-c}$.

In this paper we prove the Spielman–Teng conjecture, up to a $1 + o(1)$ factor, for all $\varepsilon \geqslant 0$. In fact, we prove the same result holds for all random matrices with iid subgaussian entries.

Theorem 1.1. *Let M be an $n \times n$ random matrix with iid subgaussian entries, with mean 0 and variance 1. Then, for all $\varepsilon \geqslant 0$,*

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2}\big) \leqslant (1 + o(1))\varepsilon + e^{-\Omega_\xi(n)}, \tag{3}$$

where the $o(1)$ term decays as $C_\xi(\log n)^{-1/16}$, where $C_\xi > 0$ depends only on ξ.

We suspect that our theorem is best possible in the sense that the $1 + o(1)$ factor cannot be removed for this wider class of matrices. Indeed, for *complex valued* iid matrices, examples were given by Edelman, Guionnet, and Péché [7], suggesting that the $1 + o(1)$ factor is indeed necessary. In other words, we believe that the specific bound (2) conjectured by Spielman and Teng, if true,

is not a "universal phenomenon", independent of the entry distribution, but depends specifically on the properties of the uniform distribution on $\{-1, 1\}$.

We also show the corresponding *lower bound* in Theorem 1.1 when $\varepsilon = o(1)$. We do this by showing that the distribution of $\sigma_n(M)$ agrees with the corresponding quantity for a Guassian random matrix, all the way down to exponentially small scales, up to a $1 + o(1)$ factor.

Theorem 1.2. *Let M an $n \times n$ random matrix with iid subgaussian entries ξ of mean 0 and variance 1 and let G be an $n \times n$ matrix with iid standard normal entries. Then, for all $\varepsilon \geqslant 0$,*

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2}\big) = \big(1 + o(1)\big)\mathbb{P}\big(\sigma_n(G) \leqslant \varepsilon n^{-1/2}\big) + e^{-\Omega_\xi(n)}, \qquad (4)$$

where the $o(1)$ term decays as $C_\xi(\log n)^{-1/16}$, where $C_\xi > 0$ depends only on ξ.

Since the distribution of the least singular value is well understood in the Gaussian case, due to the work of Edelman [5] in the 1980s, we can quickly deduce the following.

Corollary 1.3. *Let M an $n \times n$ random matrix with iid subgaussian entries ξ of mean 0 and variance 1. Then*

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2}\big) = (1 + o(1))\varepsilon + e^{-\Omega_\xi(n)},$$

for all $0 \leqslant \varepsilon \ll 1$.

1.1. Universality in random matrix theory.
These results can be cast in the context of the wider project of proving "universality" for eigenvalue statistics of random matrices. This idea goes back to the pioneering work of Wigner in the 1950s, who put forward the bold hypothesis that the behaviour of the eigenvalues of random matrices depends very little on the underlying distribution of the entries.

In recent years there have been spectacular advances towards the project of confirming and fleshing out Wigner's hypothesis. Prominent among these is the famous "circular law", which was finally settled in full by Tao and Vu [28] after a long sequence of advances [1,6,9,11,16]. This result tells us that the macroscopic "shape" of the eigenvalues M (i.e. the spectral distribution) is largely independent of the entry distribution and converges in the limit to the uniform distribution on the unit disc, when properly renormalized.

Another flagship advance has come in the context of random *symmetric* matrices: let A be an $n \times n$ random symmetric matrix where the entries on and above the diagonal are iid with mean 0 and variance 1. Also let $\Lambda(A) = \{\lambda_1(A)n^{-1/2}, \ldots, \lambda_n(A)n^{-1/2}\}$ be the set of (renormalized) eigenvalues. The extraordinary results of Tao and Vu [29] and Erdős, Schlein and Yau [8] tell us that if $I \subseteq (-2, 2)$ is an interval with $|I| = \Theta(n^{-1})$ then

$$\mathbb{P}\big(\Lambda(A) \cap I \neq \emptyset\big) = \mathbb{P}\big(\Lambda(G_0) \cap I \neq \emptyset\big) + \Theta(n^{-c}), \qquad (5)$$

where G_0 is the $n \times n$ random symmetric matrix with iid standard normal entries on and above the diagonal. Informally (5) states that the probability that a (polynomially) small interval contains a eigenvalue is independent of the underlying entry distribution. More generally, the results of [8,29] imply that the microscopic distribution of the eigenvalue gaps is independent of the underlying distribution of the entries, up to probability scales of order n^{-c}. We note that a polynomial probability error is expected in these statements; in particular the dependence on $\mathbb{E}[A_{ij}^4]$ appears when considering fine asymptotics of the mean of $\lambda_i(A)$ (see [15, Theorem 1.4]).

Another important universality result, and most relevant for us here, was proved by Tao and Vu [27], in the setting of iid matrices. They showed that the distribution of the least singular value

of $\sigma_n(M)$ is similarly universal. Specifically they proved that

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2}\big) = \mathbb{P}\big(\sigma_n(G) \leqslant \varepsilon n^{-1/2}\big) + \Theta(n^{-c}), \tag{6}$$

where G is an iid random matrix with standard normal entries. Again, since the distribution of $\sigma_n(G)$ is known exactly (due to the special symmetries in this case), Tao and Vu were able to deduce that the Spielman–Teng conjecture holds for all $\varepsilon \geqslant n^{-c}$.

In this paper we make what is perhaps the first incursion into the project of studying a "sub–microscopic" universality phenomena, by showing that this same universality phenomena persists all the way down to *exponentially small scales*, if one allows for a $1 + o(1)$ factor.

1.2. **History of the least singular value problem.** The study of the least singular value of iid random matrices goes back to the work of von Neumann (see [33, pg. 14, 477, 555]) in the context of computing approximate solutions to systems of linear equations. He suggested that if B is an $n \times n$ Rademacher random matrix then one has

$$\sigma_n(B) \approx n^{-1/2}. \tag{7}$$

This was later formally conjectured by Smale [24] and then proved by Szarek [26] and Edelman [5] when B is replaced with the $n \times n$ matrix G with iid standard Gaussian entries. In this case, Edelman found an *exact* expression for the density of the least singular value which implies that

$$\mathbb{P}\big(\sigma_n(G) \leqslant \varepsilon n^{-1/2}\big) \leqslant \varepsilon, \tag{8}$$

for all $\varepsilon \geqslant 0$ (see e.g. [25]). While this gives us an essentially complete picture in the Gaussian case, Edelman's proof relies fundamentally on the explicit formulae for the distribution of the singular values of G that are not available for other entry distributions.

One challenging feature of general entry distributions, and in particular of *discrete* distributions, is that such matrices are non-singular with non-zero probability. This singularity event corresponds to the case $\varepsilon = 0$ and exhibits a very different behaviour from the situation for $\varepsilon \geqslant e^{-o(n)}$. In fact, the problem of estimating the singularity probability of discrete random matrices has enjoyed something of its own history, going back to the pioneering work of Komlós [14] in the 1960s, and has since been the subject of intense activity [2, 13, 20, 30]. Today, the best known result on random Radamacher matrices is due to Tikhomirov who, in a breakthrough paper [32], proved

$$\mathbb{P}(\det(B) = 0) = 2^{-n+o(n)}.$$

For general $\varepsilon \geqslant 0$, a key breakthrough was made by Rudelson [18], and independently by Tao and Vu [31], who gave the first bounds on the least singular value beyond the Gaussian case. Then Rudelson and Vershynin [20] proved

$$\mathbb{P}\big(\sigma_n(B) \leqslant \varepsilon n^{-1/2}\big) \leqslant C\varepsilon + e^{-cn}, \tag{9}$$

thereby resolving the Spielman–Teng conjecture, up to constant factors and, as a consequence, Smale's original problem (7).

This was then later complemented by the work of Tao and Vu [27], mentioned above (6), who proved that the constant C can be taken to be 1 in the case $\varepsilon \geqslant n^{-c}$. In this work we, in a sense, "unite" these two results by showing the constant C is universal and equal to 1 for all $\varepsilon \geqslant 0$.

In the next section we give an outline of the proof Theorem 1.2, which implies Theorem 1.1 and Corollary 1.3. We then proceed to prove Theorem 1.2 in Section 3 through Section 9.

2. Outline of proof

Throughout we assume that M is an $n \times n$ matrix with iid subgaussian entries. We assume that $\varepsilon \geqslant e^{-cn}$, and that $\varepsilon < n^{-c}$, otherwise we can apply the theorem of Rudelson and Vershynin, or the theorem of Tao and Vu, respectively. To reduce on clutter, in our discussion here, we will not worry about tracking the exact error term and be content with multiplicative losses of $1 + o(1)$. We shall also just focus our discussion here on proving the "$\leqslant$" direction of Theorem 1.2, since this is the direction that implies Theorem 1.1. Of course, we have corresponding lower bounds for each of the crucial steps below.

2.1. A geometric reduction.
The first key idea in the work of Rudelson and Vershynin is to show that one can control the event $\sigma_n(M) \leqslant \varepsilon n^{-1/2}$ in *geometric* terms. In particular, the first key step in their proof is to show

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2}\big) \leqslant C \cdot \mathbb{E}_{M^*} \, \mathbb{P}_X\big(|\langle v, X \rangle| \leqslant \varepsilon\big) + o(\varepsilon), \tag{10}$$

where $C > 1$ is an absolute constant, X denotes the last row of the matrix M, M^* denotes the matrix M with the last row removed, and v is an arbitrary unit vector in $\ker(M^*)$. This simple and powerful idea allows one to access the least singular value σ_n through an understanding of the kernel vector v and its inner product with a independent random vector X. However, for us, there is a serious drawback in this first step as we necessarily lose a crucial constant factor $C > 1$.

In this paper, our first step draws inspiration from (10), though is much more refined and (as one might imagine) more difficult to work with. Our version of (10) takes the shape

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2}\big) \leqslant \mathbb{E}_{M^*} \, \mathbb{P}_X\left(|\langle v, X \rangle| \leqslant \big(1 + \varepsilon^{1/4}\big)\varepsilon n^{-1/2}\widetilde{\chi}(X)\right) + o(\varepsilon), \tag{11}$$

where $v \in \ker(M^*)$ is a unit vector and $\widetilde{\chi}(X) = \widetilde{\chi}_{M^*}(X)$ is the "correction" term defined by

$$\widetilde{\chi}^2(X) = 1 + \sum_{i=1}^{n-1} \frac{\langle v_i, X \rangle^2}{\sigma_i(M^*)^2}. \tag{12}$$

Here $v_{n-1}, \ldots, v_1$ are the unit singular vectors corresponding to $\sigma_{n-1}(M^*) \leqslant \cdots \leqslant \sigma_1(M^*)$, the singular values of the matrix M^*. Note here that once we fix M^*, the probability on the right hand side of (11) depends only on the inner product of X with singular directions of M^*, in the geometric spirit of (10).

2.2. A truncation step.
The major challenge in working with (11) is, perhaps unsurprisingly, getting a handle on the sum (12), which fluctuates at the scale of $\Theta(n^{1/2})$. Here our first major step in the paper is to show that we can truncate this sum to the terms corresponding to the smallest $\sqrt{\log n}$ singular values, while conceding only marginal losses. More precisely, in Sections 4 through 7, we prove the following "truncated" version of (12)

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2}\big) \leqslant \mathbb{E}_{M^*} \, \mathbb{P}_X\left(|\langle v, X \rangle| \leqslant (1 + \delta_n)\varepsilon n^{-1/2}\chi(X)\right) + o(\varepsilon), \tag{13}$$

where we define, here and throughout the paper,

$$\delta_n = (\log n)^{-c}, \qquad \ell = \sqrt{\log n}, \qquad \text{and} \qquad \chi^2(X) = \sum_{i=1}^{\ell} \frac{\langle v_{n-i}, X \rangle^2}{\sigma_{n-i}(M^*)^2}, \tag{14}$$

for some $c > 0$.

The proof of (13) is fairly involved and will consume our attention from Section 4 to Section 7. Our first step here will be to show that (13) follows fairly easily if the matrix M satisfies a "regularity" event $\mathcal{R}$. While it is a bit technical to define this event (as we do in Section 4) for now it is enough to think of $\mathcal{R}$ as an approximation of the events

$$\sigma_{n-r}(M^*) \approx \mathbb{E}\,\sigma_{n-r}(M^*) \approx rn^{-1/2} \qquad \text{and} \qquad \langle v_i, X \rangle^2 \approx \mathbb{E}\,\langle v_i, X \rangle^2 = 1.$$

From here most of the work comes from showing that the probability that $\mathcal{R}$ fails is negligible. More precisely we need to show that

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \mathcal{R}^c\big) = o(\varepsilon).$$

To prove this we shall need to bootstrap several tools, developed in previous work, to deal with this complicated intersection of events. Of particular interest is the challenge in "decoupling" the intersections

$$\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \langle X, v \rangle^2 \ll 1 \qquad \text{and} \qquad \sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \langle X, v \rangle^2 \gg 1,$$

for which we use and develop the "negative correlation inequalities" introduced recently in the work of Campos, Jenssen, Michelen, and the second author [4] and [3]. We also crucially lean on work of Rudelson and Vershynin [21] on the least singular value of random rectangular matrices to control the event that $\sigma_{n-1}(M^*)$ is atypically small.

2.3. A Gaussian replacement step.
Now to understand the right hand side of (13), we think of fixing $M^* \in \mathcal{E}^*$, where $\mathcal{E}^*$ is the event that M^* is appropriately quasi-random. We then consider the probability $\mathbb{P}_X$ on the right hand side of (13) in isolation and relate this quantity to a similar quantity where the random vector X is replaced with a *Gaussian* random vector $Z = (Z_1, \ldots, Z_n)$, with iid standard Gaussian entries. In particular, we show that

$$\mathbb{P}_X\Big(|\langle v, X \rangle| \leqslant \big(1 + \delta_n\big)\varepsilon n^{-1/2}\chi(X)\Big) \leqslant \mathbb{P}_Z\Big(|\langle v, Z \rangle| \leqslant \big(1 + 2\delta_n\big)\varepsilon n^{-1/2}\chi(Z)\Big) + o(\varepsilon). \tag{15}$$

This "replacement" maneuver echoes the so-called "Lindeberg exchange" method, which has been used to great effect in both random matrix theory (e.g. in [29]) and more widely in mathematics and computer science. The novelty in our replacement step (15) is that we show that this exchange of X for Z can be done at probability scales of order $o(\varepsilon)$, which, indeed, can be *exponentially* small. This is in contrast to the above applications, where one obtains polynomial-type losses in probability. To perform this "exchange", we rely crucially on the quasi-randomness properties of $M^* \in \mathcal{E}^*$.

Once we have (15), we see that some of the challenge in working with the sum $\chi(X)$ immediately falls away. By the rotational invariance of G and the fact that the v_i are orthonormal, we have that

$$\left(\langle X, v \rangle\,,\ \sum_i \frac{\langle v_i, G \rangle^2}{\sigma_i(M^*)^2}\right) \qquad \text{is distributed as} \qquad \left(W_n\,,\ \sum_i \frac{W_i^2}{\sigma_i(M^*)^2}\right), \tag{16}$$

where W_i are iid standard normals.

2.4. Rescaling and passing back to the singular value.
With the observation (16) in tow, it is not too difficult to see that in right hand side of (15), we are able to "replace" ε, which can be exponentially small, with a much larger (and therefore much more tractable) $\varepsilon_0 = n^{-c'}$. More precisely, we show that

$$\mathbb{P}_Z\Big(|\langle v, Z \rangle| \leqslant \big(1 + 2\delta_n\big)\varepsilon n^{-1/2}\chi(G)\Big) \leqslant (\varepsilon/\varepsilon_0)\mathbb{P}_Z\Big(|\langle v, Z \rangle| \leqslant \big(1 + 2\delta_n\big)\varepsilon_0 n^{-1/2}\chi(Z)\Big) + o(\varepsilon). \tag{17}$$

Putting these steps together shows that

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2}\big) \leqslant (\varepsilon/\varepsilon_0)\mathbb{E}_{M^*}\,\mathbb{P}_Z\Big(|\langle v, Z\rangle| \leqslant \big(1+\delta_n\big)\varepsilon_0 n^{-1/2}\chi(Z)\Big) + o(\varepsilon). \qquad (18)$$

We now turn to "mix in" the randomness of M^*. The most direct route here would be to compute the expectation in (18) by estimating the joint distribution of $\sigma_{n-1}(M^*), \ldots, \sigma_{n-\ell}(M^*)$. This is in fact possible, thanks to the work[1] of Tao and Vu, which gives us a very good understanding of the spectrum when $\varepsilon_0 \geqslant n^{-c}$. The trouble is that these results are technical to apply directly, so we instead take a simpler and more indirect route, by relating the quantity on the right hand side of (18) *back* to the probability that the least singular value of a certain matrix is small.

More precisely, we prove, using a version of the inequality (13) *in reverse*, that

$$\mathbb{E}_{M^*}\,\mathbb{P}_Z\Big(|\langle v, G\rangle| \leqslant \big(1+\delta_n\big)\varepsilon_0 n^{-1/2}\chi(Z)\Big) \leqslant \mathbb{P}\big(\sigma_n(\widetilde{M}) \leqslant \varepsilon_0 n^{-1/2}\big) + o(\varepsilon), \qquad (19)$$

where $\widetilde{M}$ is the matrix M^* with a *Gaussian* row Z appended. We may now apply the powerful results of Tao and Vu which tell us that the distribution of σ_n is *universal* at these scales. Thus we can replace $\widetilde{M}$ with an iid Gaussian matrix G, for which the distribution of σ_n is known. We have

$$\mathbb{P}\big(\sigma_n(\widetilde{M}) \leqslant \big(1+\delta_n\big)\varepsilon_0 n^{-1/2}\big) = \mathbb{P}\big(\sigma_n(G) \leqslant \big(1+\delta_n\big)\varepsilon_0 n^{-1/2}\big) + O(n^{-c}) = \varepsilon_0 + o(\varepsilon_0). \qquad (20)$$

This, when taken together with the above, proves our main theorem, Theorem 1.2.

2.5. **Organization of the paper.** In Section 3 we "warm up" by proving the first geometric reduction described at (11). We then endeavour to truncate the sum $\widetilde{\chi}$, which appears there in (11) and prove (13). This "truncation step" takes place over the course of Sections 4 to Section 7.

In Section 8, we turn to prove our Gaussian replacement step, described at (15). Then, in Section 9 we prove the "rescaling step", described in (17) and then pass back to the singular value problem in the matrix $\widetilde{M}$, a step described in (19). We finally put the pieces together and prove Theorem 1.2 and derive Theorem 1.1.

2.6. **Global notation.** In an effort to reduce clutter, this paper is written with a few global assumptions on notation. Throughout ξ with be a subgaussian random variable with subgaussian constant

$$B = \|\xi\|_{\psi_2} := \sup_{p \geqslant 1} p^{-1/2}\big(\mathbb{E}_\xi\|\xi\|^p\big)^{1/p}.$$

Throughout M will denote our $n \times n$ random matrix with entries distributed as ξ, where n is fixed but sufficiently large. We also M^* be the $(n-1) \times n$ matrix formed of the first $n-1$ rows and we will let X denote the last row.

Throughout we also assume that v is a kernel vector of M^* and that $v_{n-1}, \ldots, v_1$ are the right singular vectors of M^* corresponding to $\sigma_{n-1}(M^*) \leqslant \cdots \leqslant \sigma_1(M^*)$. As mentioned above at (14) we shall also assume that $\delta_n = (\log n)^{-c}$, for $c > 0$ sufficiently small.

3. AN EASY GEOMETRIC REDUCTION OF THE SINGULAR VALUE PROBLEM

In this short warm-up section we prove our easy initial geometric reduction of the singular value problem as described at (11). We will then go on to prepare for the rather involved proof of (13).

---

[1]Specifically, [27, Theorem 6.2].

Lemma 3.1. *Let $\varepsilon \geqslant e^{-cn}$. Then*

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2}\big) \leqslant \mathbb{P}\big(|\langle v, X \rangle| \leqslant \big(1 + \varepsilon^{1/4}\big)\varepsilon n^{-1/2}\widetilde{\chi}(X)\big) + O(\delta_n \varepsilon).$$

and conversely

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2}\big) \geqslant \mathbb{P}\big(|\langle v, X \rangle| \leqslant \big(1 - \varepsilon^{1/4}\big)\varepsilon n^{-1/2}\widetilde{\chi}(X)\big) - O(\delta_n \varepsilon).$$

We recall that $\widetilde{\chi}(X)$ is defined at (12). To prove this we use the following deterministic "rank-1 update" formula from linear algebra (see e.g. [10, (5.1)]). We include a short proof.

Lemma 3.2. *Let A be a real $n \times n$ matrix, let A^* be the matrix A with the last row Y removed, and let $u_n, \ldots, u_1$ be the orthogonal vectors corresponding to the singular values $0 = \sigma_n(A^*) \leqslant \sigma_{n-1}(A^*) \leqslant \cdots \leqslant s_1(A^*)$. Then the singular values of A are the positive solutions, in x, to the polynomial equation*

$$\prod_{i=1}^{n}(\sigma_i(A^*)^2 - x) \cdot \left(1 + \sum_{i=1}^{n} \frac{\langle u_i, Y \rangle^2}{\sigma_i(A^*)^2 - x^2}\right) = 0.$$

Proof. Note that $A^T A = (A^*)^T A^* + Y^T Y$. Thus, by the matrix determinant lemma, we have

$$\det(A^T A - xI) = \det((A^*)^T A + Y^T Y - xI) = \det((A^*)^T A - xI) \cdot (1 + Y^T((A^*)^T A - xI)^{-1}Y).$$

We then evaluate the right hand side to be

$$\prod_{i=1}^{n}(\sigma_i(A^*)^2 - x) \cdot \left(1 + \sum_{i=1}^{n}\langle u_i, Y \rangle^2(\sigma_i(A^*)^2 - x)^{-1}\right),$$

which implies the lemma. $\qquad\square$

The proof of Lemma 3.1 is based on the following deterministic lemma which says that our reduction goes through if have some basic control on the least singular value of M^*.

Lemma 3.3. *Let A be a real $n \times n$ matrix, let Y be the last row of A, let A^* be the matrix A with the last row removed, let $u \in \ker(A^*)$, and let $\widetilde{\chi}(Y) = \widetilde{\chi}_{A^*}(Y)$. Let $\varepsilon > 0$ be such that $\sigma_{n-1}(A^*) \geqslant \varepsilon^{3/4}n^{-1/2}$. Then*

$$\sigma_n(A) \leqslant \varepsilon n^{-1/2} \qquad \Longrightarrow \qquad |\langle u, Y \rangle| \leqslant \big(1 + \varepsilon^{1/4}\big)\varepsilon n^{-1/2}\widetilde{\chi}(Y). \tag{21}$$

Conversely,

$$\sigma_n(A) > \varepsilon n^{-1/2} \qquad \Longrightarrow \qquad |\langle u, Y \rangle| > \varepsilon n^{-1/2}\widetilde{\chi}(Y). \tag{22}$$

Proof. Let $\sigma_i = \sigma_i(A^*)$, for $i = 1, \ldots, n-1$ be the singular values of A^* and let u_i be the corresponding unit singular vectors. The least singular value $\sigma_n(A)$ is the unique[2] solution of

$$1 + \sum_{i=1}^{n-1} \frac{\langle u_i, Y \rangle^2}{\sigma_i^2 - x^2} = \frac{\langle u, Y \rangle^2}{x^2} \qquad \text{for} \qquad 0 \leqslant x \leqslant \sigma_{n-1}. \tag{23}$$

For x in this range, we can bound the left hand side of (23) above and below by

$$1 + \sum_{i=1}^{n-1} \frac{\langle u_i, Y \rangle^2}{\sigma_i^2} \leqslant 1 + \sum_{i=1}^{n-1} \frac{\langle u_i, Y \rangle^2}{\sigma_i^2 - x^2} = 1 + \sum_{i=1}^{n-1} \frac{\langle u_i, Y \rangle^2}{\sigma_i^2} \cdot \frac{1}{1 - (x/\sigma_i)^2}. \tag{24}$$

---

[2]Note this solution is easily seen to be unique as the left hand side is monotonically increasing to $+\infty$ in this interval while the right hand side is monotonically decreasing from $+\infty$.

Now assume $\sigma_{n-1}(A^*) \geqslant \varepsilon^{3/4} n^{-1/2}$. If $x = \sigma_n(A) \leqslant \varepsilon n^{-1/2}$ then (24) is at most $(1 + 2\varepsilon^{1/4})\widetilde{\chi}^2(Y)$, which together with (23), implies (21). Conversely, if $x = \sigma_n(A) > \varepsilon n^{-1/2}$ then (22) follows by putting the lower bound at (24) together with (23). $\qquad\square$

To finish the proof of Lemma 3.1 we need the following theorem of Rudelson and Vershynin [21].

Theorem 3.4. *For all $\gamma \geqslant 0$, we have*

$$\mathbb{P}\big(\sigma_{n-1}(M^*) \leqslant \gamma n^{-1/2}\big) \leqslant C\gamma^2 + 2e^{-\Omega(n)},$$

where $C > 0$ and the implicit constant may depend on entry distribution of M^*.

Proof of Lemma 3.1. Simply note that we may assume that $\sigma_{n-1}(M^*) > \varepsilon^{3/4} n^{-1/2}$ since we may apply Theorem 3.4 to see $\mathbb{P}\big(\sigma_{n-1}(M^*) \leqslant \varepsilon^{3/4} n^{-1/2}\big) = O(\delta_n \varepsilon)$. Thus assuming $\sigma_{n-1}(M^*) > \varepsilon^{3/4} n^{-1/2}$, we now apply Lemma 3.3. $\qquad\square$

4. Truncation step I: an initial reduction

In the next few sections we set up our main "truncation step", described in (13) in the proof outline. We then prove that this truncation is possible *assuming* that a "regularity event" $\mathcal{R}$ (to be defined in just a moment) is extremely unlikely. Indeed, we shall need that

$$\mathbb{P}(\sigma_n(M) \leqslant \varepsilon n^{1/2} \wedge \mathcal{R}^c) = O(\delta_n \varepsilon). \tag{25}$$

In the following sections (Sections 5 through 7) we shall then have to work rather hard to show that each of these properties indeed hold with this very high probability. The following lemma is the formal version of our main truncation step.

Lemma 4.1. *For $e^{-cn} < \varepsilon < n^{-c}$ we have*

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2}\big) \leqslant \mathbb{P}\big(|\langle v, X\rangle| \leqslant \big(1 + \delta_n\big)\varepsilon n^{-1/2}\chi(X)\big) + O(\delta_n \varepsilon)$$

and conversely we have

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2}\big) \geqslant \mathbb{P}\big(|\langle v, X\rangle| \leqslant \big(1 - \delta_n\big)\varepsilon n^{-1/2}\chi(X)\big) + O(\delta_n \varepsilon).$$

We now define our regularity event $\mathcal{R} = \mathcal{R}_1 \cap \mathcal{R}_2 \cap \mathcal{R}_3 \cap \mathcal{R}_4$, where $\mathcal{R}_1$ is defined to be the event

$$\sigma_{n-1}(M^*) \geqslant (\log n)^{-3} n^{-1/2}, \tag{26}$$

$\mathcal{R}_2$ is defined to be the event

$$|\langle v_{n-k}, X\rangle| \leqslant \max\{k^{1/8}, \log \log n\}, \text{ for all } 1 \leqslant k \leqslant n - 1$$

intersected with the event

$$\sigma_{n-k}(M^*) \geqslant k^{3/4} n^{-1/2}, \text{ for all } k \geqslant \log \log n.$$

We then define $\mathcal{R}_3$ and $\mathcal{R}_4$ to be the events

$$\sigma_{n-(\log \log n)^2}(M^*) \leqslant (\log \log n)^3 n^{-1/2} \quad \text{and} \quad \sum_{i \leqslant (\log \log n)^2} \langle v_{n-i}, X\rangle^2 \geqslant \log \log n,$$

respectively. The major objective of the next sections will be to show that

Lemma 4.2. *For all $e^{-cn} < \varepsilon < n^{-c}$, we have that*

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \mathcal{R}^c\big) = O(\delta_n \varepsilon).$$

8

But before we come to the proof of Lemma 4.2, we motivate the definition of the event $\mathcal{R}$ by showing that Lemma 4.2 implies Lemma 4.1. We first show the following deterministic lemma.

Lemma 4.3. *Let* $0 \leqslant \varepsilon \leqslant (\log n)^{-4}$ *and let* A *be an* $n \times n$ *matrix with* $A \in \mathcal{R}$. *Let* A^* *be* A *with the last row* Y *removed, let* $u \in \ker(A^*)$, *and write* $\chi = \chi_{A^*}$. *Then*

$$\sigma_n(A) \leqslant \varepsilon n^{-1/2} \quad \Longrightarrow \quad |\langle u, Y \rangle| \leqslant (1 + \delta_n) \varepsilon n^{-1/2} \cdot \chi(Y).$$

Conversely,

$$\sigma_n(A) > \varepsilon n^{-1/2} \quad \Longrightarrow \quad |\langle u, Y \rangle| > (1 - \delta_n) \varepsilon n^{-1/2} \cdot \chi(Y).$$

Proof. Let u_i be the unit singular vector corresponding to $\sigma_i(A^*)$. We show that if $A \in \mathcal{R}$ we have

$$\widetilde{\chi}^2(Y) = 1 + \sum_{i=1}^{n-1} \frac{\langle u_i, Y \rangle^2}{\sigma_i(A^*)^2} = (1 + O(\delta_n)) \sum_{i=1}^{\ell} \frac{\langle u_{n-i}, Y \rangle^2}{\sigma_{n-i}(A^*)^2} = (1 + O(\delta_n)) \chi^2(Y).$$

Once we have have shown this, the result follows from Lemma 3.3. Now, since $A \in \mathcal{R} \subset \mathcal{R}_3 \cap \mathcal{R}_4$,

$$1 + \sum_{i=1}^{n-1} \frac{\langle u_i, Y \rangle^2}{\sigma_i(A^*)^2} \geqslant \frac{1}{\sigma_{n-(\log \log n)^2}(A^*)^2} \sum_{i \leqslant (\log \log n)^2} \langle u_{n-i}, Y \rangle^2 \geqslant \frac{n}{(\log \log n)^5}. \tag{27}$$

Further, since $A \in \mathcal{R} \subset \mathcal{R}_2$, and since $\ell = \sqrt{\log n}$, we bound the tail of the sum $\widetilde{\chi}$ as

$$\sum_{i=\ell+1}^{n} \frac{\langle u_{n-i}, Y \rangle^2}{\sigma_{n-i}(A^*)^2} \leqslant n \sum_{i=\ell+1}^{n} \frac{i^{1/4}}{i^{3/2}} \lesssim \frac{n}{(\log n)^{3/16}},$$

thus implying (27) and thus completing the proof. $\qquad\square$

Proof of Lemma 4.1 assuming Lemma 4.2. Simply write

$$\mathbb{P}(\sigma_n(M) \leqslant \varepsilon n^{-1/2}) = \mathbb{P}(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \cap \mathcal{R}^c) + \mathbb{P}(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \cap \mathcal{R}),$$

and use Lemma 4.2 to deal with the first term on the right hand side and use Lemmas 4.3 and 3.1 to deal with the second. For the lower bound, apply Lemma 4.2 an additional time to get rid of the event $\mathcal{R}$. $\qquad\square$

5. TRUNCATION STEP II: PREPARATION FOR THE PROOF OF LEMMA 4.2

In the next sections we prove Lemma 4.2 and therefore complete the proof of Lemma 4.1. In this section we prove the following three lemmas that will be quite useful towards this goal.

Lemma 5.1. *Let* $\mathcal{E}$ *be an event which is measurable given* M^*. *Let* $\varepsilon \leqslant 1$ *and* $\delta \in [\varepsilon^{3/4}, 1]$. *Then*

$$\mathbb{P}(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \sigma_{n-1}(M^*) \geqslant \delta n^{-1/2} \wedge \mathcal{E}) \lesssim n(\varepsilon/\delta) \cdot \mathbb{P}(\mathcal{E}) + e^{-\Omega(n)}.$$

The second main lemma in this section is a refinement of Lemma 5.1, under the additional event $\mathcal{E}_r$, a cousin of $\mathcal{R}_2$, which we define to be the event

$$|\langle X, v_{n-k} \rangle| \leqslant \max\{k^{1/8}, r\}, \text{ for all } 1 \leqslant k \leqslant n-1 \tag{28}$$

intersected with the event

$$\sigma_{n-k}(M^*) \geqslant k^{3/4} n^{-1/2}, \text{ for all } k \geqslant r. \tag{29}$$

In this section we also prove the following.

Lemma 5.2. *Let* $\varepsilon \leqslant 1$, $\delta \in [\varepsilon^{3/4}, 1]$ *and let* $\mathcal{E}$ *be an event which is measurable given* M^*. *Then*

$$\mathbb{P}(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \sigma_{n-1}(M^*) \geqslant \delta n^{-1/2} \wedge \mathcal{E}_r \wedge \mathcal{E}) \lesssim r^{3/2}(\varepsilon/\delta) \cdot \mathbb{P}(\mathcal{E}) + e^{-\Omega(n)}. \tag{30}$$

The next lemma gives us an analogue of the above two lemmas in the case of the non-M^*-measurable event $\mathcal{E} = \{\langle X, w \rangle > t\}$. To prove this result, we will need a decoupling result from the work of Campos, Jenssen, Michelen, and the second author [4].

Lemma 5.3. *Let $\varepsilon \leqslant 1$, $\delta \in [\varepsilon^{3/4}, 1]$, $t \geqslant 1$ and let $w \in \mathbb{S}^{n-1}$ be measurable given M^*. Then*

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \sigma_{n-1}(M^*) \geqslant \delta n^{-1/2} \wedge |\langle w, X \rangle| \geqslant t\big) \lesssim n\big(\varepsilon/\delta\big) \cdot e^{-\Omega(t^2)} + e^{-\Omega(n)}. \qquad (31)$$

Furthermore,

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \sigma_{n-1}(M^*) \geqslant \delta n^{-1/2} \wedge \mathcal{E}_r \wedge |\langle w, X \rangle| \geqslant t\big) \lesssim r^{3/2}\big(\varepsilon/\delta\big) \cdot e^{-\Omega(t^2)} + e^{-\Omega(n)}. \quad (32)$$

In the remainder of this section we proceed to prove these lemmas. The first step in this direction is to prove the following geometric reductions, based on Lemma 3.3.

5.1. **Proof of Lemma 5.1.**

Lemma 5.4. *Let $\varepsilon > 0$, $\delta \in [\varepsilon^{3/4}, 1]$ and let $\mathcal{E}$ be an event. Then*

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \sigma_n(M^*) > \delta n^{-1/2} \wedge \mathcal{E}\big) \leqslant \mathbb{P}\big(|\langle v, X \rangle| \leqslant 2\big(\varepsilon/\delta\big)(\|X\|_2^2 + 1)^{1/2} \wedge \mathcal{E}\big).$$

Proof. Since $\delta \geqslant \varepsilon^{3/4}$ we use Lemma 3.3. We replace the event $\sigma_n(M) \leqslant \varepsilon n^{-1/2}$ with the event

$$|\langle v, X \rangle| \leqslant 2\varepsilon n^{-1/2}\bigg(1 + \sum_{i=1}^{n-1} \frac{\langle v_i, X \rangle^2}{\sigma_i(M^*)^2}\bigg)^{1/2} \leqslant 2\big(\varepsilon/\delta\big)(\|X\|_2^2 + 1)^{1/2},$$

where the last inequality holds since $\sigma_{n-1}(M^*) \geqslant \delta n^{-1/2}$ and since the v_i are orthogonal. $\qquad \square$

The proof of Lemma 5.1 now follows quickly from Lemma 5.4, along with two key results of Rudelson and Vershynin [20]. For the first we define, for a vector $v \in \mathbb{R}^n$ and $\alpha, \gamma > 0$, the *least common denominator* of v as

$$\mathrm{LCD}_{\alpha,\gamma}(v) = \inf\big\{\theta > 0 \colon \mathrm{dist}(\theta v, \mathbb{Z}^n) < \min\{\gamma\|\theta v\|_2, \sqrt{\alpha n}\}\big\}. \qquad (33)$$

The following fundamental result of Rudelson and Vershynin tells us that kernel vectors of random matrices have large least common denominator.

Theorem 5.5. *There exist constants $\alpha, \gamma, c > 0$, depending only $\|\xi\|_{\psi_2}$, for which*

$$\mathbb{P}_{M^*}\big(\exists v \in \ker(M^*) \text{ with } \mathrm{LCD}_{\alpha,\gamma}(v) < e^{cn}\big) \leqslant 2e^{-cn}.$$

Since the parameters α, γ will be uninteresting for us, we simply fix them so that Theorem 5.5 holds and write $\mathrm{LCD} = \mathrm{LCD}_{\alpha,\gamma}$.

We shall also use another fundamental result of Rudelson and Vershynin (see [19, Theorem 6.2]) which tells us that vectors with large least common denominator have good anti-concentration properties.

Theorem 5.6. *For $\varepsilon > 0$, let $u \in \mathbb{R}^n$, with $\|u\|_2 = 1$, and let $Y = (Y_1, \ldots, Y_n)$ be a random vector where the Y_i are iid, mean 0 and subgaussian. If $\mathrm{LCD}(u) \geqslant C/\varepsilon$ then*

$$\mathbb{P}_Y\big(|\langle u, Y \rangle| \leqslant \varepsilon\big) \leqslant C(\varepsilon + e^{-cn}),$$

where $C, c > 0$ depend only the the subgaussian constant $\|Y_i\|_{\psi_2}$.

We are now in a position to prove Lemma 5.1. For this, let us define the event

$$\mathcal{E}_{\mathrm{lcd}} = \big\{\forall v \in \ker(M^*) \text{ we have } \mathrm{LCM}(v) > e^{cn}\big\}, \qquad (34)$$

where the constant $c > 0$ matches that in Theorem 5.5.

Proof of Lemma 5.1. Note that $\mathbb{P}(\|X\|_2^2 \geqslant n \log n) \leqslant \exp(-\Omega(n \log n))$, since X is subgaussian. Also Theorem 5.5 tells us that $\mathbb{P}(\mathcal{E}_{\mathrm{lcd}}^c) \leqslant e^{-cn}$. Applying these observations and Lemma 5.4 gives

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \sigma_n(M^*) > \delta n^{-1/2} \wedge \mathcal{E}\big) \leqslant \mathbb{P}\big(|\langle v, X\rangle| \leqslant 4(\varepsilon/\delta)(n \log n)^{1/2} \wedge \mathcal{E} \wedge \mathcal{E}_{\mathrm{lcd}}\big) + e^{-\Omega(n)}.$$

Now using the fact that $\mathcal{E}_{\mathrm{lcd}}$ and $\mathcal{E}$ depend only on M^*, we have that the above is

$$\leqslant \max_{M^* \in \mathcal{E} \cap \mathcal{E}_{\mathrm{lcd}}} \mathbb{P}_X\big(|\langle v, X\rangle| \leqslant 4(\varepsilon/\delta)(n \log n)^{1/2} \,|\, M^*\big) \cdot \mathbb{P}_{M_*}(\mathcal{E}) + e^{-\Omega(n)} \leqslant n(\varepsilon/\delta) \cdot \mathbb{P}(\mathcal{E}) + e^{-\Omega(n)},$$

where the last inequality follows from an application of Theorem 5.6. □

5.2. **Proof of Lemma 5.2.**
We now modify the proof of the Lemma 5.4 to prove the refined Lemma 5.2.

Lemma 5.7. *Let $\varepsilon > 0$, $\delta \in [\varepsilon^{3/4}, 1]$ and $\mathcal{E}$ be an event. Then*

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \sigma_{n-1}(M^*) \geqslant \delta n^{-1/2} \wedge \mathcal{E}_r^* \wedge \mathcal{E}\big) \leqslant \mathbb{P}\big(|\langle v, X\rangle| \leqslant Cr^{3/2}(\varepsilon/\delta) \wedge \sigma_{n-1}(M^*) \geqslant \delta n^{-1/2} \wedge \mathcal{E}\big),$$

where $C > 0$ is an absolute constant.

Proof. Since $\delta > \varepsilon^{3/4}$ we use Lemma 3.3 to see that

$$\sigma_n(M) \leqslant \varepsilon n^{-1/2} \qquad \implies \qquad |\langle v, X\rangle| \leqslant 2\varepsilon n^{-1/2}\Big(1 + \sum_{i=1}^{n-1} \frac{\langle v_i, X\rangle^2}{\sigma_i(M^*)^2}\Big)^{1/2}.$$

Now the sum on the right hand side, under the event $\mathcal{E}_r^*$ and given $\sigma_{n-1}(M^*) \geqslant \delta n^{-1/2}$, is

$$1 + \sum_{i=1}^{n-1} \frac{\langle v_i, X\rangle^2}{\sigma_i(M^*)^2} \leqslant \sum_{i=1}^{r} \frac{r^2}{\sigma_{n-i}(M^*)^2} + \sum_{i>r} \frac{(r + (i+1)^{1/8})^2}{(i^{3/4}/\sqrt{n})^2} \leqslant \frac{Cnr^3}{\delta^2}.$$

Thus we may replace $\sigma_n(M) \leqslant \varepsilon n^{-1/2}$ by $|\langle v, X\rangle| \leqslant Cr^{3/2}\varepsilon/\delta$ in the probability. □

The proof of Lemma 5.2 now follows in the same manner as the proof of Lemma 5.1.

Proof of Lemma 5.2. We apply Theorem 5.5, then Lemma 5.7 and then condition on M^*, as in the proof of Lemma 5.1, to see the left hand side of (30) is

$$\leqslant \max_{M^* \in \mathcal{E} \cap \mathcal{E}_{\mathrm{lcd}}} \mathbb{P}_X\big(|\langle v, X\rangle| \leqslant 4r^{3/2}(\varepsilon/\delta) \,|\, M^*\big) \cdot \mathbb{P}_{M_*}(\mathcal{E}) + e^{-\Omega(n)} \lesssim r^{3/2}(\varepsilon/\delta) \cdot \mathbb{P}(\mathcal{E}) + e^{-\Omega(n)},$$

where the last inequality follows from an application of Theorem 5.6. □

5.3. **Proof of Lemma 5.3.**
We now turn to prove one more variant of the above lemmas. Again the proof is similar to that of Lemma 5.1, but this time we incorporate a tool developed by Campos, Jenssen, Michelen, and the second author [4], which tells us that "anti-concentration" events are approximately negatively correlated with "large-deviation" events.

Theorem 5.8. *For $\varepsilon > 0$, let $u, w \in \mathbb{R}^n$ have $\|u\|_2 = \|w\|_2 = 1$ and let $Y = (Y_1, \ldots, Y_n)$ be a random vector, where the Y_i are iid, mean 0 and subgaussian. If $t > 0$ and $\mathrm{LCD}(u) \geqslant C/\varepsilon$ then*

$$\mathbb{P}\big(|\langle u, Y\rangle| \leqslant \varepsilon \wedge \langle w, Y\rangle > t\big) \leqslant C\varepsilon \cdot e^{-ct^2} + e^{-c(n+t^2)},$$

where $C, c > 0$ depend only on the subgaussian constant $\|Y_i\|_{\psi_2}$.

We now prove Lemma 5.3.

Proof of Lemma 5.3. Here there are two items to prove. For (31), apply Theorem 5.5 to deal with $\mathcal{E}_{\mathrm{lcd}}$ then apply Lemma 5.4 with $\mathcal{E} = \{\langle v_n, X \rangle > t\}$ to the left hand side of (31) to see it is

$$\leqslant \mathbb{P}\big(|\langle v, X \rangle| \leqslant 4(\varepsilon/\delta)(n \log n)^{1/2} \wedge \mathcal{E} \wedge \mathcal{E}_{\mathrm{lcd}}\big) + e^{-\Omega(n)}.$$

Conditioning on the worst $M^* \in \mathcal{E}_{\mathrm{lcd}}$, we see, using Theorem 5.8, the above is at most

$$\max_{v,w} \mathbb{P}_X\big(|\langle v, X \rangle| \leqslant 4(\varepsilon/\delta)(n \log n)^{1/2} \wedge \langle w, X \rangle > t\big) \lesssim n(\varepsilon/\delta)e^{-ct^2} + e^{-\Omega(n)},$$

where the maximum is taken over all v, w with $\|v\|_2 = \|w\|_2 = 1$ and $\mathrm{LCD}(v) > e^{cn}$. This proves (31). The proof of (32) is the same with Lemma 5.7 applied in place of Lemma 5.4. $\square$

6. Truncation step III: taking care of the events $\mathcal{R}_1$ and $\mathcal{R}_2$

The goal of this sections will be to prove the first "half" of our main regularity lemma, Lemma 4.2.

Lemma 6.1. *For all $e^{-cn} < \varepsilon < n^{-c}$ we have*

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge (\mathcal{R}_1 \wedge \mathcal{R}_2)^c\big) = O(\delta_n \varepsilon).$$

Recall that $\mathcal{R}_1$ is the event $\sigma_{n-1}(M^*) \geqslant (\log n)^{-3} n^{-1/2}$ and that $\mathcal{R}_2$ is the event

$$|\langle v_{n-k}, X \rangle| \leqslant \max\{k^{1/8}, \log \log n\}, \text{ for all } 1 \leqslant k \leqslant n - 1,$$

intersected with the event

$$\sigma_{n-k}(M^*) \geqslant k^{3/4} n^{-1/2}, \text{ for all } k \geqslant \log \log n.$$

We prove Lemma 6.1 by first proving a weaker analogue of it which we then "bootstrap" to the full result.

6.1. **A weaker version of Lemma 6.1.** We prove a weak version of Lemma 6.1 for $\mathcal{R}_1$ by applying Theorem 3.4, discussed in Section 3, which says that

$$\mathbb{P}\big(\sigma_{n-1}(M^*) \leqslant \gamma n^{-1/2}\big) \leqslant C\gamma^2 + 2e^{-\Omega(n)}. \tag{35}$$

Lemma 6.2. *For $e^{-cn} < \varepsilon < n^{-c}$ we have*

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \sigma_{n-1}(M^*) \leqslant n^{-5/2}\big) = O(\delta_n \varepsilon). \tag{36}$$

Proof. From Theorem 3.4, we see the probability that $\sigma_{n-1}(M^*) \leqslant \varepsilon^{3/4} n^{-1/2}$ is $O(\varepsilon^{3/2})$. Thus the left hand side of (36) is at most

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \sigma_{n-1}(M^*) \in [\varepsilon^{3/4} n^{-1/2}, n^{-5/2}]\big) + O(\delta_n \varepsilon). \tag{37}$$

We now dyadically split up the interval $[\varepsilon^{3/4} n^{-1/2}, n^{-5/2}]$ by setting $\gamma_j = 2^j \cdot \varepsilon^{3/4}$ for all $j \geqslant 0$ such that $\gamma_j n^{-1/2} \leqslant n^{-5/2}$. We now put $\mathcal{E}_j = \{\sigma_{n-1}(M^*) \leqslant 2\gamma_j n^{-1/2}\}$. Thus, summing over j, (37) is

$$\leqslant \sum \mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \sigma_{n-1}(M^*) \geqslant \gamma_j n^{-1/2} \wedge \mathcal{E}_j\big) + O(\delta_n \varepsilon).$$

We now apply Lemma 5.1 to each summand to see the above is at most

$$\sum n(\varepsilon/\gamma_j)\mathbb{P}(\mathcal{E}_j) + O(\delta_n \varepsilon) \leqslant \sum n(\varepsilon/\gamma_j) \cdot \big(\gamma_j^2 + e^{-\Omega(n)}\big) + o(\varepsilon) \leqslant \varepsilon/n + O(\delta_n \varepsilon) = O(\delta_n \varepsilon),$$

as desired. Here we applied Theorem 3.4 for the first inequality. $\square$

We now prove that a weaker version of the regularity event $\mathcal{R}_2$ fails with probability $o(\varepsilon)$, under the event $\sigma_n(M) \leqslant \varepsilon n^{-1/2}$. To define this event, recall the definition of $\mathcal{E}_r$, which we defined at (28) and (29). We define our "weak" version $\widetilde{\mathcal{R}}_2$ of $\mathcal{R}_2$ to be

$$\widetilde{\mathcal{R}}_2 = \mathcal{E}_{\log n}. \tag{38}$$

To handle $\widetilde{\mathcal{R}}_2$ need the following control on the lower tails of $\sigma_i(M^*)$. The following theorem is an immediate consequence (using interlacing) of [17, Theorem 1.4].

Theorem 6.3. *For all $k \geqslant 1$, we have*

$$\mathbb{P}(\sigma_{n-k}(M^*) < ck \cdot n^{-1/2}) \leqslant e^{-k^2/8} + 2e^{-cn},$$

where $c > 0$ is a constant depending only on $\|\xi\|_{\psi_2}$.

Lemma 6.4. *For $e^{-cn} < \varepsilon < n^{-c}$ we have*

$$\mathbb{P}(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \widetilde{\mathcal{R}}_2^c) = O(\delta_n \varepsilon). \tag{39}$$

Proof. By Lemma 6.2, we may intersect the right hand side of (39) with the event $\sigma_{n-1}(M^*) > n^{-5/2}$ at a loss of $O(\delta_n \varepsilon)$ in probability. Set $r_i = \max\{(n-i+1)^{1/8}, \log n\}$ and note that, by Lemma 5.3,

$$\mathbb{P}(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \sigma_{n-1}(M^*) > n^{-5/2} \wedge |\langle v_i, X \rangle| > r_i) \lesssim n^3 \varepsilon \cdot e^{-\Omega(r_i^2)} + e^{-\Omega(n)}, \tag{40}$$

which is at most $\varepsilon n^{-\omega(1)}$.

Now note that if $\mathcal{F}_k = \{\sigma_{n-k}(M^*) < k^{3/4} n^{-1/2}\}$, from Lemma 5.1, we have

$$\mathbb{P}(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \sigma_{n-1}(M^*) > n^{-5/2} \wedge \mathcal{F}_k) \lesssim n^3 \varepsilon \cdot \mathbb{P}(\mathcal{F}_k) + e^{-\Omega(n)} \leqslant \varepsilon n^{-\omega(1)}, \tag{41}$$

where the last inequality holds by Theorem 6.3, when $k \geqslant \log n$, as in the definition of the $\widetilde{\mathcal{R}}_2$.

Now observe that the event on the left hand side of (39) is the union of n events of the type found on the left and side of (40) and of the type found on the left hand side of (41), with $k \geqslant \log n$. Thus we can union bound over (40) and (41) to complete the proof of the Lemma. $\square$

6.2. **Bootstrapping to the proof of Lemma 6.1.** We can now bootstrap these results to prove Lemma 6.1, which follows from combining Lemma 6.5 and Lemma 6.6. In both cases the proof closely follows the proof of the "weaker" counterpart.

Lemma 6.5. *For $e^{-cn} < \varepsilon < n^{-c}$ we have*

$$\mathbb{P}(\sigma(M) \leqslant \varepsilon n^{-1/2} \wedge \sigma_{n-1}(M^*) < (\log n)^{-3} n^{-1/2}) = O(\delta_n \varepsilon). \tag{42}$$

Proof. Apply Theorem 3.4, Lemma 6.2, and Lemma 6.4 to see that we may intersect the event in the left hand side of (42) with the events $\sigma_{n-1}(M^*) \geqslant \varepsilon^{3/4}$, $\sigma_{n-1}(M^*) \geqslant n^{-5/2}$, and $\widetilde{\mathcal{R}}_2$, at the loss of $O(\delta_n \varepsilon)$ in probability. Thus we are interested in the event

$$\sigma_{n-1}(M^*) \in [\max\{\varepsilon^{3/4}, n^{-5/2}\}, (\log n)^{-3} n^{-1/2}].$$

Again we dyadically partition this interval, by setting $\gamma_j n^{-1/2} = 2^j \cdot \max\{\varepsilon^{3/4}, n^{-5/2}\}$ for all $j \geqslant 0$ such that $\gamma_j \leqslant (\log n)^{-3}$. Now put $\mathcal{F}_j = \{\sigma_{n-1}(M^*) \leqslant 2\gamma_j n^{-1/2}\}$ so the probability in (42) is

$$\leqslant \sum_j \mathbb{P}(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \sigma_{n-1}(M^*) \geqslant \gamma_j n^{-1/2} \wedge \widetilde{\mathcal{R}}_2 \wedge \mathcal{F}_j) + O(\delta_n \varepsilon).$$

Thus, recalling that $\widetilde{\mathcal{R}}_2 = \mathcal{E}_{\log n}$ (defined at (38)), we apply Lemma 5.2 to each summand with $r = \log n$ to see that the above is at most

$$\lesssim (\log n)^{3/2} \sum (\varepsilon/\gamma_j) \mathbb{P}(\sigma_{n-1}(M^*) \leqslant 2\gamma_j n^{-1/2}) + O(\delta_n \varepsilon) \lesssim (\log n)^{3/2} \varepsilon \sum \gamma_j + O(\delta_n \varepsilon),$$

where we applied Theorem 3.4 to the probability in each summand. This is at most $O(\delta_n\varepsilon)$. $\qquad\square$

We now note that $\{\sigma_n(M) \leqslant \varepsilon n^{-1/2}\} \wedge \mathcal{R}_2^c$ occurs with negligible probability.

Lemma 6.6. *For $e^{-cn} < \varepsilon < n^{-c}$ we have*
$$\mathbb{P}(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \mathcal{R}_2^c) = O(\delta_n\varepsilon).$$

Proof. Follow the proof of Lemma 6.4, but use Lemma 6.5 in place of Lemma 6.2. We note that instead of taking a union bound over n events, we only need to consider at most $2(\log n)^8$ events not covered by $\widetilde{R}_2$. $\qquad\square$

7. Truncation step IV: Taking care of the events $\mathcal{R}_3$ and $\mathcal{R}_4$

Our goal in this section is to prove the following lemma which amounts to the "second half" of Lemma 4.2. This, when taken together with the work in Section 4, completes the proof of our "truncation step", Lemma 4.1.

Lemma 7.1. *For $e^{-cn} < \varepsilon < n^{-c}$ we have*
$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge (\mathcal{R}_3 \wedge \mathcal{R}_4)^c\big) = O(\varepsilon\delta_n).$$

Recall we defined $\mathcal{R}_3$ and $\mathcal{R}_4$ to be the events (respectively)
$$\sigma_{n-(\log\log n)^2}(M^*) \leqslant (\log\log n)^3 n^{-1/2} \quad \text{and} \quad \sum_{i \leqslant (\log\log n)^2} \langle v_{n-i}, X \rangle^2 \geqslant \log\log n.$$

7.1. Dealing with $\mathcal{R}_3$.
To ensure that we can assume $\mathcal{R}_3$, we need the following result on the upper tails of the σ_i.

Lemma 7.2. *Let $k \leqslant n^c$, and $t \geqslant C$ then*
$$\mathbb{P}\big(\sigma_{n-k}(M^*) \geqslant tk \cdot n^{-1/2}\big) \lesssim e^{-(tk)^2/C} + n^{-c},$$
where $C, c > 0$ depend only on the subgaussian constant.

This result concerns only the "macroscopic" properties of the spectrum and thus falls into a category of statements that we now have a good understanding of. Indeed, this result can be extracted from combining the work of Szarek [26] and Tao and Vu [27].

We now deal with $\mathcal{R}_3$ and, since it is convenient, we also deal with a related event that we will use in Section 8.

Lemma 7.3. *For $e^{-cn} < \varepsilon < n^{-c}$ we have*
$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \mathcal{R}_3^c\big) = O(\delta_n\varepsilon). \tag{43}$$
and
$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \sigma_{n-\sqrt{\log n}}(M^*) \geqslant (\log n)n^{-1/2}\big) = O(\delta_n\varepsilon). \tag{44}$$

Proof. Here we prove (43) and simply note that the proof of (44) is almost identical. From Lemma 6.4 and Lemma 6.5 we intersect with the events $\widetilde{\mathcal{R}}_2$ and $\sigma_{n-1}(M^*) \geqslant (\log n)^{-3}n^{-1/2}$ at a loss of $O(\delta_n\varepsilon)$ in probability. Now apply Lemma 5.2 with $\mathcal{E} = \mathcal{R}_3^c$ to see
$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \sigma_{n-1}(M^*) \geqslant (\log n)^{-3}n^{-1/2} \wedge \mathcal{E}_{\log n} \wedge \mathcal{R}_3^c\big) \leqslant (\log n)^{9/2}\varepsilon \cdot \mathbb{P}(\mathcal{R}_3^c) + e^{-\Omega(n)}.$$
which is $O(\delta_n\varepsilon)$, since we may apply Lemma 7.2 to see
$$\mathbb{P}(R_3^c) = \mathbb{P}\big(\sigma_{n-\log\log n}(M^*) \geqslant (\log\log n)^3 n^{-1/2}\big) \leqslant (\log n)^{-\omega(1)}. \qquad\square$$

7.2. **Dealing with $\mathcal{R}_4$.** For this we use a simpler variant of a negative correlation theorem due to Campos, Jenssen, Michelen and the second author.

Theorem 7.4. *Let $u \in \mathbb{R}^n$ satisfy $\|u\|_2 = 1$ and let $w_1, \ldots, w_k \in \mathbb{R}^n$ be orthogonal unit vectors. Furthermore let $Y = (Y_1, \ldots, Y_n)$ be a random vector where the Y_i are iid, mean 0 and subgaussian. If $\mathrm{LCD}(u) \geqslant C/\varepsilon$ and $k \cdot \|u\|_\infty + \sum_{i=1}^k \|w_i\|_\infty \leqslant (\log n)^{-3}$ then*

$$\mathbb{P}\left(|\langle u, Y \rangle| \leqslant \varepsilon \wedge \sum_{j=1}^k |\langle u_i, Y \rangle|^2 \leqslant ck \right) \leqslant C(\varepsilon e^{-ck} + e^{-cn}),$$

where $C, c > 0$ depend only on the subgaussian constant $\|Y_i\|_{\psi_2}$.

This theorem can be thought of as a weaker version of Theorem 1.2 in the paper of Campos, Jenssen, Michelen, and the second author [4]. Interestingly, this theorem admits a significantly simpler proof which we give in Appendix B.

We also need the following lemma that we will use to say that the unit vectors corresponding to the smallest singular values are relatively flat.

Lemma 7.5. *We have*

$$\mathbb{P}_{M^*}\left(\exists v \in \mathbb{R}^n \text{ with } \|v\|_2 = 1, \|v\|_\infty > Cn^{-c}, \text{ and } \|M^*v\|_2 < 1 \right) < n^{-\omega(1)},$$

where $c > 0$ is absolute and $C > 0$ depends only on the subgaussian constant.

As the proof of this lemma is based on somewhat standard techniques, we defer the proof to Appendix A.

Now recall the event $\mathcal{E}_{\mathrm{lcd}} = \{\exists v \in \ker(M^*) \colon \mathrm{LCD}(v) > e^{cn}\}$, which we defined at (34), and recall that $\mathbb{P}(\mathcal{E}_{\mathrm{lcd}}^c) < e^{-cn}$ by Theorem 5.5.

Lemma 7.6. *For $e^{-cn} < \varepsilon < n^c$, we have*

$$\mathbb{P}\left(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge \mathcal{R}_4 \right) = O(\varepsilon \delta_n). \tag{45}$$

Proof. Let $\mathcal{E}_{\mathrm{flat}}$ denote the event in Lemma 7.5. By Lemma 6.4, Lemma 6.5, Lemma 7.3, Theorem 5.5 and Lemma 7.5 we may intersect with the events

$$\widetilde{\mathcal{R}}_2, \quad \sigma_{n-1}(M^*) \geqslant (\log n)^{-3} n^{-1/2}, \quad \sigma_{n-(\log\log n)^2}(M^*) \leqslant (\log\log n)^3 n^{-1/2}, \quad \mathcal{E}_{\mathrm{lcd}}, \quad \text{and} \quad \mathcal{E}_{\mathrm{flat}},$$

with only $O(\delta_n \varepsilon)$ loss in probability. Let $\mathcal{E}$ be the intersection of these events and let $\mathcal{E}'$ be the intersection of all but the first event. Note that $\mathcal{E}'$ is M^* measurable.

Now apply Lemma 5.7 to see that the left hand side of (45) is at most

$$\mathbb{E}_{M^*} \mathbb{1}(M^* \in \mathcal{E}') \cdot \mathbb{P}_X\left(|\langle v, X \rangle| \leqslant \varepsilon(\log n)^{10} \wedge \sum_{i \leqslant (\log\log n)^2} \langle v_{n-i}, X \rangle^2 \leqslant \log\log n \right) + O(\delta_n \varepsilon). \tag{46}$$

We now look to apply Theorem 7.4 to deal with the probability in (46). For this, note that 1) $\mathrm{LCD}(v) \geqslant C/\varepsilon$ since $M^* \in \mathcal{E}_{\mathrm{lcd}}$, and 2) we have $\|v\|_\infty, \|v_{n-1}\|_\infty, \ldots, \|v_{n-(\log\log n)^2}\|_\infty \leqslant n^{-\Omega(1)}$, which holds by the fact that $\|Mv\|_2, \|Mv_{n-1}\|_2, \ldots, \|Mv_{n-\log\log n}\|_2 \leqslant (\log\log n)n^{-1/2} < 1$ and the assumption that $M^* \in \mathcal{E}_{\mathrm{flat}}$. Thus we apply Theorem 7.4 to see (46) is

$$\lesssim \varepsilon(\log n)^{10} \cdot e^{-\Omega((\log\log n)^2)} + e^{-\Omega(n)} = O(\delta_n \varepsilon). \qquad \square$$

8. A Gaussian replacement step

In this section we show that we can "replace", without much loss, the random vector X with the a Gaussian random vector $Z = (Z_1, \ldots, Z_n)$ in our quantities of interest. Throughout this section we will think of the $(n-1) \times n$ matrix M^* as *fixed* with $M^* \in \mathcal{E}^*$, where $\mathcal{E}^*$ is a "quasi-randomness" event (to be defined shortly) that fails with negligible probability. Our main objective in this section is to prove the following.

Lemma 8.1. *Let $M^* \in \mathcal{E}^*$ and let $\chi = \chi_{M^*}$. Then*

$$\mathbb{P}_X\big(|\langle v, X\rangle| \leqslant \varepsilon n^{-1/2} \cdot \chi(X)\big) \leqslant \mathbb{P}_Z\big(|\langle v, Z\rangle| \leqslant (1 + \delta_n) \cdot \varepsilon n^{-1/2} \cdot \chi(Z)\big) + O(\delta_n \varepsilon) \qquad (47)$$

and

$$\mathbb{P}_X\big(|\langle v, X\rangle| \leqslant \varepsilon n^{-1/2} \cdot \chi(X)\big) \geqslant \mathbb{P}_Z\big(|\langle v, Z\rangle| \leqslant (1 - \delta_n) \cdot \varepsilon n^{-1/2} \cdot \chi(Z)\big) - O(\delta_n \varepsilon) \qquad (48)$$

The proofs of the two inequalities are similar to each other and are proved in two steps. To show (47), we let $\mathcal{F}(X) = \mathcal{F}_{M^*}(X)$ be the event on the left hand side of (47) and let $\mathcal{G}(Z) = \mathcal{G}_{M^*}(Z)$ be the event on the right hand side. We first construct a pair of (very similar) smooth bump functions f^+, f^- for which

$$\mathbb{P}_X\big(\mathcal{F}(X)\big) \leqslant \mathbb{E}_X\, f^+(X) + O(\delta_n \varepsilon) \qquad \text{and} \qquad \mathbb{E}_Z\, f^-(Z) \leqslant \mathbb{P}_Z\big(\mathcal{G}(Z)\big) + O(\delta_n \varepsilon)$$

and for which $|\mathbb{E}_Z\left(f^+(Z) - f^-(Z)\right)| = O(\delta_n \varepsilon)$. We then show, in the most subtle of the steps, that we can replace X with Z in the context of this smooth bump function

$$\big|\mathbb{E}_X\, f^+(X) - \mathbb{E}_Z\, f^+(Z)\big| = O(\delta_n \varepsilon). \qquad (49)$$

These together imply (47). The proof of the lower bound at (48) is similar.

8.1. Definition of $\mathcal{E}^*$.
This "microscopic" replacement is possible at these exponentially small probability scales as a consequence of several quasirandomness properties that we may assume of M^* and which we isolate here. We define $\mathcal{E}^*$ to be the intersection of the events

$$\sigma_{n-1}(M^*) \geqslant (\log n)^{-3} n^{-1/2} \quad \text{and} \quad \sigma_{n-k}(M^*) \geqslant k^{3/4} n^{-1/2} \text{ for all } k \geqslant \log \log n.$$

along with the events

$$\sigma_{n-(\log \log n)^2}(M^*) \leqslant (\log \log n)^3 n^{-1/2} \quad \text{and} \quad \sigma_{n-\sqrt{\log n}}(M^*) \leqslant (\log n) n^{-1/2} \qquad (50)$$

intersected with the events $\mathcal{E}_{\mathrm{lcd}}$ and $\mathcal{E}'_{\mathrm{flat}}$, which are (respectively)

$$\mathrm{LCD}(v) > e^{cn} \qquad \text{and} \qquad \|v\|_\infty, \|v_{n-1}\|, \ldots, \|v_{n-\ell}\|_\infty < n^{-\Omega(1)}. \qquad (51)$$

From our work in earlier sections, it is not hard to see the following.

Lemma 8.2. *For $e^{-cn} < \varepsilon < n^{-c}$ we have*

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2} \wedge (\mathcal{E}^*)^c\big) = O(\delta_n \varepsilon).$$

8.2. Approximation by smooth bump functions.

For M^* fixed we note that the event $\mathcal{F}(X) = \mathcal{F}_{M^*}(X)$ is defined in terms of a smooth function of the random vector

$$H = \big(\langle v, X\rangle, \langle v_{n-1}, X\rangle, \ldots, \langle v_{n-\ell}, X\rangle\big),$$

while the right hand side is a smooth function of the random vector

$$K = \big(\langle v, G\rangle, \langle v_{n-1}, G\rangle, \ldots, \langle v_{n-\ell}, G\rangle\big),$$

where $G \sim \mathcal{N}(0,1)^{\otimes n}$. Note that in fact $K \sim \mathcal{N}(0,1)^{\otimes(\ell+1)}$, since $v_n, \ldots, v_{n-\ell}$ are orthonormal. Thus we can think of the probabilities in Lemma 8.1 as expectations of the indicators $\mathbb{1}(H \in S)$ and $\mathbb{1}(K \in S')$ for some regions $S, S' \subset \mathbb{R}^{\ell+1}$. Before we do the replacement we find smooth bump functions to approximate these regions, which will be amenable to the Fourier methods that follow.

Lemma 8.3. *Let $M^* \in \mathcal{E}^*$. There exist functions $f^-, f^+ : \mathbb{R}^{\ell+1} \to \mathbb{R}$ for which*

$$|\mathbb{E}_G\, f^+(G) - f^-(G)| = O(\delta_n \varepsilon)$$

and for $f \in \{f^+, f^-\}$ we have

$$\mathbb{P}_X(\mathcal{F}(X)) \leqslant \mathbb{E}_X\, f(H) + O(\delta_n \varepsilon) \qquad \text{and} \qquad \mathbb{E}_G\, f(K) \leqslant \mathbb{P}_G(\mathcal{F}(G)) + O(\delta_n \varepsilon),$$

and for all $\theta \in \mathbb{R}$, $\xi = (\xi_1, \ldots, \xi_\ell) \in \mathbb{R}^\ell$, we have

$$|\widehat{f}(\theta, \xi)| \leqslant \varepsilon \cdot \exp\big((\log n)^{2/3}\big) \exp\big(-c(\log n)^{-26}\big(|\theta/\varepsilon|^{1/2} + |\xi_1|^{1/2} + \cdots + |\xi_\ell|^{1/2}\big)\big),$$

where $c > 0$ is an absolute constant.

As this construction is based on fairly standard techniques, we postpone it to Appendix C.

8.3. Replacement step.

Here we prove the replacement step described at (15) above. For this, we will bolster a weaker version of this replacement step known as the Lindeberg exchange method that gives us the analogue of our result up to probability scales of $n^{-\Omega(1)}$. We include a short proof for completeness.

Theorem 8.4. *Let $X_1, \ldots, X_n$, be independent random variables with mean 0, variance 1, let $Z_1, \ldots, Z_n$ be iid standard normals, and let $f : \mathbb{R}^n \to \mathbb{C}$ be a smooth function. Then*

$$\big|\mathbb{E}\, f(X_1, \ldots, X_n) - \mathbb{E} f(Z_1, \ldots, Z_n)\big| \leqslant \Big(\max_{1 \leqslant i \leqslant n} \mathbb{E}\, |X_i|^3 + \mathbb{E}\, |Z_i|^3\Big) \cdot \sum_{i=1}^n \left\|\frac{d^3 f}{dx_i^3}\right\|_\infty.$$

Proof. We have that

$$\big|\mathbb{E}[f(X_1, \ldots, X_n) - f(Z_1, \ldots, Z_n)]\big|$$

$$\leqslant \sum_{i=1}^n \big|\mathbb{E}[f(X_1, \ldots, X_{i-1}, Z_i, \ldots, Z_n)] - \mathbb{E}[f(X_1, \ldots, X_i, Z_{i+1}, \ldots, Z_n)]\big|$$

$$\leqslant \sum_{i=1}^n \max_{t_j \in \mathbb{R}} \big|\mathbb{E}[f(t_1, \ldots, t_{i-1}, Z_i, t_{i+1}, \ldots, t_n) - f(t_1, \ldots, t_{i-1}, X_i, t_{i+1}, \ldots, t_n)]\big|.$$

Let $g_{\vec{t}}(y) = f(t_1, \ldots, t_{i-1}, y, t_{i+1}, \ldots, t_n)$ and note via Taylor's theorem (with integral form for the remainder) that

$$\left|g_{\vec{t}}(y) - g_{\vec{t}}(0) - y \cdot g'_{\vec{t}}(0) - \frac{y^2}{2} g''_{\vec{t}}(0)\right| \leqslant y^3 \cdot \sup_{z \in \mathbb{R}} g'''_{\vec{t}}(z).$$

The result follows immediately. $\qquad\square$

We now prove our main replacement result.

Lemma 8.5. *Let $M^* \in \mathcal{E}^*$, let $\varepsilon \geqslant e^{-cn}$ and let $f \in \{f^+, f^-\}$ be as above. Then*

$$\left| \mathbb{E}_X f(H) - \mathbb{E}_K f(K) \right| \leqslant \varepsilon \cdot n^{-\Omega(1)}, \tag{52}$$

Proof. We write $\xi = (\xi_1, \ldots, \xi_\ell)$, and bound the left hand side of (52) above by

$$\int_{(\theta,\xi) \in \mathbb{R}^{\ell+1}} |\widehat{f}(\theta, \xi)| \cdot \left| \mathbb{E}\, e^{-2\pi i(H_0 \theta + \langle \xi, H \rangle)} - \mathbb{E}\, e^{-2\pi i(\theta K_0 + \langle \xi, K \rangle)} \right| d\theta d\xi, \tag{53}$$

by using Fourier inversion and the triangle inequality. We now define the rectangular region

$$\Omega = \left\{ (\theta, \xi) \colon |\theta| \leqslant \varepsilon^{-1}(\log n)^{28} \text{ and } \|\xi\|_\infty \leqslant (\log n)^{30} \right\},$$

and truncate (53) to Ω. In particular, using the decay condition on $\widehat{f}$, we see that (53) is

$$\lesssim \varepsilon \cdot e^{2(\log n)^{2/3}} \cdot \int_{(\theta,\xi) \in \Omega} \left| \mathbb{E}\, e^{-2\pi i(H_0 \theta + \langle \xi, H \rangle)} - \mathbb{E}\, e^{-2\pi i(\theta K_0 + \langle \xi, K \rangle)} \right| d\theta d\xi + \varepsilon n^{-\omega(1)}. \tag{54}$$

We now write

$$\left| \mathbb{E}_X\, e^{-2\pi i(H_0 \theta + \langle \xi, H \rangle)} - \mathbb{E}_G\, e^{-2\pi i(\theta K_0 + \langle \xi, K \rangle)} \right| = \left| \psi(\theta, \xi) - \widetilde{\psi}(\theta, \xi) \right|$$

and address the integral in (54) in two different ranges.

In the case $|\theta| \leqslant (\log n)^{30}$, we look to apply Theorem 8.4, for each fixed (θ, ξ) to bound the integrand. Write $x = (x_1, \ldots, x_n)$ and $u = u(\theta, \xi) = \theta v + \sum_i \xi_i v_i$ and define

$$g(x) = g_{\theta,\xi}(x) = \exp\left(-2\pi i \left\langle x, \theta v + \sum_{i=1}^{\ell} \xi_i v_i \right\rangle \right) = \exp\left(-2\pi i \langle x, u \rangle \right).$$

so that

$$\left| \psi(\theta, \xi) - \widetilde{\psi}(\theta, \xi) \right| = \left| \mathbb{E}_X\, g(X) - \mathbb{E}_G\, g(G) \right|. \tag{55}$$

Now note that

$$\|u\|_2^2 = \left\langle \theta v + \sum_{i=1}^{\ell} \xi_i v_i, \theta v + \sum_{i=1}^{\ell} \xi_i v_i \right\rangle = \|\theta\|_2 + \sum_{i=1}^{\ell} |\xi_i|_2^2 \leqslant (\log n)^{60} \ell.$$

Since $|\theta| \leqslant (\log n)^{30}$ and $M^* \in \mathcal{E}$ we have that

$$\|u\|_\infty = \left\| \theta v + \sum_{i=1}^{\ell} \xi_i v_i \right\|_\infty \leqslant |\theta| \|v\|_\infty + \sum_{i=1}^{\ell} |\xi_i| \|v_i\|_\infty \leqslant n^{-\Omega(1)}.$$

So we apply Theorem 8.4 to see that

$$\left| \mathbb{E}_X\, g(X) - \mathbb{E}_G\, g(G) \right| \lesssim \sum_{i=1}^{n} \left\| \frac{d^3 g}{dx_i^3} \right\| \leqslant \sum_{i=1}^{n} u_i^3 \leqslant \|u\|_\infty \|u\|_2 \leqslant n^{-\Omega(1)}.$$

Thus

$$\int_{(\theta,\xi) \in \Omega, \, |\theta| \leqslant (\log n)^{30}} \left| \psi(\theta, \xi) - \widetilde{\psi}(\theta, \xi) \right| d\theta d\xi \leqslant (\log n)^{28\ell + 30} n^{-\Omega(1)} = n^{-\Omega(1) + o(1)}.$$

We now deal with the case $|\theta| \geqslant (\log n)^{30}$, write $\Omega^+ = \Omega \cap \{|\theta| \geqslant (\log n)^{30}\}$, and show that the contributions from both $\psi, \widetilde{\psi}$ are small. First note

$$\int_{(\theta,\xi) \in \Omega^+} |\psi(\theta, \xi)| = \int_{(\theta,\xi) \in \Omega^+} |\mathbb{E}\, e^{-2\pi i(K_0 \theta + \langle \xi, K \rangle)}| \leqslant \int_{(\theta,\xi) \in \Omega^+} \exp(-\Omega(\|\xi\|_2^2 + |\theta|^2)) \leqslant n^{-\omega(1)},$$

which is small enough for us. To bound the contribution from ψ, we use that

$$I = \int_{(\theta,\xi)\in\Omega^+} |\psi(\theta,\xi)|\, d\theta d\xi \leqslant \int_{(\theta,\xi)\in\Omega} \exp\left(-c\left\|\xi v + \sum_{i=1}^{\ell} \xi_i v_i\right\|_{\mathbb{T}}^2\right) d\theta d\xi. \qquad (56)$$

Now note that since $|\theta| \leqslant \varepsilon^{-1} \cdot (\log n)^{30} < e^{cn}$ and thus $\|v\theta\|_{\mathbb{T}} \gtrsim \min\{|\theta|, \sqrt{n}\}$, by property $\mathcal{E}_{\mathrm{lcd}}$. Therefore

$$\left\|\theta v + \sum_{i=1}^{\ell} \xi_i v_i\right\|_{\mathbb{T}} \geqslant \|\theta v\|_{\mathbb{T}} - \left\|\sum_{i=1}^{\ell} \xi_i v_i\right\|_2 \geqslant \|\theta v\|_{\mathbb{T}} - (\log n)^{28} \cdot \ell \gtrsim \min\{|\theta|, \sqrt{n}\}$$

and therefore the integral on the left hand side of (56) is $I \leqslant n^{-\omega(1)}$, which again is small. $\qquad\square$

8.4. Proof of Lemma 8.1. We now finish the proof of Lemma 8.1.

Proof of Lemma 8.1. Apply the properties guaranteed by Lemma 8.3 and Lemma 8.5 to get

$$\mathbb{P}_X(\mathcal{F}(X)) \leqslant \mathbb{E}_X\, f^+(H) + O(\delta_n\varepsilon) = \mathbb{E}_G\, f^+(K) + O(\delta_n\varepsilon) = \mathbb{E}\, f^-(K) + O(\delta_n\varepsilon) \leqslant \mathbb{P}(\mathcal{G}(G)) + O(\delta_n\varepsilon),$$

thus proving (47). The proof of (48) is similar. $\qquad\square$

9. RETURN TO THE SINGULAR VALUE PROBLEM AND PROOF OF THEOREM 1.2

In this section we prove our main theorem, Theorem 1.2. We do this by first proving the following lemma, which, relates our main quantity of interest to the probability that the least singular value of an $n \times n$ *Gaussian* matrix is small. Together with our work from earlier sections and the work of Edelman [5], this will allow us to complete the proof of our main theorem.

Lemma 9.1. *For $0 \leqslant \varepsilon < n^{-c}$, we have*

$$\mathbb{E}_{M^*}\, \mathbb{P}_Z\left(|\langle v, Z\rangle| \leqslant \varepsilon n^{-1/2} \cdot \chi(Z)\right) = \left(1 + O(\delta_n)\right)\mathbb{P}_G\left(\sigma_n(G) \leqslant \varepsilon n^{-1/2}\right) + O(\delta_n\varepsilon). \qquad (57)$$

To prove this we use the symmetry and smoothness of the distribution of Z to show that we can "replace" the small ε on the left hand side of (57) with a much larger (and therefore more tractable) $\varepsilon_0 = n^{-c}$. This step we described at (17) in the proof sketch.

Lemma 9.2. *Let $0 \leqslant \varepsilon \leqslant n^{-c}$, let $\varepsilon_0 \geqslant n^{-c}$ and let $\sigma_{n-1}(M^*) \geqslant (\log n)^{-1} n^{-1/2}$. Then*

$$\mathbb{P}_Z\left(|\langle v, Z\rangle| \leqslant \varepsilon n^{-1/2} \cdot \chi(Z)\right) = (\varepsilon/\varepsilon_0) \cdot \mathbb{P}_Z\left(|\langle v, Z\rangle| \leqslant \varepsilon_0 n^{-1/2} \chi(Z)\right) + O(\delta_n\varepsilon).$$

Proof. Since $Z \sim N(0,1)^{\oplus n}$ and the v_i are orthonormal, we can express $|\langle v, Z\rangle| \leqslant \varepsilon n^{-1/2} \cdot \chi(Z)$ as

$$|W_0| \leqslant \varepsilon n^{-1/2}\left(\sum_{i=1}^{\ell} \frac{W_i^2}{\sigma_{n-i}(M^*)^2}\right)^{\frac{1}{2}} =: f(W_1, \dots W_\ell)$$

where $(W_0, W_1, \dots, W_\ell) \sim N(0,1)^{\oplus(\ell+1)}$. Now note that, since W_0 is independent of f and since the Gaussian density function is flat about its mean, we have that

$$\mathbb{P}\left(W_0 \leqslant f \mid f < (\log n)^{-1}\right) = (\varepsilon/\varepsilon_0)(1 + O(\delta_n)) \cdot \mathbb{P}(W_0 \leqslant (\varepsilon_0/\varepsilon)f \mid f < (\log n)^{-1}).$$

So to finish we only need to show that $\mathbb{P}(f \geqslant (\log n)^{-1}) = O(\delta\varepsilon)$. For this, simply note that under the event $\sigma_{n-1}(M^*) \geqslant (\log n)^{-1} n^{-1/2}$ we have that

$$\mathbb{P}\left(f \geqslant (\log n)^{-1}\right) \leqslant \mathbb{P}\left(\varepsilon(\log n)\left(\sum_{i=1}^{\ell} W_i^2\right)^{\frac{1}{2}} \geqslant (\log n)^{-1}\right) = O(\delta_n\varepsilon),$$

where the last equality follows by standard concentration estimates and the assumption that $\varepsilon < n^{-c}$. Taking expectations $\mathbb{E}_{Z_1,\ldots,Z_\ell}$ finishes the proof of the lemma. $\qquad\square$

We now properly state the universality theorem of Tao and Vu, for the least singular value. The following is Theorem 6.7 in [27].

Theorem 9.3. *Let $\widetilde{M}$ be an $n \times n$ random matrix, where $\widetilde{M}_{ij}$ are independent with mean 0, variance 1 and subgaussian constants $\|\widetilde{M}_{ij}\|_{\psi_2} \leqslant B$. Then*

$$\mathbb{P}\big(\sigma_n(\widetilde{M}) \leqslant \varepsilon n^{-1/2}\big) = \mathbb{P}(\sigma_n(G) \leqslant \varepsilon n^{-1/2}) + O(n^{-c}),$$

for some $c > 0$, depending only on B.

We also need the following theorem of Edelman [5], mentioned in the introduction.

Theorem 9.4. *Let G be an $n \times n$ random matrix with iid entries distributed as $N(0,1)$. Then*

$$\mathbb{P}\big(\sigma_n(G) \leqslant \varepsilon n^{-1/2}\big) \leqslant \varepsilon.$$

We are now in a position to prove Lemma 9.1.

Proof of Lemma 9.1. Write $\mathcal{F}_\varepsilon = \{|\langle v, Z\rangle| \leqslant \varepsilon n^{-1/2} \cdot \chi(Z)\}$. Using Lemma 9.2 with the right hand side of (57) with $\varepsilon_0 = n^{-c}$, we have

$$\mathbb{E}_{M^*}\mathbb{P}_Z\big(\mathcal{F}_\varepsilon\big) = (\varepsilon/\varepsilon_0) \cdot \mathbb{E}_{M^*}\mathbb{1}\big(M^* \in \mathcal{E}^*\big) \cdot \mathbb{P}_Z\big(\mathcal{F}_{\varepsilon_0}\big) + O(\delta_n \varepsilon).$$

Here we have used Lemma 8.2 to deal with the case that $\sigma_{n-1}(M^*) < (\log n)^{-1} n^{-1/2}$. Now apply Lemma 4.1 to see that

$$\mathbb{P}_{\widetilde{M}}\big(\sigma_n(\widetilde{M}) \leqslant \big(1-\delta_n\big)\varepsilon_0 n^{-1/2}\big) + O(\delta_n \varepsilon) \leqslant \mathbb{E}_{M^*}\mathbb{P}_Z\big(\mathcal{F}_{\varepsilon_0}\big) \leqslant \mathbb{P}_{\widetilde{M}}\big(\sigma_n(\widetilde{M}) \leqslant \big(1+\delta_n\big)\varepsilon_0 n^{-1/2}\big) + O(\delta_n \varepsilon),$$

where $\widetilde{M}$ is the random matrix M^* with a Gaussian row Z added. We now apply Theorem 9.3 to the right hand side of the above to see

$$\mathbb{P}_{\widetilde{M}}\big(\sigma_n(\widetilde{M}) \leqslant \big(1+\delta_n\big)\varepsilon_0 n^{-1/2}\big) = \mathbb{P}_G(\sigma_n(G) \leqslant (1+\delta_n)\varepsilon_0 n^{-1/2}) + O(n^{-c}) = \big(1 + O(\delta_n)\big)\varepsilon_0,$$

and similarly for the left hand side. Now apply Theorem 9.4 to obtain the desired result. $\qquad\square$

9.1. **Proof of Theorem 1.2 and Theorem 1.1.** Here we prove our main theorem, Theorem 1.2, by snapping together the results developed in the previous sections.

Proof of Theorem 1.2 . We prove the "$\leqslant$" direction of our main theorem and note that the other direction is similar. We may assume that $e^{-cn} < \varepsilon < n^{-c}$, otherwise we can apply the results of Rudelson and Vershynin or Tao and Vu, respectively. We first apply Lemma 4.1 and write

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2}\big) \leqslant \mathbb{E}_{M^*}\mathbb{P}_X\big(|\langle v, X\rangle| \leqslant \big(1+\delta_n\big)\varepsilon n^{-1/2}\chi(X)\big) + O(\delta_n \varepsilon).$$

Now apply our Gaussian replacement step, Lemma 8.1, to write

$$\mathbb{E}_{M^*}\mathbb{P}_X\big(|\langle v, X\rangle| \leqslant \varepsilon n^{-1/2}\chi(X)\big) \leqslant \mathbb{E}_{M^*}\mathbb{P}_Z\big(|\langle v, Z\rangle| \leqslant \big(1+\delta_n\big)\varepsilon n^{-1/2} \cdot \chi(Z)\big) + O(\delta_n \varepsilon),$$

where we used Lemma 8.2, to deal with the event $M^* \notin \mathcal{E}^*$. Now we apply Lemma 9.1 to write

$$\mathbb{E}_{M^*}\mathbb{P}_Z\Big(|\langle v, Z\rangle| \leqslant \big(1+\delta_n\big)\varepsilon n^{-1/2} \cdot \chi(Z)\Big) = \big(1 + O(\delta_n)\big)\mathbb{P}_G\big(\sigma_n(G) \leqslant \big(1+\delta_n\big)\varepsilon n^{-1/2}\big) + O(\delta_n \varepsilon),$$

thus completing the proof. $\qquad\square$

We now easily derive our approximate Spielman–Teng conjecture.

Proof of Theorem 1.1. Apply Theorem 1.2 and Theorem 9.4 and write

$$\mathbb{P}\big(\sigma_n(M) \leqslant \varepsilon n^{-1/2}\big) \leqslant \big(1 + o(\delta_n)\big)\mathbb{P}\big(\sigma_n(G) \leqslant \varepsilon n^{-1/2}\big) + e^{-\Omega(n)} \leqslant \big(1 + \delta_n\big)\varepsilon + e^{-\Omega(n)},$$

where G is a random $n \times n$ matrix with iid standard normal entries. $\qquad\square$

Appendix A. The flatness of the v_i

We first require which follows from interlacing and the proof of [27, Lemma 4.1].

Lemma A.1. *There exist $c \in (0, 1/4)$ and $c' > 0$ (potentially depending on the subgaussian constant of ξ) such that with probability $1 - \exp(-c'n^c)$, there are at least $c'n^{1-c}$ singular values of M_n in $[n^{1/2-c}/2, n^{1/2-c}]$.*

We now prove Lemma 7.5.

Proof of Lemma 7.5. Let c be as in lemma A.1; and suppose that there exists $v \in \mathbb{S}^n$ with $\|v\|_\infty \geqslant n^{-c/4}$ and $\|M_n^* v\|_2 \leqslant 1$. Applying the union bound by symmetry we may assume that $|v_n| \geqslant n^{-c/4}$ (at the cost of paying a negligible factor n in the probability). Let $\widetilde{M}$ denote the first $n - 1$ columns of M_n^*, $\widetilde{X}$ denote the last column of M_n^*, and $v' = (v_1, \ldots, v_{n-1})$. Since $\|\widetilde{M}\|_{\mathrm{op}} \leqslant 5\sqrt{n}$ with exponentially good probability by [3, Theorem C.1], if $\|M_n^* v\|_2 \leqslant 1$ then

$$\|\widetilde{M}^T(\widetilde{M}v' + \widetilde{X}v_n)\|_2 \leqslant 5\sqrt{n}.$$

Let F denote the set of singular values of size between $[2^{-2}n^{1/2-c}, 2n^{1/2-c}]$. By lemma A.1 applied to $\widetilde{M}$, with very high probability, we have $|F| \geqslant c'n^{1-c}$. Letting π_F denote the projection onto the corresponding left-singular vectors of $\widetilde{M}$ and note that

$$\|\pi_F \widetilde{M}^T(\widetilde{M}v' + \widetilde{X}v_n)\|_2 = \|\pi_F \widetilde{M}^T \widetilde{M}v' + \pi_F \widetilde{M}^T \widetilde{X}v_n\|_2 \leqslant 5\sqrt{n}.$$

As $v_n \geqslant n^{-c/4}$, we have

$$\|\pi_F \widetilde{M}^T \widetilde{X}\|_2 \leqslant v_n^{-1}(\|\pi_F \widetilde{M}^T \widetilde{M}v'\|_2 + 5\sqrt{n}) \leqslant 5n^{1-7c/4}.$$

Furthermore note that

$$\mathbb{E}[\|\pi_F \widetilde{M}^T \widetilde{X}\|_2^2] = \mathbb{E}[\mathrm{tr}(\pi_F \widetilde{M}^T \widetilde{X} \widetilde{X}^T \widetilde{M} \pi_F^T)] = n\,\mathrm{tr}(\pi_F \widetilde{M}^T \widetilde{M} \pi_F^T) \gtrsim c'n^{2-3c}.$$

The desired result then follows via the Hanson–Wright inequality (see [22, Theorem 2.1]). $\qquad\square$

Theorem B.1. *Let $u \in \mathbb{R}^n$ satisfy $\|u\|_2 = 1$ and let $w_1, \ldots, w_k \in \mathbb{R}^n$ be orthogonal unit vectors. Furthermore let $Y = (Y_1, \ldots, Y_n)$ be a random vector where the Y_i are iid, mean 0 and subgaussian. If $\mathrm{LCD}(u) \geqslant C/\varepsilon$ and $k \cdot \|u\|_\infty + \sum_{i=1}^k \|w_i\|_\infty \leqslant (\log n)^{-3}$ then*

$$\mathbb{P}\Big(|\langle u, Y \rangle| \leqslant \varepsilon \wedge \sum_{j=1}^k |\langle u_i, Y \rangle|^2 \leqslant ck \Big) \leqslant C(\varepsilon e^{-ck} + e^{-cn}),$$

where $C, c > 0$ depend only on the subgaussian constant $\|Y_i\|_{\psi_2}$.

For this, we shall need the following basic estimate. If $Y = (Y_1, \ldots, Y_n)$ is a random vector where Y_i are iid and subgaussian and $u \in \mathbb{R}^n$, we have

$$\mathbb{E}_Y \exp(i\langle u, Y \rangle) \leqslant \exp\Big(-c_0 \inf_{r \in [1, c_0^{-1}]} \|rv\|_{\mathbb{T}}^2 \Big), \tag{58}$$

where $c_0 > 0$ depends only on the random variable Y_1.

Proof. We may assume that $\varepsilon \leqslant e^{-ck/2}$; else adjusting c the result follows immediately from the Hanson–Wright inequality (see [22, Corollary 3.1]). Furthermore we may assume that $\varepsilon \geqslant e^{-c\alpha n/2}$ by adjusting c appropriately. Furthermore note that we may assume that n is a sufficiently large absolute constant else the estimate is vacous. For the remainder of the argument, let c be a sufficiently small constant to be chosen later. We first bound

$$\mathbb{P}\Big(|\langle u, Y \rangle| \leqslant \varepsilon \wedge \sum_{j=1}^k |\langle w_i, Y \rangle|^2 \leqslant ck \Big) \leqslant 2^k \max_{|S|=k/2} \mathbb{P}\Big(|\langle u, Y \rangle| \leqslant \varepsilon \wedge \bigwedge_{j \in S} |\langle w_i, Y \rangle| \leqslant 2\sqrt{c} \Big). \tag{59}$$

We now approximate the event on the right hand side using the smooth bump function

$$f(t) = (\sin(2\pi t)/(\pi t))^2 \quad \text{for which} \quad \widehat{f}(\theta) = (\mathbb{1}_{[-1,1]} * \mathbb{1}_{[-1,1]})(\theta).$$

Assume, without loss, that $S = \{1, \ldots, k/2\}$. Note that $f(t) \geqslant 1$ for $t \leqslant 1/4$. Thus we have

$$\mathbb{P}\Big(|\langle u, Y \rangle| \leqslant \varepsilon \wedge \bigwedge_{j \in S} |\langle w_i, Y \rangle| \leqslant 2\sqrt{c} \Big) \leqslant \mathbb{E}\, f\big(\langle u, Y \rangle (4\varepsilon)^{-1}\big) \prod_{j \in S} f\big(\langle w_i, Y \rangle (64c)^{-1/2}\big). \tag{60}$$

Now note that the characteristic function of the random variable $\big(\langle v, Y \rangle, \langle w_1, Y \rangle, \ldots, \langle w_{k/2}, Y \rangle\big)$ is

$$\varphi(\theta, \xi_1, \ldots, \xi_{k/2}) = \mathbb{E} \exp\Big(i\theta \langle X, v \rangle + \sum_{1 \leqslant j \leqslant k/2} i\xi_j \langle w_j, Y \rangle \Big).$$

After putting $\xi = (\xi_1, \ldots, \xi_{k/2})$, we may rewrite the expression on the right hand side of (60) using the Fourier transform as

$$\int_{\mathbb{R}^{k/2+1}} 4\varepsilon \widehat{f}\big(4\varepsilon\theta\big) \prod_{j \in S} \big((64c)^{1/2} \cdot \widehat{f}(8\sqrt{c}\xi_j)\big) \cdot \varphi(\theta, \xi)\, d\theta d\xi \leqslant \varepsilon 16^{k+1} c^{k/4} \int_{(\theta, \xi) \in B} |\varphi(\theta, \xi)|\, d\theta d\xi, \tag{61}$$

where we define $B = \{(\theta, \xi) \colon |\theta| \leqslant \varepsilon^{-1}, |\xi_j| \leqslant c^{-1/2}, \forall j\}$ and used that $\widehat{f}$ is supported on $[-2, 2]$.

We now bound the integral in (61) in two regimes, based on $|\theta|$. If $|\theta| \leqslant k \cdot \log n$ we define

$$u(\theta) = \theta u + \sum_{j \leqslant k/2} \xi_j w_j \quad \text{and note} \quad \|u(\theta)\|_\infty \leqslant (\log n)^{-1},$$

22

by using the bounds on k, $\|u\|_\infty$ and $\|w_i\|_\infty$. Thus, $\|r \cdot u(\theta,\xi)\|_\mathbb{T} = \|r \cdot u(\theta,\xi)\|_2$ for all scalars $r < (\log n)/2$. Thus, using the inequality in (58), we have

$$|\varphi(\theta,\xi)| = \left|\mathbb{E}_Y \exp\left(i\langle u(\theta,\xi), Y\rangle\right)\right| \leqslant \exp\left(-c_0 \inf_{r \in [1,c_0^{-1}]} \|ru(\theta,\xi)\|_\mathbb{T}^2\right) \leqslant \exp\left(-c_0|\theta|^2 - c_0\|\xi\|_2^2\right),$$

where $c_0 > 0$ is a constant depending only on the distribution of Y_1 and were we have used that $u,,w_1,\ldots,w_k$ are orthonormal for the last inequality. As a result, if we set $B^- = \{(\theta,\xi) \in B \colon |\theta| \leqslant k\log n\}$, we have

$$\int_{B^-} |\varphi(\theta,\xi)|\, d\theta d\xi = \exp(O(k)), \tag{62}$$

where the implicit constant depends only on the distribution of Y_i.

We now consider the other range of θ, where $k\log n \leqslant |\theta| \leqslant \varepsilon^{-1}$. We have that

$$\left\|u(\theta,\xi)\right\|_\mathbb{T} \geqslant \|\theta v\|_\mathbb{T} - \left\|\sum_j \xi_j w_j\right\|_2 \geqslant \min\{\gamma\|\theta v\|_2, \sqrt{\alpha n}\} - k^{1/2}c^{-1/2} \geqslant \min\{\gamma\|\theta v\|_2, \sqrt{\alpha n}\}/2,$$

where the second inequality holds by the condition on the last common denomonator of u and the L_∞ bound on the w_i. The last inequality holds when n is sufficiently large with respect to γ.

We now use this information along with the inequality in (58) to write

$$|\varphi(\theta,\xi)| \leqslant \exp\left(-c_0 \inf_{r \in [1,c_0^{-1}]} \|r \cdot u(\theta,\xi)\|_\mathbb{T}^2\right) \leqslant \exp(-\Omega(n)).$$

Integrating this expression over $B^+ = \{(\theta,\xi) \colon k\log n \leqslant |\theta| \leqslant \varepsilon^{-1}\}$ gives

$$\int_{B^+} |\varphi(\theta,\xi)|\, d\theta d\xi \leqslant |B^+| \cdot e^{-\Omega(n)} \leqslant \varepsilon^{-1}(1/c)^{O(k)}e^{-\Omega(n)} = o(1). \tag{63}$$

Here we have used that $\varepsilon > e^{-c\alpha n}$

We now put (59) together with (61) and our estimates on the integral (62) and (63) to get

$$\mathbb{P}\left(|\langle u, Y\rangle| \leqslant \varepsilon \wedge \sum_{j=1}^k |\langle w_i, Y\rangle|^2 \leqslant ck\right) \leqslant \varepsilon c^{k/4}e^{\Omega(k)}.$$

Thus choosing $c > 0$ to be sufficiently small completes the proof. $\qquad\square$

Appendix C. Construction of smooth bump functions

We will need the construction of a bump function with sufficiently fast Fourier decay.

Lemma C.1. *There exists a function $\psi \colon \mathbb{R} \to \mathbb{R}$ such that $\psi \geqslant 0$, $\hat\psi \geqslant 0$, $\mathrm{supp}(\hat\psi) \in [-1,1]$, $\int_\mathbb{R} \psi(x)\, dx = 1$, and $\psi(x) \leqslant \exp(-c(|x|+1)^{1/2})$, where $c > 0$ is an absolute constant.*

Proof. Omitting the first and fourth bullet point, $\widehat{\psi^*}(x) = \exp(-(1-4x^2)^{-1})\mathbb{1}_{|x|\leqslant 1/2}$ is sufficient. The necessary Fourier decay property is proved in [12]. Taking $\hat\psi(x) = (\widehat{\psi^*} * \widehat{\psi^*})(x)$ and normalizing fixes the remaining conditions. $\qquad\square$

We now work towards defining the bump functions that are required Lemma 8.3. To construct these functions, let $(x, y_1, \ldots, y_\ell) \in \mathbb{R}^{\ell+1}$; and let $y = (y_1, \ldots, y_\ell)$. Define the functions

$$\chi(y) = \chi_{M^*} = \left(\sum_{i=1}^\ell \frac{y_i^2}{\sigma_{n-i}(M^*)^2}\right)^{1/2} \quad \text{and} \quad B_r(x,y) = \mathbb{1}\left((1+r\delta_n)\cdot \varepsilon n^{-1/2}\cdot \chi(y) \geqslant x\right).$$

23

Here we will only use $|r| \leqslant 1$. Note that the probabilities in Lemma 8.1 can be expressed as

$$\mathbb{E}_H B_1(H) = \mathbb{P}_X\big(|\langle v, X\rangle| \leqslant (1+\delta_n)\varepsilon n^{-1/2}\cdot\chi(X)\big); \mathbb{E}_K B_{-1}(K) = \mathbb{P}_X\big(|\langle v, X\rangle| \leqslant (1-\delta_n)\varepsilon n^{-1/2}\cdot\chi(X)\big).$$

We now truncate the indicators $B_r(x,y)$ in two ways. Define $\tau = (\log n)^7$, $\tau' = (\log n)^{25}$ and

$$Q(x,y) = \mathbb{1}\big(\|y\|_\infty \leqslant \log n \wedge |x| \leqslant \varepsilon \cdot (\log n)^5\big) \quad \text{and} \quad C(x,y) = \mathbb{1}\Big(\sum_{i \leqslant (\log\log n)^2} y_i^2 \geqslant 1\Big).$$

We now define the "smoothed rectangle"

$$\rho(x, y_1, \ldots, y_\ell) = (\varepsilon^{-1} \cdot \tau') \cdot \psi(x\tau'\varepsilon^{-1}) \prod_{i=1}^\ell \tau' \cdot \psi(y_i \cdot \tau')$$

where ψ is as in Lemma C.1. We then define our functions f^+, f^- as

$$f^+(y) = (B_{r+\tau^{-1}}(x,y)Q(x,y)C(x,y)) * \rho \quad \text{and} \quad f^-(y) = (B_{r-\tau^{-1}}(x,y)Q(x,y)C(x,y)) * \rho.$$

We now prove Lemma 8.3 by checking that the functions f^+, f^- satisfy the desired properties. It is easiest to check the final property in Lemma 8.3 first.

Claim C.2. *For $f \in \{f^-, f^+\}$ and all $\theta \in \mathbb{R}$, $\xi = (\xi_1, \ldots, \xi_\ell) \in \mathbb{R}^\ell$, we have*

$$|\widehat{f}(\theta, \xi)| \leqslant \varepsilon \cdot \exp\big((\log n)^{2/3}\big) \exp\big(-c(\log n)^{-26}\big(|\theta/\varepsilon|^{1/2} + |\xi_1|^{1/2} + \cdots + |\xi_\ell|^{1/2}\big)\big),$$

where $c > 0$ is an absolute constant.

Proof. We check the claim for $f = f^+$ and note the case $f = f^-$ is similar. We have that

$$|\widehat{f^+}(\theta, \xi)| \leqslant \|B_{r+\tau^{-1}}(x,y)Q(x,y)C(x,y)\|_{L^1(\mathbb{R}^{\ell+1})} \cdot |\widehat{\rho}(\theta,\xi)| \leqslant \|Q(x,y)\|_{L^1(\mathbb{R}^{\ell+1})} \cdot |\widehat{\rho}(\theta,\xi)|,$$

where we used $\widehat{f * g} = \widehat{f} \cdot \widehat{g}$ and that $\|\widehat{f}\|_\infty \leqslant \|f\|_1$. Recalling the definition of Q, the above is

$$\leqslant \varepsilon \cdot (\log n)^{\ell+10} \cdot |\widehat{\rho}(\theta,\xi)| \leqslant \varepsilon \cdot \exp\big((\log n)^{2/3}\big) \exp\big(-c(\log n)^{-26}\big(|\theta/\varepsilon|^{1/2} + |\xi_1|^{1/2} + \cdots + |\xi_\ell|^{1/2}\big)\big),$$

where we have used the definition of ρ for the last inequality. $\square$

For the next two claims we use the following point-wise inequalities.

$$B_r(x,y)Q(2x,2y)C(x/2,y/2) \leqslant f^+(y) \leqslant B_{r+2\tau^{-1}}(x,y)Q(x/2,y/2)C(2x,2y) + E(x,y) \quad (64)$$

and

$$(1 - n^{-\omega(1)})B_{r-2\tau^{-1}}(x,y)Q(2x,2y)C(x/2,y/2) \leqslant f^-(y) \leqslant B_r(x,y)Q(x/2,y/2)C(2x,2y) + E(x,y). \quad (65)$$

where $E(x,y)$ is the "error" function

$$E(x,y) = \sum_{t \geqslant (\log n)^4} e^{-t^{1/3}} Q(x/t, y/t).$$

Claim C.3. *Let $G \sim N(0,1)^{\otimes n}$. We have $|\mathbb{E}_G f^+(G) - f^-(G)| = O(\delta_n\varepsilon)$.*

Proof. Recall that $K = (\langle v, X\rangle, \langle v_{n-1}, X\rangle, \ldots, \langle v_{n-\ell}, X\rangle)$ which is distributed as $N(0,1)^{\otimes(\ell+1)}$, since the v_i are orthonormal. Now note that

$$\mathbb{E}\, Q(K) \lesssim \varepsilon \cdot (\log n)^4 \quad \text{and similarly} \quad \mathbb{E}\, E(K) \leqslant \varepsilon \cdot n^{-\omega(1)}, \quad (66)$$

via standard Gaussian estimates. So from (64), (65) and (66) it is enough to show

$$\mathbb{E}\,\big|B_{r+2\tau^{-1}}(K)Q(K/2)C(2K) - B_{r-2\tau^{-1}}(K)Q(2K)C(K/2)\big| = O(\delta_n\varepsilon).$$

24

Towards this goal, we note that the left hand side is at most

$$\mathbb{E}\big|Q(K/2)-Q(2K)\big|+\mathbb{E}\big|Q(2K)(C(2K)-C(K/2))\big|+\mathbb{E}\big|Q(2K)C(K/2)(B_{r+2\tau^{-1}}(K)-B_{r-2\tau^{-1}}(K))\big|.$$

The first two terms are $O(\delta_n\varepsilon)$ by a direct Gaussian computation. So the above is

$$\leqslant \mathbb{P}\big(|\langle v,G\rangle - r\chi(K)|\leqslant 2\tau^{-1}\delta_n\varepsilon n^{-1/2}\chi(K)\big)+O(\delta_n\varepsilon)\leqslant 2\tau^{-1}\delta_n\cdot\varepsilon n^{-1/2}\mathbb{E}_K\chi(K)+O(\delta_n\varepsilon).$$

Now since $M^*\in\mathcal{E}^*$, we have that $\sigma_{n-i}\geqslant (\log n)^{-3}n^{-1/2}$ and thus

$$\mathbb{E}_K\chi(K)\leqslant \big(\mathbb{E}_K\chi(K)^2\big)^{\frac{1}{2}}\leqslant n^{1/2}(\log n)^3\ell^{1/2}.$$

Using this in the above completes the proof of the claim. $\qquad\qquad\square$

We now complete the proof of Lemma 8.3 with the following claim.

Claim C.4. *For $f\in\{f^+,f^-\}$, we have*

$$\mathbb{P}_X(\mathcal{F}(X))\leqslant \mathbb{E}_X\, f(H)+O(\delta_n\varepsilon)\qquad\text{and}\qquad \mathbb{E}_G\, f(K)\leqslant \mathbb{P}_G(\mathcal{F}(G))+O(\delta_n\varepsilon).\qquad (67)$$

Proof. Here we prove the first inequality in (67) for $f=f^+$ and note that the other cases are similar or easier. Note it is enough to show that

$$\mathbb{E}[f^+(H)-B_r(H)]\geqslant -\varepsilon(\log n)^{-\omega(1)}\qquad\text{and}\qquad \mathbb{E}[B_r(H)-f^-(H)]\geqslant -\varepsilon(\log n)^{\omega(1)}.$$

For the first inequality in the above, we write

$$\mathbb{E}[f^+(H)-B_r(H)]\geqslant \mathbb{E}[B_r(H)(Q(2H)C(H/2)-1)],$$

and note that

$$\chi(y)\leqslant (\sigma_{n-1}(M^*))^{-1}\ell\max_{i\leqslant\ell}|y_i|\leqslant (\log n)^{11/4}\max_{i\leqslant\ell}|y_i|,$$

where we used that $M^*\in\mathcal{E}^*$ for the second inequality and thus

$$B_r(x,y)\leqslant \mathbb{1}\big(\varepsilon(\log n)^{11/4}\max_{1\leqslant i\leqslant\ell}|y_i|\geqslant x\big).$$

Thus if $B_r(H)=1$ and $|\langle v,X\rangle|\geqslant \varepsilon\cdot(\log n)^5/2$ we have that $\max_{1\leqslant i\leqslant\ell}|\langle H,v\rangle|\geqslant (\log n)$. Therefore
$\mathbb{E}[f^+(H)-B_r(H)]\geqslant \mathbb{E}[B_r(H)(Q(2H)C(H/2)-1)]$

$$\geqslant -\mathbb{E}\Big[\varepsilon(\log n)^{11/4}\max_{1\leqslant i\leqslant\ell}|\langle H,v_i\rangle|\geqslant |\langle H,v\rangle|\wedge\Big(\bigvee_{1\leqslant i\leqslant\ell}|\langle H,v_i\rangle|\geqslant (\log n)/2\vee\sum_{1\leqslant i\leqslant(\log\log n)^2}|\langle H,v_i\rangle|^2\leqslant 1/2\Big)\Big]$$

$$\geqslant -\mathbb{E}\Big[\varepsilon(\log n)^{11/4}\max_{1\leqslant i\leqslant\ell}|\langle H,v_i\rangle|\geqslant |\langle H,v\rangle|\wedge\bigvee_{1\leqslant i\leqslant\ell}|\langle H,v_i\rangle|\geqslant \log n\Big]$$

$$-\mathbb{E}\Big[\varepsilon(\log n)^{15/4}\geqslant |\langle H,v\rangle|\vee\sum_{1\leqslant i\leqslant(\log\log n)^2}|\langle H,v_i\rangle|^2\leqslant 1/2\Big]$$

$$\geqslant -\varepsilon\cdot(\log n)^{-\omega(1)};$$

here we have used Theorem 5.8 and Theorem 7.4. For the corresponding lower bound, note that

$$\mathbb{E}[B_r(H)-f^-(H)]\geqslant \mathbb{E}[B_r(H)(1-Q(H/2)C(2H))]-\mathbb{E}[E(H)]\geqslant -\mathbb{E}[E(H)]$$

$$=-\sum_{t\geqslant(\log n)^4}e^{-t^{1/3}}\mathbb{E}[Q(H/t)]\geqslant -\sum_{t\geqslant(\log n)^4}e^{-t^{1/3}}\mathbb{E}[\mathbb{1}[|\langle v,H\rangle|\leqslant t\varepsilon(\log n)^4]]\geqslant -\varepsilon\cdot(\log n)^{-\omega(1)}.$$

$$\qquad\qquad\square$$

References

[1] Z. D. Bai. Circular law. *Ann. Probab.*, 25(1):494–529, 1997.

[2] J. Bourgain, V. H. Vu, and P. M. Wood. On the singularity probability of discrete random matrices. *J. Funct. Anal.*, 258(2):559–603, 2010.

[3] M. Campos, M. Jenssen, M. Michelen, and J. Sahasrabudhe. The least singular value of a random symmetric matrix. arXiv:2203.06141.

[4] M. Campos, M. Jenssen, M. Michelen, and J. Sahasrabudhe. The singularity probability of a random symmetric matrix is exponentially small. *Journal of the American Mathematical Society*, 2024.

[5] A. Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(4):543–560, 1988.

[6] A. Edelman. The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law. *J. Multivariate Anal.*, 60(2):203–232, 1997.

[7] A. Edelman, A. Guionnet, and S. Péché. Beyond universality in random matrix theory. *Ann. Appl. Probab.*, 26(3):1659–1697, 2016.

[8] L. Erdős, B. Schlein, and H.-T. Yau. Universality of random matrices and local relaxation flow. *Invent. Math.*, 185(1):75–119, 2011.

[9] V. L. Girko. The circular law. *Teor. Veroyatnost. i Primenen.*, 29(4):669–679, 1984.

[10] G. H. Golub. Some modified matrix eigenvalue problems. *SIAM Rev.*, 15:318–334, 1973.

[11] F. Götze and A. Tikhomirov. The circular law for random matrices. *Ann. Probab.*, 38(4):1444–1491, 2010.

[12] S. G. Johnson. Saddle-point integration of C_∞ "bump" functions.

[13] J. Kahn, J. Komlós, and E. Szemerédi. On the probability that a random ± 1-matrix is singular. *J. Amer. Math. Soc.*, 8(1):223–240, 1995.

[14] J. Komlós. On the determinant of $(0, 1)$ matrices. *Studia Sci. Math. Hungar.*, 2:7–21, 1967.

[15] B. Landon and P. Sosoe. Applications of mesoscopic CLTs in random matrix theory. *Ann. Appl. Probab.*, 30(6):2769–2795, 2020.

[16] M. L. Mehta. *Random matrices and the statistical theory of energy levels*. Academic Press, New York-London, 1967.

[17] H. H. Nguyen. Random matrices: overcrowding estimates for the spectrum. *J. Funct. Anal.*, 275(8):2197–2224, 2018.

[18] M. Rudelson. Invertibility of random matrices: norm of the inverse. *Ann. of Math. (2)*, 168(2):575–600, 2008.

[19] M. Rudelson. Recent developments in non-asymptotic theory of random matrices. In *Modern aspects of random matrix theory*, volume 72 of *Proc. Sympos. Appl. Math.*, pages 83–120. Amer. Math. Soc., Providence, RI, 2014.

[20] M. Rudelson and R. Vershynin. The Littlewood-Offord problem and invertibility of random matrices. *Adv. Math.*, 218(2):600–633, 2008.

[21] M. Rudelson and R. Vershynin. Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.*, 62(12):1707–1739, 2009.

[22] M. Rudelson and R. Vershynin. Hanson-Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.*, 18:no. 82, 9, 2013.

[23] A. Sankar, D. A. Spielman, and S.-H. Teng. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM J. Matrix Anal. Appl.*, 28(2):446–476, 2006.

[24] S. Smale. On the efficiency of algorithms of analysis. *Bull. Amer. Math. Soc. (N.S.)*, 13(2):87–121, 1985.

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

[26] S. a. J. Szarek. Condition numbers of random matrices. *J. Complexity*, 7(2):131–149, 1991.

[27] T. Tao and V. Vu. Random matrices: the distribution of the smallest singular values. *Geom. Funct. Anal.*, 20(1):260–297, 2010.

[28] T. Tao and V. Vu. Random matrices: universality of ESDs and the circular law. *Ann. Probab.*, 38(5):2023–2065, 2010. With an appendix by Manjunath Krishnapur.

[29] T. Tao and V. Vu. Random matrices: universality of local eigenvalue statistics. *Acta Math.*, 206(1):127–204, 2011.

[30] T. Tao and V. H. Vu. On the singularity probability of random Bernoulli matrices. *J. Amer. Math. Soc.*, 20(3):603–628, 2007.

[31] T. Tao and V. H. Vu. Inverse Littlewood-Offord theorems and the condition number of random discrete matrices. *Ann. of Math. (2)*, 169(2):595–632, 2009.

[32] K. Tikhomirov. Singularity of random Bernoulli matrices. *Ann. of Math. (2)*, 191(2):593–634, 2020.

[33] J. von Neumann. *Collected works. Vol. V: Design of computers, theory of automata and numerical analysis*. The Macmillan Company, New York, 1963. General editor: A. H. Taub.

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
Email address: asah@mit.edu

University of Cambridge. Department of Pure Mathematics and Mathematical Statistics.
Email address: jdrs2@cam.ac.uk

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
Email address: msawhney@mit.edu