

## INVERTIBILITY OF RANDOM MATRICES: UNITARY AND ORTHOGONAL PERTURBATIONS

MARK RUDELSON AND ROMAN VERSHYNIN

*In memory of Joram Lindenstrauss*

### CONTENTS

1. Introduction	294
1.1. The smallest singular values of random matrices	294
1.2. The main results	294
1.3. A word about the proofs	296
1.4. An application to the Single Ring Theorem	296
1.5. Organization of the paper	298
1.6. Notation	298
2. Strategy of the proofs	298
2.1. Unitary perturbations	298
2.2. Orthogonal perturbations	299
3. Unitary perturbations: Proof of Theorem 1.1	302
3.1. Decomposition of the problem; local and global perturbations	302
3.2. Invertibility via quadratic forms	304
3.3. When the denominator is small	306
3.4. When the denominator is large and $\ M\ $ is small	306
3.5. When $\ M\ $ is large	308
3.6. Combining the three cases	309
4. Orthogonal perturbations: Proof of Theorem 1.3	310
4.1. Initial reductions of the problem	310
4.2. Local perturbations and decomposition of the problem	310
4.3. When a minor is well invertible: Going 3 dimensions up	313
4.4. When all minors are poorly invertible: Going 1 + 2 dimensions up	320
4.5. Combining the results for well-invertible and poorly invertible minors	322
5. Application to the Single Ring Theorem: Proof of Corollary 1.4	323
Appendix A. Orthogonal perturbations in low dimensions	327
A.1. Remez-type inequalities	327
A.2. Vanishing determinant	328
A.3. Proof of Theorem 4.1	330
Appendix B. Some tools used in the proof of Theorem 1.3	334
B.1. Small ball probabilities	334
B.2. Invertibility of random Gaussian perturbations	334
B.3. Breaking complex orthogonality	336

---

Received by the editors June 22, 2012 and, in revised form, January 30, 2013.

2010 *Mathematics Subject Classification*. Primary 60B20.

The first author was partially supported by NSF grant DMS 1161372.

The second author was partially supported by NSF grant DMS 1001829.

©2013 American Mathematical Society  
Reverts to public domain 28 years from publication

Acknowledgements	337
References	337

## 1. INTRODUCTION

**1.1. The smallest singular values of random matrices.** Singular values capture important metric properties of matrices. For an  $N \times n$  matrix  $A$  with real or complex entries,  $n \leq N$ , the *singular values*  $s_j(A)$  are the eigenvalues of  $|A| = (A^*A)^{1/2}$  arranged in a non-decreasing order; thus  $s_1(A) \geq \dots \geq s_n(A) \geq 0$ . The smallest and the largest singular values play a special role.  $s_1(A)$  is the operator norm of  $A$ , while  $s_{\min}(A) := s_n(A)$  is the distance in the operator norm from  $A$  to the set of singular matrices (those with rank smaller than  $n$ ). For square matrices, where  $N = n$ , the smallest singular value  $s_n(A)$  provides a *quantitative measure of invertibility* of  $A$ . It is natural to ask whether typical matrices are well invertible; one often models “typical” matrices as random matrices. This is one of the reasons why the smallest singular values of different classes of random matrices have been extensively studied (see [17] and the references therein).

On a deeper level, questions about the behavior of  $s_{\min}(A)$  for random  $A$  arise in several intrinsic problems of random matrix theory. Quantitative estimates of  $s_{\min}(A)$  for square random matrices  $A$  with independent entries [15], [18], [16], [19] were instrumental in proving the Circular Law, which states that the distribution of the eigenvalues of such matrices converges as  $n \rightarrow \infty$  to the uniform probability measure on the disc [9], [20]. Quantitative estimates on  $s_{\min}(A)$  of random Hermitian matrices  $A$  with independent entries above the diagonal were necessary in the proof of the local semicircle law for the limit spectrum of such matrices [4], [21]. Stronger bounds for the tail distribution of the smallest singular value of a Hermitian random matrix were established in [23], [5]; see also [14].

**1.2. The main results.** In the present paper we study the smallest singular value for a natural class of random matrices, namely for random unitary and orthogonal perturbations of a fixed matrix. Let us consider the complex case first. Let  $D$  be any fixed  $n \times n$  complex matrix, and let  $U$  be a random matrix uniformly distributed over the unitary group  $U(n)$  with respect to the Haar measure. Then the matrix  $D + U$  is non-singular with probability 1, which can be easily observed considering its determinant. However, this observation does not give any useful quantitative information on the degree of non-singularity. A quantitative estimate of the smallest singular value of  $D + U$  is one of the two main results of this paper.

**Theorem 1.1** (Unitary perturbations). *Let  $D$  be an arbitrary fixed  $n \times n$  matrix,  $n \geq 2$ . Let  $U$  be a random matrix uniformly distributed in the unitary group  $U(n)$ . Then*

$$\mathbb{P} \{s_{\min}(D + U) \leq t\} \leq t^c n^C, \quad t > 0.$$

In the statement above and thereafter  $C, c$  denote positive absolute constants. As a consequence of Theorem 1.1, the random matrix  $D + U$  is well invertible,  $\|(D + U)^{-1}\| = n^{O(1)}$  with high probability.

An important point in Theorem 1.2 is that the bound is independent of the deterministic matrix  $D$ . This feature is essential in the application to the Single Ring Theorem, which we shall discuss in Section 1.4 below.

To see that Theorem 1.2 is a subtle result, note that in general it *fails over the reals*. Indeed, suppose  $n$  is odd. If  $-D, U \in SO(n)$ , then  $-D^{-1}U \in SO(n)$  has eigenvalue 1 and as a result  $D + U = D(I_n + D^{-1}U)$  is singular. Therefore, if  $D \in O(n)$  is any fixed matrix and  $U \in O(n)$  is random uniformly distributed,  $s_{\min}(D + U) = 0$  with probability at least  $1/2$ . However, it turns out that this example is essentially the only obstacle for Theorem 1.1 in the real case. Indeed, our second main result states that if  $D$  is not close to  $O(n)$ , then  $D + U$  is well invertible with high probability.

**Theorem 1.2** (Orthogonal perturbations). *Let  $D$  be a fixed  $n \times n$  real matrix,  $n \geq 2$ . Assume that*

$$(1.1) \quad \|D\| \leq K, \quad \inf_{V \in O(n)} \|D - V\| \geq \delta$$

for some  $K \geq 1, \delta \in (0, 1)$ . Let  $U$  be a random matrix uniformly distributed in the orthogonal group  $O(n)$ . Then

$$\mathbb{P} \{s_{\min}(D + U) \leq t\} \leq t^c(Kn/\delta)^C, \quad t > 0.$$

Similarly to the complex case, this bound is uniform over all matrices  $D$  satisfying (1.1). This condition is relatively mild: in the case when  $K = n^{C_1}$  and  $\delta = n^{-C_2}$  for some constants  $C_1, C_2 > 0$ , we have

$$\mathbb{P} \{s_{\min}(D + U) \leq t\} \leq t^c n^C, \quad t > 0,$$

as in the complex case. It is possible that the condition  $\|D\| \leq K$  can be eliminated from Theorem 1.2; we have not tried this in order to keep the argument more readable and because such a condition already appears in the Single Ring Theorem.

Motivated by an application to the Single Ring Theorem, we shall prove the following more general version of Theorem 1.2, which is valid for *complex* diagonal matrices  $D$ .

**Theorem 1.3** (Orthogonal perturbations, full version). *Consider a fixed matrix  $D = \text{diag}(d_1, \dots, d_n)$ ,  $n \geq 2$ , where  $d_i \in \mathbb{C}$ . Assume that*

$$(1.2) \quad \max_i |d_i| \leq K, \quad \max_{i,j} |d_i^2 - d_j^2| \geq \delta$$

for some  $K \geq 1, \delta \in (0, 1)$ . Let  $U$  be a random matrix uniformly distributed in the orthogonal group  $O(n)$ . Then

$$\mathbb{P} \{s_{\min}(D + U) \leq t\} \leq t^c(Kn/\delta)^C, \quad t > 0.$$

Let us show how this result implies Theorem 1.2.

*Proof of Theorem 1.2 from Theorem 1.3.* Without loss of generality, we can assume that  $t \leq \delta/2$ . Further, using rotation invariance of  $U$  we can assume that  $D = \text{diag}(d_1, \dots, d_n)$  where all  $d_i \geq 0$ . The assumptions in (1.1) then imply that

$$(1.3) \quad \max_i |d_i| \leq K, \quad \max_i |d_i - 1| \geq \delta.$$

If  $\max_{i,j} |d_i^2 - d_j^2| \geq \delta^2/4$ , then we can finish the proof by applying Theorem 1.3 with  $\delta^2/4$  instead of  $\delta$ . In the remaining case we have

$$\max_{i,j} |d_i - d_j|^2 \leq \max_{i,j} |d_i^2 - d_j^2| < \delta^2/4,$$

which implies that  $\max_{i,j} |d_i - d_j| < \delta/2$ . Using (1.3), we can choose  $i_0$  so that  $|d_{i_0} - 1| \geq \delta$ . Thus either  $d_{i_0} \geq 1 + \delta$  or  $d_{i_0} \leq 1 - \delta$  holds.

If  $d_{i_0} \geq 1 + \delta$ , then  $d_i > d_{i_0} - \delta/2 \geq 1 + \delta/2$  for all  $i$ . In this case

$$s_{\min}(D + U) \geq s_{\min}(D) - \|U\| > 1 + \delta/2 - 1 \geq t,$$

and the conclusion holds trivially with probability 1.

If  $d_{i_0} \leq 1 - \delta$ , then similarly  $d_i < d_{i_0} + \delta/2 \leq 1 - \delta/2$  for all  $i$ . In this case

$$s_{\min}(D + U) \geq s_{\min}(U) - \|D\| > 1 - (1 - \delta/2) = \delta/2 \geq t,$$

and the conclusion follows trivially again.  $\square$

**1.3. A word about the proofs.** The proofs of Theorems 1.1 and 1.3 are significantly different from those of corresponding results for random matrices with i.i.d. entries [15], [16] and for symmetric random matrices [23]. The common starting point is the identity  $s_{\min}(A) = \min_{x \in S^{n-1}} \|Ax\|_2$ . The critical step of the previous arguments [15], [16], [23] was the analysis of the small ball probability  $\mathbb{P}\{\|Ax\|_2 < t\}$  for a fixed vector  $x \in S^{n-1}$ . The decay of this probability as  $t \rightarrow 0$  is determined by the *arithmetic structure* of the coordinates of the vector  $x$ . An elaborate covering argument was used to treat the set of the vectors with a “bad” arithmetic structure. In contrast to this, arithmetic structure plays no role in Theorems 1.1 and 1.3. The difficulty lies elsewhere—the entries of the matrix  $D + U$  are not independent. This motivates one to seek a way to introduce some independence into the model. The independent variables have to be chosen in such a way that one can tractably express the smallest singular value in terms of them. We give an overview of this procedure in Section 2 below.

The proof of Theorem 1.3 is harder than that of Theorem 1.1. To make the arguments more transparent, the proofs of the two theorems are organized in such a way that they are essentially self-contained and independent of each other. The reader is encouraged to start from the proof of Theorem 1.1.

**1.4. An application to the Single Ring Theorem.** The invertibility problem studied in this paper was motivated by a limit law of the random matrix theory, namely the Single Ring Theorem. This is a result about the eigenvalues of random matrices with prescribed singular values. The problem was studied by Feinberg and Zee [6] on the physical level of rigor, and mathematically by Guionnet, Krishnapur, and Zeitouni [10]. Let  $D_n = \text{diag}(d_1^{(n)}, \dots, d_n^{(n)})$  be an  $n \times n$  diagonal matrix with non-negative diagonal. If we choose  $U_n, V_n$  to be independent, random, and uniformly distributed in  $U(n)$  or  $O(n)$ , then  $A_n = U_n D_n V_n$  constitutes the most natural model of a *random matrix with prescribed singular values*. The matrices  $D_n$  can be deterministic or random; in the latter case we assume that  $U_n$  and  $V_n$  are independent of  $D_n$ .

The Single Ring Theorem [10] describes the typical behavior of the eigenvalues of  $A_n$  as  $n \rightarrow \infty$ . To state this result, we consider the empirical measures of the singular values and the eigenvalues of  $A_n$ :

$$\mu_s^{(n)} := \frac{1}{n} \sum_{j=1}^n \delta_{d_j^{(n)}}, \quad \mu_e^{(n)} := \frac{1}{n} \sum_{j=1}^n \delta_{\lambda_j^{(n)}}$$

where  $\delta_x$  stands for the  $\delta$ -measure at  $x$  and  $\lambda_1^{(n)}, \dots, \lambda_n^{(n)}$  denote the eigenvalues of  $A_n$ . Assume that the measures  $\mu_s^{(n)}$  converge weakly in probability to a measure  $\mu_s$  compactly supported in  $[0, \infty)$ . The Single Ring Theorem [10] states that, under certain conditions, the empirical measures of the eigenvalues  $\mu_e^{(n)}$  converge in

probability to an absolutely continuous rotationally symmetric probability measure  $\mu_e$  on  $\mathbb{C}$ . Haagerup and Larsen [12] previously computed the density of  $\mu_e$  in terms of  $\mu_s$  in the context of operator algebras.

In the formulation of this result,  $\sigma_n(z) := s_n(A_n - zI_n)$  denotes the smallest singular value of the shifted matrix, and  $S_\mu$  denotes the Stieltjes transform of a Borel measure  $\mu$  on  $\mathbb{R}$ :

$$S_\mu(z) = \int_{\mathbb{R}} \frac{d\mu(x)}{z - x}.$$

**Single Ring Theorem** ([10]). *Assume that the sequence  $\{\mu_s^{(n)}\}_{n=1}^\infty$  converges weakly to a probability measure  $\mu_s$  compactly supported on  $\mathbb{R}_+$ . Assume further:*

- (SR1) *There exists  $M > 0$  such that  $\mathbb{P} \{ \|D_n\| > M \} \rightarrow 0$  as  $n \rightarrow \infty$ .*
- (SR2) *There exist constants  $\kappa, \kappa_1 > 0$  such that for any  $z \in \mathbb{C}$ ,  $\text{Im}(z) > n^{-\kappa}$ ,*

$$| \text{Im}(S_{\mu_s^{(n)}}(z)) | \leq \kappa_1.$$

- (SR3) *There exist a sequence of events  $\Omega_n$  with  $\mathbb{P}(\Omega_n) \rightarrow 1$  and constants  $\delta, \delta' > 0$  such that for Lebesgue almost any  $z \in \mathbb{C}$ ,*

$$\mathbb{E} [ \mathbf{1}_{\Omega_n} \mathbf{1}_{\sigma_n(z) < n^{-\delta}} \log^2 \sigma_n(z) ] \leq \delta'.$$

*Then the sequence  $\{\mu_e^{(n)}\}_{n=1}^\infty$  converges in probability to a probability measure  $\mu_e$ . The measure  $\mu_e$  has density, which can be explicitly calculated<sup>1</sup> in terms of the measure  $\mu_s$  and whose support coincides with a single ring  $\{z \in \mathbb{C} : a \leq |z| \leq b\}$  for some  $0 \leq a < b < \infty$ .*

The explicit formula for the density of the measure  $\mu_e$  shows that it is strictly positive in the interior of the ring. This is surprising since the support of the measure  $\mu_s$  can have gaps. Informally, this means that there are no forbidden zones for the eigenvalues, even in the case when there are such zones for singular values.

The inner and outer radii of the ring can be easily calculated [11]:

$$(1.4) \quad a = \left( \int_0^\infty x^{-2} d\mu_s(x) \right)^{-1/2}, \quad b = \left( \int_0^\infty x^2 d\mu_s(x) \right)^{1/2}.$$

The first two conditions of the Single Ring Theorem are effectively checkable for a given sequence  $d_1^{(n)}, \dots, d_n^{(n)}$ . Indeed, condition (SR1) is readily reformulated in terms of this sequence, since  $\|D_n\| = s_1(D_n) = \max(d_1^{(n)}, \dots, d_n^{(n)})$ . Condition (SR2) is already formulated in terms of this sequence; it means that the singular values of the matrices  $D_n$  cannot concentrate on short intervals.

Since the relation between the singular values of the original and shifted matrices is not clear, condition (SR3) is much harder to check. It has only been verified in [10] for the original setup of Feinberg and Zee [6], namely for the case when the singular values of  $A_n$  are random variables with density

$$f(d_1^{(n)}, \dots, d_n^{(n)}) \sim \prod_{j < k} |(d_j^{(n)})^2 - (d_k^{(n)})^2|^\beta \cdot \exp \left( - \sum_{j=1}^n P((d_j^{(n)})^2) \right) \cdot \left( \prod_{j=1}^n d_j^{(n)} \right)^{\beta-1},$$

where  $P$  is a polynomial with positive leading coefficient and where  $\beta = 1$  in the real case and  $\beta = 2$  in the complex case. The proof of condition (SR3) for this model

<sup>1</sup>See [12, Theorem 4.4] and [10, Theorem 1] for a precise description of  $\mu_e$ .

is based on adding small Gaussian noise to the matrix  $A_n$  and using coupling to compare the eigenvalue distributions of random matrices with and without noise. Such an approach does not seem to be extendable to more general distributions of singular values.

As an application of Theorems 1.1 and 1.3, one can show the following:

**Corollary 1.4.** *Condition (SR3) can always be eliminated from the Single Ring Theorem.*

The remaining conditions (SR1) and (SR2) are formulated in terms of the singular values of the original matrix  $D_n$ . This means that the validity of the Single Ring Theorem for a concrete sequence of matrices  $D_n$  can now be effectively checked.

**1.5. Organization of the paper.** The rest of this paper is organized as follows. In Section 2 we give an overview and heuristics of the proofs of both main results. Theorem 1.1 is proved in Section 3. Theorem 1.3 is proved in Section 4 with the exception of the low dimensions  $n = 2, 3$  that are treated separately in Appendix A; some standard tools in the proof of Theorem 1.3 are isolated in Appendix B. In Section 5 we prove Corollary 1.4 concerning the Single Ring Theorem.

**1.6. Notation.** We will use the following notation.

Positive absolute constants are denoted  $C, C_1, c, c_1, \dots$ ; their values may be different in different instances. The notation  $a \lesssim b$  means  $a \leq Cb$  where  $C$  is an absolute constant, and similarly for  $a \gtrsim b$ .

The intervals of integers are denoted by  $[n] := \{1, 2, \dots, n\}$  and  $[k : n] := \{k, k + 1, \dots, n\}$  for  $k \leq n$ .

Given a matrix  $A$ , the minor obtained by removing the first, second, and fifth rows and columns of  $A$  is denoted  $A_{(1,2,5)}$ , and similarly for other subsets of rows and columns.

The identity matrix on  $\mathbb{R}^n$  and  $\mathbb{C}^n$  is denoted  $I_n$ ; we often simply write  $I$  if the ambient dimension is clear.

Since we will be working with several sources of randomness at the same time, we denote by  $\mathbb{P}_{X,Y}(\mathcal{E})$  the conditional probability of the event  $\mathcal{E}$  given all random variables except  $X, Y$ .

The operator norm of a matrix  $A$  is denoted  $\|A\|$ , and the Hilbert-Schmidt (Frobenius) norm is denoted  $\|A\|_{\text{HS}}$ .

The diagonal matrix with diagonal entries  $d_1, \dots, d_n$  is denoted  $\text{diag}(d_1, \dots, d_n)$ .

Finally, without loss of generality we may assume in Theorems 1.1 and 1.3 that  $t < c\delta$  for an arbitrarily small absolute const  $c > 0$ .

## 2. STRATEGY OF THE PROOFS

Let us present the heuristics of the proofs of Theorems 1.1 and 1.3. Both proofs are based on the idea of using local and global structures of the Lie groups  $U(n)$  and  $O(n)$ , but the argument for  $O(n)$  is more difficult.

**2.1. Unitary perturbations.** Our proof of Theorems 1.1 uses both global and local structures of the Lie group  $U(n)$ . The local structure is determined by the infinitesimally small perturbations of the identity in  $U(n)$ , which are given by skew-Hermitian matrices. This allows us to essentially replace  $D + U$  (up to  $O(\varepsilon^2)$  error) by

$$VD + I + \varepsilon S$$

where  $V$  is a random matrix uniformly distributed in  $U(n)$ ,  $S$  is an independent skew-Hermitian matrix, and  $\varepsilon > 0$  is a small number. The distribution of  $S$  can be arbitrary. For example, one may choose  $S$  to have independent normal above-diagonal entries. (In the actual proof, we populate just one row and column of  $S$  by random variables leaving the other entries zero; see (3.4).) After conditioning on  $V$ , we are left with a random matrix with a lot of independent entries—the quality that was missing from the original problem.

However, this local argument is not powerful enough, in particular because real skew-Hermitian (i.e. skew-symmetric) matrices themselves are singular in odd dimensions  $n$ . This forces us to use some global structure of  $U(n)$  as well. A simplest random global rotation is a random complex rotation  $R$  in one coordinate in  $\mathbb{C}^n$  (given by multiplication of that coordinate by a random unit complex number). Thus we can essentially replace  $D + U$  by

$$A = RVD + I + \varepsilon S,$$

and we again condition on  $V$ . A combination of the two sources of randomness, a local perturbation  $S$  and a global perturbation  $R$ , produces enough power to conclude that  $A$  is typically well invertible, which leads to Theorem 1.1.

The formal proof of Theorem 1.1 is presented in Section 3.

**2.2. Orthogonal perturbations.** Our proof of Theorem 1.3 will also make use of both global and local structures of the Lie group  $O(n)$ . The local structure is determined by the skew-symmetric matrices. As before, we can use it to replace  $D + U$  by  $VD + I + \varepsilon S$  where  $V$  is a random matrix uniformly distributed in  $O(n)$  and  $S$  is a random independent Gaussian skew-symmetric matrix (with i.i.d.  $N_{\mathbb{R}}(0, 1)$  above-diagonal entries).

Regarding the global structure, the simplest random global rotation in  $O(n)$  is a random rotation  $R$  of some *two* coordinates in  $\mathbb{R}^n$ , say the first two. Still,  $R$  alone does not appear to be powerful enough, so we supplement it with a further *random change of basis*. Specifically, we replace  $D$  with  $\tilde{D} = QDQ^T$  where  $Q$  is a random independent rotation of the first two coordinates. Overall, we have changed  $D + U$  to

$$\tilde{A} = RV\tilde{D} + I + \varepsilon S, \quad \text{where } \tilde{D} = QDQ^T.$$

Only now do we condition on  $V$ , and we will work with three sources of randomness—a local perturbation given by  $S$  and two global perturbations given by  $R$  and  $Q$ .

**2.2.1. Decomposition of the problem.** By rotation invariance, we can assume that  $D$  is diagonal; thus  $D = \text{diag}(d_1, \dots, d_n)$ . By assumption,  $d_i^2$  and  $d_j^2$  are not close to each other for some pair of indices  $i, j$ ; without loss of generality we can assume that  $d_1^2$  and  $d_2^2$  are not close to each other. Recall that our task is to show that

$$(2.1) \quad s_{\min}(\tilde{A}) = \inf_{x \in S^{n-1}} \|\tilde{A}x\|_2 \gtrsim \varepsilon$$

with high probability. (In this informal presentation, we suppress the dependence on  $n$ ; it should always be polynomial.) Each  $x \in S^{n-1}$  has a coordinate whose magnitude is at least  $n^{-1/2}$ . By decomposing the sphere according to which coordinate is large, without loss of generality we can replace our task (2.1) by showing

that

$$\inf_{x \in S_{1,2}} \|\tilde{A}x\|_2 \gtrsim \varepsilon$$

where  $S_{1,2}$  consists of the vectors  $x \in S^{n-1}$  with  $|x_1|^2 + |x_2|^2 \geq 1/n$ .

In order to use the rotations  $R, Q$  which act on the first two coordinates, we decompose  $\tilde{A}$  as follows:

$$(2.2) \quad \tilde{A} = \begin{bmatrix} A_0 & Y \\ X & A_{(1,2)} \end{bmatrix}, \quad \text{where } A_0 \in \mathbb{C}^{2 \times 2}, A_{(1,2)} \in \mathbb{C}^{(n-2) \times (n-2)}.$$

We condition on everything except  $Q, R$  and the first two rows and columns of  $S$ . This fixes the minor  $A_{(1,2)}$ . We will proceed differently depending on whether  $A_{(1,2)}$  is well invertible or not.

2.2.2. *When the minor is well invertible.* Let us assume that

$$\|M\| \lesssim \frac{1}{\varepsilon}, \quad \text{where } M := (A_{(1,2)})^{-1}.$$

It is not difficult to show (see Lemma 4.4) that in this case

$$\inf_{x \in S_{1,2}} \|\tilde{A}x\|_2 \gtrsim \varepsilon \cdot s_{\min}(A_0 - YMX).$$

So our task becomes to prove that

$$s_{\min}(A_0 - YMX) \gtrsim 1.$$

We have reduced our problem to the invertibility of a  $2 \times 2$  random matrix.

The argument in this case will only rely on the global perturbations  $Q$  and  $R$  and will not use the local perturbation  $S$ . So let us assume for simplicity that  $S = 0$ , although removing  $S$  will take some effort in the formal argument. Expressing the matrix  $A_0 - YMX$  as a function of  $R$ , we see that

$$A_0 - YMX = I + R_0B$$

where  $B \in \mathbb{C}^{2 \times 2}$  and  $R_0 \in O(2)$  is the part of  $R$  restricted to the first two coordinates (recall that  $R$  is the identity on the other coordinates).

Note that  $I + R_0B$  has the same distribution as  $R_0^{-1} + B$  and  $R_0$  is uniformly distributed in  $O(2)$ . But the invertibility of the latter matrix is the same problem that we are studying in this paper, only in dimension two. One can prove Theorem 1.3 in dimension two (and even for non-diagonal matrices) by a separate argument based on Remez-type inequalities; see Appendix A. It yields that unless  $B$  is approximately *complex orthogonal*, i.e.  $\|BB^T - I\| \ll \|B\|^2$ , the random matrix  $I + R_0B$  is well invertible with high probability in  $R_0$ , leading to the desired conclusion. We have thus reduced the problem to showing that  $B$  is not approximately complex orthogonal.

To this end we use the remaining source of randomness, the random rotation  $Q$ . Expressing  $b$  as a function of  $Q$ , we see that

$$B = T\tilde{D}_0$$

where  $T \in \mathbb{C}^{2 \times 2}$  is a fixed matrix,  $\tilde{D}_0 = Q_0D_0Q_0^T$ , and  $Q_0, D_0$  are the  $2 \times 2$  minors of  $Q$  and  $D$ , respectively. Thus  $Q_0$  is a random rotation in  $SO(2)$  and  $D_0 = \text{diag}(d_1, d_2)$ .

Now we recall our assumption that  $d_1^2$  and  $d_2^2$  are not close to each other. It is fairly easy to show for such  $D_0$  that, whatever the matrix  $T$  is, the random matrix  $B = T\tilde{D}_0 = TQ_0D_0Q_0^T$  is not approximately complex orthogonal with high



probability in  $Q_0$  (see Lemma 4.6). This concludes the argument in this case. The formal analysis is presented in Section 4.3.

2.2.3. *When the minor is poorly invertible.* The remaining case is when

$$\|M\| \gg \frac{1}{\varepsilon}, \quad \text{where } M := (A_{(1,2)})^{-1}.$$

We will only use the local perturbation  $S$  in this case.

Here we encounter a new problem. Imagine for a moment that we were working with decompositions into dimensions  $1 + (n - 1)$  rather than  $2 + (n - 2)$ ; thus in (2.2) we had  $A_0 \in \mathbb{C}^{1 \times 1}$ ,  $A_{(1,2)} \in \mathbb{C}^{(n-1) \times (n-1)}$ . Using the Gaussian random vector  $X$ , one could quickly show (see Lemma 4.8) that in this case

$$(2.3) \quad \inf_{x \in S_1} \|\tilde{A}x\|_2 \gtrsim \varepsilon$$

with high probability, where  $S_1$  consists of the vectors  $x \in S^{n-1}$  with  $|x_1| \geq n^{-1/2}$ .

Unfortunately, this kind of argument fails for decompositions into dimensions  $2 + (n - 2)$  which we are working with. In other words, we can step *one* rather than two dimensions up—from a poor invertibility of an  $(n - 1) \times (n - 1)$  minor to the good invertibility of the  $n \times n$  matrix (on vectors with the large corresponding coordinate). The failure of stepping two dimensions up has a good reason. Indeed, one can show that Gaussian skew symmetric matrices are well invertible in even dimensions  $n$  and singular in odd dimensions  $n$ . Since our argument in the current case only relies on the local perturbation given by a Gaussian skew symmetric matrix, nothing seems to prevent both the  $(n - 2) \times (n - 2)$  minor and the full  $n \times n$  matrix to be poorly invertible if  $n$  is odd.

To circumvent this difficulty, we shall redefine the two cases that we have worked with, as follows.

**Case 1:** There exists an  $(n - 3) \times (n - 3)$  minor  $A_{(1,2,i)}$  of  $A_{(1,2)}$  which is well invertible. In this case one proceeds by the same argument as in Section 2.2.2, but for the decomposition into dimensions  $3 + (n - 3)$  rather than  $2 + (n - 2)$ . The formal argument is presented in Section 4.3.

**Case 2:** All  $(n - 3) \times (n - 3)$  minors  $A_{(1,2,i)}$  of  $A_{(1,2)}$  are poorly invertible. Let us fix  $i$  and apply the reasoning described above, which allows us to move *one* dimension up, this time from  $n - 3$  to  $n - 2$ . We conclude that  $A_{(1,2)}$  is well invertible on the vectors whose  $i$ -th coordinate is large. Doing this for each  $i$  and recalling that each vector has at least one large coordinate, we conclude that  $A_{(1,2)}$  is well invertible on all vectors. Now we are in the same situation that we have already analyzed in Section 2.2.2, as the minor  $A_{(1,2)}$  is well invertible. So we proceed by the same argument as there. The formal analysis of this case is presented in Section 4.4.

Summarizing, in Case 1 we move three dimensions up, from  $n - 3$  to  $n$ , in one step. In Case 2 we take two steps, first moving one dimension up (from poor invertibility in dimension  $n - 3$  to good invertibility in dimension  $n - 2$ ), then two dimensions up (from good invertibility in dimension  $n - 2$  to good invertibility in dimension  $n$ ).

This concludes the informal presentation of the proof of Theorem 1.3.

3. UNITARY PERTURBATIONS: PROOF OF THEOREM 1.1

In this section we give a formal proof of Theorem 1.1.

3.1. Decomposition of the problem; local and global perturbations.

3.1.1. *Decomposition of the sphere.* By definition, we have

$$s_{\min}(D + U) = \inf_{x \in S^{n-1}} \|(D + U)x\|_2.$$

Since for every  $x \in S^{n-1}$  there exists a coordinate  $i \in [n]$  such that  $|x_i| \geq 1/\sqrt{n}$ , a union bound yields

$$(3.1) \quad \mathbb{P} \{ s_{\min}(D + U) \leq t \} \leq \sum_{i=1}^n \mathbb{P} \left\{ \inf_{x \in S_i} \|(D + U)x\|_2 \leq t \right\}$$

where

$$S_i = \{ x \in S^{n-1} : |x_i| \geq 1/\sqrt{n} \}.$$

So, without loss of generality, our goal is to bound

$$(3.2) \quad \mathbb{P} \left\{ \inf_{x \in S_1} \|(D + U)x\|_2 \leq t \right\}.$$

3.1.2. *Introducing local and global perturbations.* We can express  $U$  in distribution as

$$(3.3) \quad U = V^{-1}R^{-1}W$$

where  $V, R, W \in U(n)$  are random independent matrices, such that  $V$  is uniformly distributed in  $U(n)$  while  $R$  and  $W$  may have arbitrary distributions. In a moment, we shall choose  $R$  as a random diagonal matrix (a “global perturbation”),  $W$  as a small perturbation of identity with a random skew-Hermitian matrix (a “local perturbation”), and we shall then condition on  $V$ .

So we let  $V$  be uniform in  $U(n)$  and let

$$R = \text{diag}(r, 1, \dots, 1),$$

where  $r$  is a random variable uniformly distributed on the unit torus  $\mathbb{T} \subset \mathbb{C}$ . Finally,  $W$  will be defined with the help of the following standard lemma. It expresses quantitatively the local structure of the unitary group  $U(n)$ , namely that the tangent space to  $U(n)$  at the identity matrix is given by the skew-Hermitian matrices.

**Lemma 3.1** (Perturbations of identity in  $U(n)$ ). *Let  $S$  be an  $n \times n$  skew-Hermitian matrix (i.e.  $S^* = -S$ ), let  $\varepsilon > 0$ , and define*

$$W_0 = I + \varepsilon S.$$

*Then there exists  $W \in U(n)$  which depends only on  $W_0$  and such that*

$$\|W - W_0\| \leq 2\varepsilon^2 \|S^2\| \quad \text{whenever} \quad \varepsilon^2 \|S^2\| \leq 1/4.$$

*Proof.* We write the singular value decomposition  $W_0 = U_0 \Sigma V_0$  where  $U_0, V_0 \in U(n)$  and  $\Sigma$  is a diagonal matrix with non-negative entries, and we define  $W := U_0 V_0$ . Since  $S$  is skew-Hermitian, we see that

$$W_0^* W_0 = (I + \varepsilon S)^*(I + \varepsilon S) = I - \varepsilon^2 S^2,$$

so  $W_0^*W_0 - I = \varepsilon^2 S^2$ . On the other hand, the singular value decomposition of  $W_0$  yields  $W_0^*W_0 - I = V_0^*(\Sigma^2 - I)V_0$ . Combining these, we obtain

$$\|\Sigma^2 - I\| \leq \varepsilon^2 \|S^2\|.$$

Assuming  $\varepsilon^2 \|S^2\| \leq 1/4$  and recalling that  $\Sigma$  is a diagonal matrix with non-negative entries, we conclude that

$$\|\Sigma - I\| \leq 2\varepsilon^2 \|S^2\|.$$

It follows that

$$\|W - W_0\| = \|U_0(I - \Sigma)V_0\| = \|I - \Sigma\| \leq 2\varepsilon^2 \|S^2\|,$$

as claimed. □

Now we define the random skew-Hermitian matrix as

$$(3.4) \quad S = \begin{bmatrix} \sqrt{-1} s & -Z^T \\ Z & 0 \end{bmatrix}$$

where  $s \sim N_{\mathbb{R}}(0, 1)$  and  $Z \sim N_{\mathbb{R}}(0, I_{n-1})$  are the independent standard normal random variable and vector, respectively. Clearly,  $S$  is skew-Hermitian.

Let  $\varepsilon \in (0, 1)$  be an arbitrarily small number. We define  $W_0$  and  $W$  as in Lemma 3.1, and finally we recall that a random uniform  $U$  is represented as in (3.3).

**3.1.3. Replacing  $D+U$  by  $RVD+I+\varepsilon S$ .** Let us rewrite the quantity to be estimated (3.2) in terms of the global and local perturbations. Applying Lemma 3.1 for the random matrix  $W_0 = I + \varepsilon S$ , we obtain a random matrix  $W \in U(n)$ , which satisfies the following for every  $x \in S_1$ :

$$\begin{aligned} \|(D + U)x\|_2 &= \|(D + V^{-1}R^{-1}W)x\|_2 = \|(RVD + W)x\|_2 \\ &\geq \|(RVD + W_0)x\|_2 - \|W - W_0\| \\ &\geq \|(RVD + I + \varepsilon S)x\|_2 - 2\varepsilon^2 \|S^2\| \quad \text{whenever } \varepsilon^2 \|S^2\| \leq 1/4. \end{aligned}$$

Further,  $\mathbb{E} \|S\|^2 \leq s^2 + 2\|Z\|_2^2 = 2n - 1$ , so  $\|S\| = O(\sqrt{n})$  with high probability. More precisely, let  $K_0 > 1$  be a parameter to be chosen later and which satisfies

$$(3.5) \quad \varepsilon^2 K_0^2 n \leq 1/4.$$

Consider the event

$$(3.6) \quad \mathcal{E}_S = \{\|S\| \leq K_0 \sqrt{n}\}; \quad \text{then } \mathbb{P}(\mathcal{E}_S^c) \leq 2 \exp(-cK_0^2 n)$$

by a standard large deviation inequality (see e.g. [22, Corollary 5.17]). On  $\mathcal{E}_S$ , one has  $\varepsilon^2 \|S^2\| \leq 1/4$  due to (3.5), and thus

$$\|(D + U)x\|_2 \geq \|(RVD + I + \varepsilon S)x\|_2 - 2\varepsilon^2 K_0^2 n.$$

Denote

$$A := RVD + I + \varepsilon S.$$

Let  $\mu \in (0, 1)$  be a parameter to be chosen later and which satisfies

$$(3.7) \quad \mu \geq 2\varepsilon K_0^2 n.$$

By the above, our goal is to estimate

$$\begin{aligned}
 \mathbb{P} \left\{ \inf_{x \in S_1} \|(D + U)x\|_2 \leq \mu\varepsilon \right\} &\leq \mathbb{P} \left\{ \inf_{x \in S_1} \|Ax\|_2 \leq \mu\varepsilon + 2\varepsilon K_0^2 n \wedge \mathcal{E}_S \right\} + \mathbb{P}(\mathcal{E}_S^c) \\
 (3.8) \qquad \qquad \qquad &\leq \mathbb{P} \left\{ \inf_{x \in S_1} \|Ax\|_2 \leq 2\mu\varepsilon \wedge \mathcal{E}_S \right\} + 2 \exp(-cK_0^2 n).
 \end{aligned}$$

Summarizing, we now have control of the first coordinate of  $x$ , we have introduced the global perturbation  $R$  and the local perturbation  $S$ , and we replaced the random matrix  $D + U$  by  $A = RVD + I + \varepsilon S$ .

**3.1.4. Decomposition into  $1 + (n - 1)$  dimensions.** Next, we would like to expose the first row and first column of the matrix  $A = RVD + I + \varepsilon S$ . We do so first for the matrix

$$VD = \begin{bmatrix} (VD)_{11} & v^\top \\ u & (VD)_{(1,1)} \end{bmatrix} \quad \text{where } u, v \in \mathbb{C}^{n-1}.$$

Recalling the definition (3.4) of  $S$ , we can express

$$(3.9) \qquad A = RVD + I + \varepsilon S = \begin{bmatrix} r(VD)_{11} + 1 + \sqrt{-1} \varepsilon s & (rv - \varepsilon Z)^\top \\ u + \varepsilon Z & (I + VD)_{(1,1)} \end{bmatrix} =: \begin{bmatrix} A_{11} & Y^\top \\ X & B^\top \end{bmatrix}.$$

We condition on an arbitrary realization of the random matrix  $V$ . This fixes the number  $(VD)_{11}$ , the vectors  $u, v$ , and the matrix  $B^\top$  involved in (3.9). All randomness thus remains in the independent random variables  $r$  (which is chosen uniformly in  $\mathbb{T}$ ),  $s \sim N_{\mathbb{R}}(0, 1)$ , and the independent random vector  $Z \sim N_{\mathbb{R}}(0, I_{n-1})$ . We regard the random variable  $r$  as a global perturbation, and  $s, Z$  as local perturbations.

**3.2. Invertibility via quadratic forms.** Recall from (3.8) that our goal is to bound below the quantity

$$\inf_{x \in S_1} \|Ax\|_2.$$

Let  $A_1, \dots, A_n$  denote the columns of  $A$ . Let  $h \in \mathbb{C}^n$  be such that

$$\|h\|_2 = 1, \quad h^\top A_i = 0, \quad i = 2, \dots, n.$$

For every  $x \in \mathbb{C}^n$  we have

$$\|Ax\|_2 = \left\| \sum_{i=1}^n x_i A_i \right\|_2 \geq \left| h^\top \sum_{i=1}^n x_i A_i \right| = |x_1| \cdot |h^\top A_1|.$$

Since  $|x_1| \geq 1/\sqrt{n}$  for all vectors  $x \in S_1$ , this yields

$$(3.10) \qquad \inf_{x \in S_1} \|Ax\|_2 \geq \frac{1}{\sqrt{n}} |h^\top A_1|.$$

We thus reduced the problem to finding a lower bound on  $|h^\top A_1|$ . Let us express this quantity as a function of  $X$  and  $Y$  in the decomposition as in (3.9). The following lemma shows that  $|h^\top A_1|$  is essentially a quadratic form in  $X, Y$ , which is ultimately a quadratic form in  $Z$ .

**Lemma 3.2** (Quadratic form). *Consider an arbitrary square matrix*

$$A = \begin{bmatrix} A_{11} & Y^\top \\ X & B^\top \end{bmatrix}, \quad A_{11} \in \mathbb{C}, \quad X, Y \in \mathbb{C}^{n-1}, \quad B \in \mathbb{C}^{(n-1) \times (n-1)}.$$

Assume that  $B$  is invertible. Let  $A_1, \dots, A_n$  denote the columns of  $A$ . Let  $h \in \mathbb{C}^n$  be such that

$$\|h\|_2 = 1, \quad h^\top A_i = 0, \quad i = 2, \dots, n.$$

Then

$$|h^\top A_1| = \frac{|A_{11} - X^\top B^{-1}Y|}{\sqrt{1 + \|B^{-1}Y\|_2^2}}.$$

*Proof.* The argument is from [23, Proposition 5.1]. We express  $h$  by exposing its first coordinate as

$$h = \begin{bmatrix} h_1 \\ \bar{h} \end{bmatrix}.$$

Then

$$(3.11) \quad h^\top A_1 = (h_1 \ \bar{h}^\top) \begin{bmatrix} A_{11} \\ X \end{bmatrix} = A_{11}h_1 + \bar{h}^\top X = A_{11}h_1 + X^\top \bar{h}.$$

The assumption that  $h^\top A_i = 0$  for  $i \geq 2$  can be stated as

$$0 = (h_1 \ \bar{h}^\top) \begin{bmatrix} Y^\top \\ B^\top \end{bmatrix} = h_1 Y^\top + \bar{h}^\top B^\top.$$

Equivalently,  $h_1 Y + B\bar{h} = 0$ . Hence

$$(3.12) \quad \bar{h} = -h_1 \cdot B^{-1}Y.$$

To determine  $h_1$ , we use the assumption  $\|h\|_2 = 1$ , which implies

$$1 = |h_1|^2 + \|\bar{h}\|_2^2 = |h_1|^2 + |h_1|^2 \cdot \|B^{-1}Y\|_2^2.$$

So

$$(3.13) \quad |h_1| = \frac{1}{\sqrt{1 + \|B^{-1}Y\|_2^2}}.$$

Combining (3.11) and (3.12), we obtain

$$|h^\top A_1| = A_{11}h_1 - h_1 \cdot X^\top B^{-1}Y.$$

This and (3.13) complete the proof. □

Let us use this lemma for our random matrix  $A$  in (3.9). One can check that the minor  $B$  is invertible almost surely. To facilitate the notation, denote

$$M := B^{-1}.$$

Recall that  $M$  is a fixed matrix. Then

$$|h^\top A_1| = \frac{|A_{11} - X^\top MY|}{\sqrt{1 + \|MY\|_2^2}}.$$

Since as we know from (3.9),

$$A_{11} = r(VD)_{11} + 1 + \sqrt{-1}\varepsilon s, \quad X = u + \varepsilon Z, \quad Y = rv - \varepsilon Z,$$

we can expand

$$(3.14) \quad |h^\top A_1| = \frac{|r(VD)_{11} + 1 + \sqrt{-1}\varepsilon s - ru^\top Mv - \varepsilon r(Mv)^\top Z + \varepsilon u^\top MZ + \varepsilon^2 Z^\top MZ|}{\sqrt{1 + \|rMv - \varepsilon MZ\|_2^2}}.$$

Recall that  $(VD)_{11}$ ,  $u$ ,  $v$ ,  $M$  are fixed, while  $r$ ,  $s$ ,  $Z$  are random.

Our difficulty in controlling this ratio is that the typical magnitudes of  $\|M\|$  and of  $\|Mv\|_2$  are unknown to us. So we shall consider all possible cases depending on these magnitudes.

**3.3. When the denominator is small.** We start with the case where the denominator in (3.14) is  $O(1)$ . The argument in this case will rely on the local perturbation given by  $s$ .

Let  $K \geq 1$  be a parameter to be chosen later, and let us consider the event

$$\mathcal{E}_{\text{denom}} = \{\|rMv - \varepsilon MZ\|_2 \leq K\}.$$

This event depends on random variables  $r$  and  $Z$  and is independent of  $s$ . Let us condition on realizations of  $r$  and  $Z$  which satisfy  $\mathcal{E}_{\text{denom}}$ . We can rewrite (3.14) as

$$|h^\top A_1| \geq \frac{|ra + \sqrt{-1}\varepsilon s|}{\sqrt{1 + K^2}} \geq \frac{|ra + \sqrt{-1}\varepsilon s|}{2K}$$

where  $a \in \mathbb{C}$  and  $r \in \mathbb{T}$  (and of course  $K$ ) are fixed numbers and  $s \sim N_{\mathbb{R}}(0, 1)$ . Since the density of  $s$  is bounded by  $1/\sqrt{2\pi}$ , it follows that

$$\mathbb{P}_s \left\{ |h^\top A_1| \leq \frac{\lambda\varepsilon}{K} \right\} \leq C\lambda, \quad \lambda \geq 0.$$

Therefore, a similar bound holds for the unconditional probability:

$$\mathbb{P} \left\{ |h^\top A_1| \leq \frac{\lambda\varepsilon}{K} \text{ and } \mathcal{E}_{\text{denom}} \right\} \leq C\lambda, \quad \lambda \geq 0.$$

Finally, using (3.10) this yields

$$(3.15) \quad \mathbb{P} \left\{ \inf_{x \in S_1} \|Ax\|_2 \leq \frac{\lambda\varepsilon}{K\sqrt{n}} \text{ and } \mathcal{E}_{\text{denom}} \right\} \leq C\lambda, \quad \lambda \geq 0.$$

This is a desired form of the invertibility estimate, which is useful when the event  $\mathcal{E}_{\text{denom}}$  holds. Next we will analyze the case where it does not.

**3.4. When the denominator is large and  $\|M\|$  is small.** If  $\mathcal{E}_{\text{denom}}$  does not occur, either  $\|Mv\|_2$  or  $\|MZ\|_2$  must be large. Furthermore, since  $M$  is fixed and  $Z \sim N_{\mathbb{R}}(0, I_{n-1})$ , we have  $\|MZ\|_2 \sim \|M\|_{\text{HS}}$  with high probability. We shall consider the cases where  $\|M\|_{\text{HS}}$  is small and large separately. In this section we analyze the case where  $\|M\|_{\text{HS}}$  is small. The argument will rely on the global perturbation  $r$  and the local perturbation  $Z$ .

To formalize this, assume that  $\mathcal{E}_{\text{denom}}$  does not occur. Then we can estimate the denominator in (3.14) as

$$\sqrt{1 + \|rMv - \varepsilon MZ\|_2^2} \leq 2\|rMv - \varepsilon MZ\|_2 \leq 2\|Mv\|_2 + \varepsilon\|MZ\|_2.$$

Note that  $\mathbb{E}\|MZ\|_2^2 = \|M\|_{\text{HS}}^2$ . This prompts us to consider the event

$$\mathcal{E}_{MZ} := \{\|MZ\|_2 \leq K_1\|M\|_{\text{HS}}\},$$

where  $K_1 \geq 1$  is a parameter to be chosen later. This event is likely. Indeed, the map  $f(Z) = \|MZ\|_2$  defined on  $\mathbb{R}^{n-1}$  has Lipschitz norm bounded by  $\|M\|$ , so a concentration inequality in the Gauss space (see e.g. [13, (1.5)]) implies that

$$\mathbb{P}(\mathcal{E}_{MZ}) \geq 1 - \exp\left(-\frac{cK_1^2\|M\|_{\text{HS}}^2}{\|M\|^2}\right) \geq 1 - \exp(-cK_1^2).$$

On the event  $\mathcal{E}_{MZ} \cap \mathcal{E}_{\text{denom}}^c$  one has

$$(3.16) \quad \sqrt{1 + \|rMv - \varepsilon MZ\|_2^2} \leq 2\|Mv\|_2 + \varepsilon K_1 \|M\|_{\text{HS}}.$$

Now we consider the case where  $\|M\|_{\text{HS}}$  is small. This can be formalized by the event

$$(3.17) \quad \mathcal{E}_M := \left\{ \|M\|_{\text{HS}} \leq \frac{K}{2\varepsilon K_1} \right\}.$$

On the event  $\mathcal{E}_{MZ} \cap \mathcal{E}_{\text{denom}}^c \cap \mathcal{E}_M$ , the inequality (3.16) yields the following bound on the denominator in (3.14):

$$\sqrt{1 + \|rMv - \varepsilon MZ\|_2^2} \leq 2\|Mv\|_2 + \frac{K}{2}.$$

On the other hand, the left side of this inequality is at least  $K$  by  $\mathcal{E}_{\text{denom}}^c$ . Therefore

$$(3.18) \quad \sqrt{1 + \|rMv - \varepsilon MZ\|_2^2} \leq 4\|Mv\|_2.$$

To estimate the numerator in (3.14), let us condition for a moment on all random variables but  $r$ . The numerator then takes the form  $|ar + b|$  where  $a = (VD)_{11} - u^T Mv - \varepsilon(Mv)^T Z$  and  $b = 1 + \sqrt{-1}\varepsilon s + \varepsilon u^T MZ + \varepsilon^2 Z^T MZ$  are fixed numbers and  $r$  is uniformly distributed in  $\mathbb{T}$ . A quick calculation yields a general bound on the conditional probability:

$$\mathbb{P}_r \{ |ar + b| \geq \lambda_1 |a| \} \geq 1 - C\lambda_1, \quad \lambda_1 \geq 0.$$

Therefore a similar bound holds unconditionally. Let  $\lambda_1 \in (0, 1)$  be a parameter to be chosen later. We showed that the event

$$\mathcal{E}_{\text{num}} := \{ \text{numerator in (3.14)} \geq \lambda_1 |(VD)_{11} - u^T Mv - \varepsilon(Mv)^T Z| \}$$

is likely:

$$\mathbb{P}(\mathcal{E}_{\text{num}}) \geq 1 - C\lambda_1.$$

Assume that the event  $\mathcal{E}_{\text{num}} \cap \mathcal{E}_{MZ} \cap \mathcal{E}_{\text{denom}}^c \cap \mathcal{E}_M$  occurs. (Here the first two events are likely, while the other two specify the case being considered in this section.) We substitute the bounds on the denominator (3.18) and the numerator (given by the definition of  $\mathcal{E}_{\text{num}}$ ) into (3.14) to obtain

$$|h^T A_1| \geq \frac{\lambda_1 |(VD)_{11} - u^T Mv - \varepsilon(Mv)^T Z|}{4\|Mv\|_2}.$$

We can rewrite this inequality as

$$|h^T A_1| \geq \left| d + \frac{\lambda_1 \varepsilon}{4} \cdot w^T Z \right|, \quad \text{where } d = \lambda_1 \cdot \frac{(VD)_{11} - u^T Mv}{4\|Mv\|_2}, \quad w := -\frac{Mv}{\|Mv\|_2}.$$

Here  $d$  is a fixed number and  $w$  is a fixed unit vector, while  $Z \sim N_{\mathbb{R}}(0, I_{n-1})$ . Therefore  $w^T Z = \theta\gamma$ , where  $\gamma \sim N_{\mathbb{R}}(0, 1)$ , and  $\theta \in \mathbb{C}$ ,  $|\theta| = 1$ . A quick density calculation yields the following bound on the conditional probability:

$$\mathbb{P}_Z \left\{ \left| d + \frac{\lambda_1 \varepsilon}{4} \cdot w^T Z \right| \leq \lambda \lambda_1 \varepsilon \right\} \leq C\lambda, \quad \lambda > 0.$$

Hence a similar bound holds unconditionally:

$$\mathbb{P} \{ |h^T A_1| \leq \lambda \lambda_1 \varepsilon \text{ and } \mathcal{E}_{\text{num}} \cap \mathcal{E}_{MZ} \cap \mathcal{E}_{\text{denom}}^c \cap \mathcal{E}_M \} \leq C\lambda, \quad \lambda > 0.$$

Therefore,

$$\begin{aligned} \mathbb{P} \{ |h^\top A_1| \leq \lambda \lambda_1 \varepsilon \text{ and } \mathcal{E}_{\text{denom}}^c \cap \mathcal{E}_M \} &\leq \mathbb{P}(\mathcal{E}_{\text{num}}^c) + P(\mathcal{E}_{MZ}^c) + C\lambda \\ &\leq C\lambda_1 + \exp(-cK_1^2) + C\lambda, \quad \lambda > 0. \end{aligned}$$

Using (3.10), we conclude that

$$(3.19) \quad \mathbb{P} \left\{ \inf_{x \in S_1} \|Ax\|_2 \leq \frac{\lambda \lambda_1 \varepsilon}{\sqrt{n}} \text{ and } \mathcal{E}_{\text{denom}}^c \cap \mathcal{E}_M \right\} \leq C\lambda_1 + \exp(-cK_1^2) + C\lambda, \quad \lambda > 0.$$

**3.5. When  $\|M\|$  is large.** The remaining case to analyze is where  $\|M\|$  is large, i.e. where  $\mathcal{E}_M$  does not occur. Here we shall estimate the desired quantity  $\inf_{x \in S_1} \|Ax\|_2$  directly, without using Lemma 3.2. The local perturbation  $Z$  will do the job.

Indeed, on  $\mathcal{E}_M^c$  we have

$$\|B^{-1}\| \geq \frac{1}{\sqrt{n}} \|B^{-1}\|_{\text{HS}} = \frac{1}{\sqrt{n}} \|M\|_{\text{HS}} \geq \frac{K}{2\varepsilon K_1 \sqrt{n}}.$$

Therefore there exists a vector  $\tilde{w} \in \mathbb{C}^{n-1}$  such that

$$(3.20) \quad \|\tilde{w}\|_2 = 1, \quad \|B\tilde{w}\|_2 \leq \frac{2\varepsilon K_1 \sqrt{n}}{K}.$$

Note that  $\tilde{w}$  can be chosen depending only on  $B$  and thus is fixed.

Let  $x \in S_1$  be arbitrary; we can express it as

$$(3.21) \quad x = \begin{bmatrix} x_1 \\ \tilde{x} \end{bmatrix}, \quad \text{where } |x_1| \geq \frac{1}{\sqrt{n}}.$$

Set

$$w = \begin{bmatrix} 0 \\ \tilde{w} \end{bmatrix} \in \mathbb{C}^n.$$

Using the decomposition of  $A$  given in (3.9), we obtain

$$\begin{aligned} \|Ax\|_2 &\geq |w^\top Ax| = \left| \begin{bmatrix} 0 & \tilde{w}^\top \end{bmatrix} \begin{bmatrix} A_{11} & Y^\top \\ X & B^\top \end{bmatrix} \begin{bmatrix} x_1 \\ \tilde{x} \end{bmatrix} \right| \\ &= |x_1 \cdot \tilde{w}^\top X + \tilde{w}^\top B^\top \tilde{x}| \\ &\geq |x_1| \cdot |\tilde{w}^\top X| - \|B\tilde{w}\|_2 \quad (\text{by the triangle inequality}) \\ &\geq \frac{1}{\sqrt{n}} |\tilde{w}^\top X| - \frac{2\varepsilon K_1 \sqrt{n}}{K} \quad (\text{using (3.21) and (3.20)}). \end{aligned}$$

Recalling from (3.9) that  $X = u + \varepsilon Z$  and taking the infimum over  $x \in S_1$ , we obtain

$$\inf_{x \in S_1} \|Ax\|_2 \geq \frac{1}{\sqrt{n}} |\tilde{w}^\top u + \varepsilon \tilde{w}^\top Z| - \frac{2\varepsilon K_1 \sqrt{n}}{K}.$$

Recall that  $\tilde{w}, u$  are fixed vectors,  $\|\tilde{w}\|_2 = 1$ , and  $Z \sim N_{\mathbb{R}}(0, I_{n-1})$ . Then  $\tilde{w}^\top Z = \theta \gamma$ , where  $\gamma \sim N_{\mathbb{R}}(0, 1)$ , and  $\theta \in \mathbb{C}, |\theta| = 1$ . A quick density calculation yields the following bound on the conditional probability:

$$\mathbb{P}_Z \{ |\tilde{w}^\top u + \varepsilon \tilde{w}^\top Z| \leq \varepsilon \lambda \} \leq C\lambda, \quad \lambda > 0.$$

Therefore, a similar bound holds unconditionally, after intersection with the event  $\mathcal{E}_M^c$ . So we conclude that

$$(3.22) \quad \mathbb{P} \left\{ \inf_{x \in S_1} \|Ax\|_2 \leq \frac{\varepsilon \lambda}{\sqrt{n}} - \frac{2\varepsilon K_1 \sqrt{n}}{K} \text{ and } \mathcal{E}_M^c \right\} \leq C\lambda, \quad \lambda > 0.$$



**3.6. Combining the three cases.** We have obtained lower bounds on  $\inf_{x \in S_1} \|Ax\|_2$  separately in each possible case:

- inequality (3.15) in the case of small denominator (event  $\mathcal{E}_{\text{denom}}$ );
- inequality (3.19) in the case of large denominator, small  $\|M\|$  (event  $\mathcal{E}_{\text{denom}}^c \cap \mathcal{E}_M$ );
- inequality (3.22) in the case of large  $\|M\|$  (event  $\mathcal{E}_M^c$ ).

To combine these three inequalities, we set

$$\mu := \frac{1}{2} \min \left( \frac{\lambda}{K\sqrt{n}}, \frac{\lambda\lambda_1}{\sqrt{n}}, \frac{\lambda}{\sqrt{n}} - \frac{2K_1\sqrt{n}}{K} \right).$$

We conclude that if the condition (3.5) on  $K_0$  and the condition (3.7) on  $\mu$  are satisfied, then

$$\begin{aligned} \mathbb{P} \left\{ \inf_{x \in S_1} \|Ax\|_2 \leq 2\mu\varepsilon \right\} &\leq C\lambda + (C\lambda_1 + \exp(-cK_1^2) + C\lambda) + C\lambda \\ &= 3C\lambda + C\lambda_1 + \exp(-cK_1^2). \end{aligned}$$

Substituting into (3.8), we obtain

$$\mathbb{P} \left\{ \inf_{x \in S_1} \|(D + U)x\|_2 \leq \mu\varepsilon \right\} \leq 3C\lambda + C\lambda_1 + \exp(-cK_1^2) + 2\exp(-cK_0^2n).$$

The same holds for each  $S_i$ ,  $i \in [n]$ . Substituting into (3.1), we get

$$\mathbb{P} \{ s_{\min}(D + U) \leq \mu\varepsilon \} \leq 3C\lambda n + C\lambda_1 n + \exp(-cK_1^2)n + 2\exp(-cK_0^2n)n.$$

This estimate holds for all  $\lambda, \lambda_1, \varepsilon \in (0, 1)$  and all  $K, K_0, K_1 \geq 1$  provided that the conditions (3.5) on  $K_0$  and (3.7) on  $\mu$  are satisfied. So for a given  $\varepsilon \in (0, 1)$ , let us choose

$$\lambda = \lambda_1 = \varepsilon^{0.1}, \quad K_0 = K_1 = \log(1/\varepsilon), \quad K = \frac{4K_1n}{\lambda} = 4\log(1/\varepsilon)n\varepsilon^{-0.1}.$$

Then

$$\mu \gtrsim \frac{\lambda\lambda_1}{K\sqrt{n}} = \frac{\varepsilon^{0.3}}{4\log(1/\varepsilon)n^{3/2}}.$$

Assume that  $\varepsilon \leq c'n^{-4}$  for a sufficiently small absolute constant  $c' > 0$ ; then one quickly checks that the conditions (3.5) and (3.7) are satisfied. For any such  $\varepsilon$  and for the choice of parameters made above, our conclusion becomes

$$\begin{aligned} \mathbb{P} \left\{ s_{\min}(D + U) \leq \frac{\varepsilon^{0.3}}{4\log(1/\varepsilon)n^{3/2}} \right\} \\ \lesssim \varepsilon^{0.1}n + \exp(-c\log^2(1/\varepsilon)) + \exp(-c\log^2(1/\varepsilon)n)n. \end{aligned}$$

Since this estimate is valid for all  $\varepsilon \leq c'n^{-4}$ , this quickly leads to the conclusion of Theorem 1.1. □

## 4. ORTHOGONAL PERTURBATIONS: PROOF OF THEOREM 1.3

In this section we give a formal proof of Theorem 1.3.

## 4.1. Initial reductions of the problem.

4.1.1. *Eliminating dimensions  $n = 2, 3$ .* Since our argument will make use of  $(n - 3) \times (n - 3)$  minors, we would like to assume that  $n > 3$  from now on. This calls for a separate argument in dimensions  $n = 2, 3$ . The following is a somewhat stronger version Theorem 1.3 in these dimensions.

**Theorem 4.1** (Orthogonal perturbations in low dimensions). *Let  $B$  be a fixed  $n \times n$  complex matrix, where  $n \in \{2, 3\}$ . Assume that*

$$(4.1) \quad \|BB^T - I\| \geq \delta \|B\|^2$$

for some  $\delta \in (0, 1)$ . Let  $U$  be a random matrix uniformly distributed in  $O(n)$ . Then

$$\mathbb{P} \{s_{\min}(B + U) \leq t\} \leq C(t/\delta)^c, \quad t > 0.$$

We defer the proof of Theorem 4.1 to Appendix A.

Theorem 4.1 readily implies Theorem 1.3 in dimensions  $n = 2, 3$ . Indeed, the assumptions in (1.2) yield

$$\|DD^T - I\| = \max_i |d_i^2 - 1| \geq \frac{1}{2} \max_{i,j} |d_i^2 - d_j^2| \geq \frac{\delta}{2} \geq \frac{\delta}{2K^2} \|D\|^2.$$

Therefore we can apply Theorem 4.1 with  $\delta/2K^2$  instead of  $\delta$  and obtain

$$\mathbb{P} \{s_{\min}(D + U) \leq t\} \leq C(2K^2t/\delta)^c, \quad t > 0.$$

Thus we conclude a slightly stronger version of Theorem 1.3 in dimensions  $n = 2, 3$ .

*Remark 4.2* (Complex orthogonality). The factor  $\|B\|^2$  cannot be removed from the assumption (4.1). Indeed, it can happen that  $\|BB^T - I\| = 1$  while  $B + U$  is arbitrarily poorly invertible. Such an example is given by the matrix  $B = M \cdot \begin{bmatrix} 1 & i \\ i & -1 \end{bmatrix}$  where  $M \rightarrow \infty$ . Then  $\det(B + U) = 1$  for all  $U \in O(2)$ ;  $\|B + U\| \sim M$ , and thus  $s_{\min}(B + U) \lesssim 1/M \rightarrow 0$ . On the other hand,  $BB^T = 0$ .

This example shows that, surprisingly, staying away from the set of complex orthogonal matrices at (any) constant distance may not guarantee good invertibility of  $B + U$ . It is worthwhile noting that this difficulty does not arise for real matrices  $B$ . For such matrices one can show that the factor  $\|B\|^2$  can be removed from (4.1).

*Remark 4.3.* Theorem 4.1 will be used not only to eliminate the low dimensions  $n = 2$  and  $n = 3$  in the beginning of the argument. We will use it one more time in the heart of the proof, in Subsection 4.3.6 where the problem in higher dimensions  $n$  will get reduced to the invertibility of certain matrices in dimensions  $n = 2, 3$ .

**4.2. Local perturbations and decomposition of the problem.** We can represent  $U$  in Theorem 1.3 as  $U = V^{-1}W$  where  $V, W \in O(n)$  are random independent matrices and  $V$  is uniformly distributed in  $O(n)$  while  $W$  may have arbitrary distribution. We are going to define  $W$  as a small random perturbation of identity.

4.2.1. *The local perturbation  $S$ .* Let  $S$  be an independent random Gaussian skew-symmetric matrix; thus the above-diagonal entries of  $S$  are i.i.d.  $N_{\mathbb{R}}(0, 1)$  random variables and  $S^{\top} = -S$ . By Lemma 3.1,  $W_0 = I + \varepsilon S$  is approximately orthogonal up to error  $O(\varepsilon^2)$ . Although this lemma was stated over the complex numbers, it is evident from the proof that the same result holds over the reals as well (skew-Hermitian is replaced by skew-symmetric, and  $U(n)$  by  $O(n)$ ).

More formally, fix an arbitrary number  $\varepsilon \in (0, 1)$ . Applying the real analog of Lemma 3.1 for the random matrix  $W_0 = I + \varepsilon S$ , we obtain a random matrix  $W \in O(n)$  that satisfies

$$\begin{aligned} s_{\min}(D + U) &= s_{\min}(D + V^{-1}W) = s_{\min}(VD + W) \\ &\geq s_{\min}(VD + W_0) - \|W - W_0\| \\ &\geq s_{\min}(VD + W_0) - 2\varepsilon^2\|S^2\| \quad \text{whenever } \varepsilon^2\|S^2\| \leq 1/4. \end{aligned}$$

Further,  $\|S\| = O(\sqrt{n})$  with high probability. Indeed, let  $K_0 > 1$  be a parameter to be chosen later, which satisfies

$$(4.2) \quad \varepsilon^2 K_0^2 n \leq 1/4.$$

Consider the event

$$(4.3) \quad \mathcal{E}_S := \{\|S\| \leq K_0\sqrt{n}\}; \quad \text{then } \mathbb{P}(\mathcal{E}_S^c) \leq 2 \exp(-cK_0^2 n)$$

provided that

$$(4.4) \quad K_0 > C_0$$

for an appropriately large constant  $C_0$ . Indeed, by rotation invariance  $S$  has the same distribution as  $(\hat{S} - \hat{S}^{\top})/\sqrt{2}$  where  $\hat{S}$  is the matrix with all independent  $N_{\mathbb{R}}(0, 1)$  entries. But for the matrix  $\hat{S}$ , a version of (4.3) is a standard result on random matrices with i.i.d. entries; see [22, Theorem 5.39]. Thus by triangle inequality, (4.3) also holds for  $S$ .

On  $\mathcal{E}_S$ , one has  $\varepsilon^2\|S^2\| \leq 1/4$  due to (4.2), and thus

$$s_{\min}(D + U) \geq s_{\min}(VD + W_0) - 2\varepsilon^2 K_0^2 n.$$

Next, let  $\mu \in (0, 1)$  be a parameter to be chosen later and which satisfies

$$(4.5) \quad \mu \geq 2\varepsilon K_0^2 n.$$

Our ultimate goal will be to estimate

$$p := \mathbb{P} \{s_{\min}(D + U) \leq \mu\varepsilon\}.$$

By the above, we have

$$\begin{aligned} p &\leq \mathbb{P} \{s_{\min}(VD + W_0) \leq \mu\varepsilon + 2\varepsilon^2 K_0^2 n \wedge \mathcal{E}_S\} + \mathbb{P}(\mathcal{E}_S^c) \\ &\leq \mathbb{P} \{s_{\min}(VD + W_0) \leq 2\mu\varepsilon \wedge \mathcal{E}_S\} + \mathbb{P}(\mathcal{E}_S^c) \\ (4.6) \quad &\leq \mathbb{P} \{s_{\min}(A) \leq 2\mu\varepsilon \wedge \mathcal{E}_S\} + 2 \exp(-cK_0^2 n), \end{aligned}$$

where

$$A := VD + I + \varepsilon S.$$

Summarizing, we have introduced a local perturbation  $S$ , which we can assume to be well bounded due to  $\mathcal{E}_S$ . Moreover,  $S$  is independent from  $V$ , which is uniformly distributed in  $O(n)$ .

4.2.2. *Decomposition of the problem.* We are trying to bound below

$$s_{\min}(A) = \inf_{x \in S^{n-1}} \|Ax\|_2.$$

Our immediate task is to reduce the set of vectors  $x$  in the infimum to those with considerable energy in the first two coordinates,  $|x_1|^2 + |x_2|^2 \geq 1/n$ . This will allow us to introduce global perturbations  $R$  and  $Q$ , which will be rotations of the first few (two or three) coordinates.

To this end, note that for every  $x \in S^{n-1}$  there exists a coordinate  $i \in [n]$  such that  $|x_i| \geq n^{-1/2}$ . Therefore

$$(4.7) \quad S^{n-1} = \bigcup_{i \in [n]} S_i \quad \text{where } S_i := \{x \in S^{n-1} : |x_i| \geq n^{-1/2}\}.$$

More generally, given a subset of indices  $J \in [n]$ , we shall work with the set of vectors with considerable energy on  $J$ :

$$S_J := \left\{ x \in S^{n-1} : \sum_{j \in J} x_j^2 \geq 1/n \right\}.$$

Note that the sets  $S_J$  increase by inclusion:

$$J_1 \subseteq J_2 \quad \text{implies} \quad S_{J_1} \subseteq S_{J_2}.$$

To simplify the notation, we shall write  $S_{1,2,3}$  instead of  $S_{\{1,2,3\}}$ , etc.

Using (4.7), we decompose the event we are trying to estimate as follows:

$$\{s_{\min}(A) \leq 2\mu\varepsilon\} = \bigcup_{i \in [n]} \left\{ \inf_{x \in S_i} \|Ax\|_2 \leq 2\mu\varepsilon \right\}.$$

Next, for every  $i \in [n]$ ,

$$\max_{j \in [n]} |d_i^2 - d_j|^2 \geq \frac{1}{2} \max_{i,j \in [n]} |d_i^2 - d_j|^2,$$

so the second assumption in (1.3) implies that there exists  $j = j(i) \in [n]$ ,  $j \neq i$ , such that  $|d_i^2 - d_{j(i)}^2| \geq \delta$ .

Since  $S_i \subseteq S_{i,j(i)}$ , we obtain from the above and (4.6) that

$$(4.8) \quad \begin{aligned} p &\leq \sum_{i=1}^n \mathbb{P} \left\{ \inf_{x \in S_{i,j(i)}} \|Ax\|_2 \leq 2\mu\varepsilon \wedge \mathcal{E}_S \right\} + 2 \exp(-cK_0^2 n) \\ &=: \sum_{i=1}^n p_i + 2 \exp(-cK_0^2 n). \end{aligned}$$

We reduced the problem to estimating each term  $p_i$ . This task is similar for each  $i$ , so without loss of generality we can focus on  $i = 1$ . Furthermore, without loss of generality we can assume that  $j(1) = 2$ . Thus our goal is to estimate

$$p_1 = \mathbb{P} \left\{ \inf_{x \in S_{1,2}} \|Ax\|_2 \leq 2\mu\varepsilon \wedge \mathcal{E}_S \right\}$$

under the assumption that

$$(4.9) \quad |d_1^2 - d_2^2| \geq \delta.$$

Finally, we further decompose the problem according to whether there exists a well invertible  $(n - 3) \times (n - 3)$  minor of  $A_{(1,2)}$  or not. Why we need to consider these cases was explained informally in Section 2.2.3.

Let  $K_1 \geq 1$  be a parameter to be chosen later. By a union bound, we have

$$\begin{aligned}
 p_1 &\leq \sum_{i=3}^n \mathbb{P} \left\{ \inf_{x \in S_{1,2}} \|Ax\|_2 \leq 2\mu\varepsilon \wedge \|(A_{(1,2,i)})^{-1}\| \leq \frac{K_1}{\varepsilon} \wedge \mathcal{E}_S \right\} \\
 &\quad + \mathbb{P} \left\{ \inf_{x \in S_{1,2}} \|Ax\|_2 \leq 2\mu\varepsilon \wedge \|(A_{(1,2,i)})^{-1}\| > \frac{K_1}{\varepsilon} \forall i \in [3 : n] \wedge \mathcal{E}_S \right\} \\
 (4.10) \quad &=: \sum_{i=3}^n p_{1,i} + p_{1,0}.
 \end{aligned}$$

**4.3. When a minor is well invertible: Going 3 dimensions up.** In this section we shall estimate the probabilities  $p_{1,i}$ ,  $i = 3, \dots, n$ , in the decomposition (4.10). All of them are similar, so without loss of generality we can focus on estimating  $p_{1,3}$ . Since  $S_{1,2} \subseteq S_{1,2,3}$ , we have

$$(4.11) \quad p_{1,3} \leq \mathbb{P} \left\{ \inf_{x \in S_{1,2,3}} \|Ax\|_2 \leq 2\mu\varepsilon \wedge \|(A_{(1,2,3)})^{-1}\| \leq \frac{K_1}{\varepsilon} \wedge \mathcal{E}_S \right\}.$$

This is the same as the original invertibility problem, except now we have three extra pieces of information: (a) the minor  $A_{(1,2,3)}$  is well invertible; (b) the vectors in  $S_{1,2,3}$  over which we are proving invertibility have large energy in the first three coordinates; (c) the local perturbation  $S$  is well bounded.

**4.3.1. The global perturbations  $Q, R$ .** The core argument in this case will rely on global perturbations (rather than the local perturbation  $S$ ), which we shall now introduce into the matrix  $A = VD + I + \varepsilon S$ . Define

$$Q := \begin{bmatrix} Q_0 & 0 \\ 0 & \tilde{I} \end{bmatrix}, \quad R := \begin{bmatrix} R_0 & 0 \\ 0 & \tilde{I} \end{bmatrix}$$

where  $Q_0 \in SO(3)$  and  $R_0 \in O(3)$  are independent uniform random matrices and  $\tilde{I}$  denotes the identity on  $\mathbb{C}^{[3:n]}$ .

Let us condition on  $Q$  and  $R$  for a moment. By the rotation invariance of the random orthogonal matrix  $V$  and of the Gaussian skew-symmetric matrix  $S$ , the (conditional) joint distribution of the pair  $(V, S)$  is the same as that of  $(Q^T RVQ, Q^T SQ)$ . Therefore, the conditional distribution of  $A$  is the same as that of

$$Q^T RVQD + I + \varepsilon Q^T SQ = Q^T (RVQDQ^T + I + \varepsilon S)Q =: \hat{A}.$$

Let us go back to estimating  $p_{1,3}$  in (4.11). Since  $A$  and  $\hat{A}$  are identically distributed and the event  $\mathcal{E}_S$  does not change when  $S$  is replaced by  $Q^T SQ$ , the conditional probability

$$\mathbb{P} \left\{ \inf_{x \in S_{1,2,3}} \|Ax\|_2 \leq 2\mu\varepsilon \wedge \|(A_{(1,2,3)})^{-1}\| \leq \frac{K_1}{\varepsilon} \wedge \mathcal{E}_S \mid Q, R \right\}$$

does not change when  $A$  is replaced by  $\hat{A}$ . Taking expectations with respect to  $Q$  and  $R$ , we see that the full (unconditional) probability does not change either, so

$$(4.12) \quad p_{1,3} \leq \mathbb{P} \left\{ \inf_{x \in S_{1,2,3}} \|\hat{A}x\|_2 \leq 2\mu\varepsilon \wedge \|(\hat{A}_{(1,2,3)})^{-1}\| \leq \frac{K_1}{\varepsilon} \wedge \mathcal{E}_S \right\}.$$

4.3.2. *Randomizing D.* Let us try to understand the terms appearing in  $\widehat{A}$ . We think of

$$\widetilde{D} := QDQ^\top$$

as a randomized version of  $D$  obtained by a random change of basis in the first three coordinates. Then we can express

$$\widehat{A} = Q^\top \widetilde{A} Q \quad \text{where} \quad \widetilde{A} = RV\widetilde{D} + I + \varepsilon S.$$

Compared to  $A = VD + I + \varepsilon S$ , the random matrix  $\widetilde{A}$  incorporates the global perturbations  $Q$  and  $R$ . Thus we seek to replace  $A$  with  $\widetilde{A}$  in our problem. To this end, let us simplify two quantities that appear in (4.12).

First,

$$(4.13) \quad \inf_{x \in S_{1,2,3}} \|\widehat{A}x\|_2 = \inf_{x \in S_{1,2,3}} \|\widetilde{A}x\|_2$$

since  $Q(S_{1,2,3}) = S_{1,2,3}$  by definition and  $Q^\top \in SO(n)$ . Second, using that  $Q$  and  $R$  affect only the first three coordinates and since  $D$  is diagonal, one checks that

$$(4.14) \quad \widehat{A}_{(1,2,3)} = \widetilde{A}_{(1,2,3)} = (V\widetilde{D} + I + \varepsilon S)_{(1,2,3)} = (VD + I + \varepsilon S)_{(1,2,3)}.$$

Similarly to previous matrices, we decompose  $S$  as

$$(4.15) \quad S = \begin{bmatrix} S_0 & -Z^\top \\ Z & \check{S} \end{bmatrix}, \quad \text{where } S_0 \in \mathbb{R}^{3 \times 3}, \check{S} \in \mathbb{R}^{(n-3) \times (n-3)}.$$

Note that  $S_0$ ,  $\check{S}$ , and  $Z$  are independent and that  $Z \in \mathbb{R}^{(n-3) \times 3}$  is a random matrix with all i.i.d.  $N_{\mathbb{R}}(0, 1)$  entries.

By (4.14),  $\widehat{A}_{(1,2,3)}$  is independent of  $S_0$ ,  $Z$ ,  $Q$ ,  $R$ ; it only depends on  $V$  and  $\check{S}$ . Let us condition on  $S_0$ ,  $\check{S}$ , and  $V$  and thus fix  $\widehat{A}_{(1,2,3)}$  such that the invertibility condition in (4.12) is satisfied, i.e. such that

$$\|(\widehat{A}_{(1,2,3)})^{-1}\| \leq \frac{K_1}{\varepsilon}$$

(otherwise the corresponding conditional probability is automatically zero). All randomness remains in the local perturbation  $Z$  and the global perturbations  $Q$ ,  $R$ .

Let us summarize our findings. Recalling (4.13), we have shown that

$$(4.16) \quad p_{1,3} \leq \inf_{S_0, \check{S}, V} \mathbb{P}_{Z, Q, R} \left\{ \inf_{x \in S_{1,2,3}} \|\widetilde{A}x\|_2 \leq 2\mu\varepsilon \wedge \mathcal{E}_S \right\}$$

where

$$\widetilde{A} = RV\widetilde{D} + I + \varepsilon S, \quad \widetilde{D} := QDQ^\top,$$

where  $S$  is decomposed as in (4.15) and where the infimum is over all  $\check{S}$ ,  $V$  satisfying

$$(4.17) \quad \|(VD + I + \varepsilon S)_{(1,2,3)}^{-1}\| \leq \frac{K_1}{\varepsilon}.$$

Compared with (4.11), we have achieved the following: we introduced into the problem global perturbations  $Q$ ,  $R$  acting on the first three coordinates.  $Q$  randomizes the matrix  $D$  and  $R$  serves as a further global rotation.

4.3.3. *Reducing to the invertibility of random  $3 \times 3$  matrices.* Let us decompose the matrix  $\tilde{A} = RV\tilde{D} + I + \varepsilon S$  by revealing its first three rows and columns as before. To this end, recall that  $R = \begin{bmatrix} R_0 & 0 \\ 0 & \check{I} \end{bmatrix}$  and  $Q = \begin{bmatrix} Q_0 & 0 \\ 0 & \check{I} \end{bmatrix}$ . We similarly decompose

$$V =: \begin{bmatrix} V_0 & v \\ u & \check{V} \end{bmatrix}$$

and

$$(4.18) \quad D =: \begin{bmatrix} D_0 & 0 \\ 0 & \check{D} \end{bmatrix}; \quad \text{then} \quad \tilde{D} = \begin{bmatrix} \tilde{D}_0 & 0 \\ 0 & \check{D} \end{bmatrix} \quad \text{where} \quad \tilde{D}_0 := Q_0 D Q_0^\top.$$

Using these and the decomposition of  $S$  in (4.15), we decompose

$$(4.19) \quad \tilde{A} = RV\tilde{D} + I + \varepsilon S = \begin{bmatrix} R_0 V_0 \tilde{D}_0 + I_0 + \varepsilon S_0 & R_0 v \check{D} - \varepsilon Z^\top \\ u \tilde{D}_0 + \varepsilon Z & \check{V} \check{D} + \check{I} + \varepsilon \check{S} \end{bmatrix} =: \begin{bmatrix} H_0 & Y \\ X & \check{H} \end{bmatrix}$$

where  $I_0$  denotes the identity in  $\mathbb{C}^3$ .

Note that

$$\check{H} = \check{V} \check{D} + \check{I} + \varepsilon \check{S} = (VD + I + \varepsilon S)_{(1,2,3)}$$

is a well-invertible matrix by (4.17); namely

$$(4.20) \quad \|\check{H}^{-1}\| \leq \frac{K_1}{\varepsilon}.$$

The next lemma reduces the invertibility of  $\tilde{A}$  to the invertibility of a  $3 \times 3$  matrix  $H_0 - Y\check{H}^{-1}X$ .

**Lemma 4.4** (Invertibility of a matrix with a well-invertible minor). *Consider a matrix*

$$H = \begin{bmatrix} H_0 & Y \\ X & \check{H} \end{bmatrix} \quad \text{where} \quad H_0 \in \mathbb{C}^{3 \times 3}, \quad \check{H} \in \mathbb{C}^{(n-3) \times (n-3)}.$$

Assume that

$$\|\check{H}^{-1}\| \leq L_1, \quad \|Y\| \leq L_2$$

for some  $L_1, L_2 > 0$ . Then

$$\inf_{x \in S_{1,2,3}} \|Hx\|_2 \geq \frac{1}{\sqrt{n}(1 + L_1 L_2)} s_{\min}(H_0 - Y\check{H}^{-1}X).$$

*Proof.* Choose  $x \in S_{1,2,3}$  which attains  $\inf_{x \in S_{1,2,3}} \|Hx\|_2 =: \delta$  and decompose it as

$$x =: \begin{bmatrix} x_0 \\ \check{x} \end{bmatrix} \quad \text{where} \quad x_0 \in \mathbb{C}^3, \quad \check{x} \in \mathbb{C}^{n-3}.$$

Then

$$Hx = \begin{bmatrix} H_0 x_0 + Y \check{x} \\ X x_0 + \check{H} \check{x} \end{bmatrix}.$$

The assumption  $\|Hx\|_2 = \delta$  then leads to the system of inequalities

$$\begin{cases} \|H_0 x_0 + Y \check{x}\|_2 \leq \delta, \\ \|X x_0 + \check{H} \check{x}\|_2 \leq \delta. \end{cases}$$

We solve these inequalities in a standard way. Multiplying the second inequality by  $\|\check{H}^{-1}\|$ , we obtain

$$\|\check{H}^{-1} X x_0 + \check{x}\|_2 \leq \delta \|\check{H}^{-1}\| \leq \delta L_1,$$

which informally means that  $\tilde{x} \approx -\check{H}^{-1}Xx_0$ . Replacing  $\tilde{x}$  with  $-\check{H}^{-1}Xx_0$  in the first equation and estimating the error by the triangle inequality, we arrive at

$$\begin{aligned} \|H_0x_0 - Y\check{H}^{-1}Xx_0\|_2 &\leq \|H_0x_0 + Y\tilde{x}\|_2 + \|Y\tilde{x} + Y\check{H}^{-1}Xx_0\|_2 \\ &\leq \delta + \|Y\| \cdot \|\tilde{x} + \check{H}^{-1}Xx_0\|_2 \\ &\leq \delta + L_2 \cdot \delta L_1 = \delta(1 + L_1L_2). \end{aligned}$$

Note that the left-hand side is  $\|(H_0 - Y\check{H}^{-1}X)x_0\|_2$  and that  $\|x_0\|_2 \geq 1/\sqrt{n}$  since  $x \in S_{1,2,3}$ . By the definition of the smallest singular value, it follows that

$$s_{\min}(H_0 - Y\check{H}^{-1}X) \leq \sqrt{n} \delta(1 + L_1L_2).$$

Rearranging the terms concludes the proof of the lemma. □

In order to apply Lemma 4.4 for the matrix  $\tilde{A}$  in (4.19), let us check that the boundedness assumptions are satisfied. We already know that  $\|\check{H}^{-1}\| \leq K_1/\varepsilon$  from (4.20). Further,

$$\|Y\| = \|R_0v\check{D} - \varepsilon Z^\top\|.$$

Here,  $\|R_0\| = 1$ ,  $\|v\| \leq \|V\| \leq 1$ ,  $\|\check{D}\| \leq \|D\| \leq K$  by the assumption of the theorem, and  $\|Z\| \leq \|S\| \leq K_0\sqrt{n}$  if the event  $\mathcal{E}_S$  holds. Putting these together, we have

$$\|Y\| \leq K + \varepsilon K_0\sqrt{n} \leq 2K$$

where the last inequality follows from (4.2). An application of Lemma 4.4 yields that on the event  $\mathcal{E}_S$  one has

$$(4.21) \quad \inf_{x \in S_{1,2,3}} \|\tilde{A}x\|_2 \geq \frac{\varepsilon}{3KK_1\sqrt{n}} s_{\min}(H_0 - YMX)$$

where

$$(4.22) \quad M := \check{H}^{-1}, \quad \|M\| \leq \frac{K_1}{\varepsilon}.$$

We have reduced our problem to the invertibility of the  $3 \times 3$  matrix  $H_0 - YMX$ .

*4.3.4. Dependence on the global perturbation R.* Let us write our random matrix  $H_0 - YMX$  as a function of the global perturbation  $R_0$  (which determines  $R$ ). Recalling (4.19), we have

$$H_0 - YMX = R_0V_0\tilde{D}_0 + I_0 + \varepsilon S_0 - (R_0v\check{D} - \varepsilon Z^\top)M(u\tilde{D}_0 + \varepsilon Z) = a + R_0b,$$

where

$$(4.23) \quad \begin{aligned} a &:= I_0 + \varepsilon S_0 + \varepsilon Z^\top Mu\tilde{D}_0 + \varepsilon^2 Z^\top MZ, \\ b &:= V_0\tilde{D}_0 - v\check{D}M(u\tilde{D}_0 + \varepsilon Z). \end{aligned}$$

It will be helpful to simplify  $a$  and  $b$ . We shall first remove the terms  $\varepsilon S_0$  and  $\varepsilon^2 Z^\top MZ$  from  $a$  and then (in the next subsection) remove all other terms from  $a$  and  $b$  that depend on  $Z$ .



To achieve the first step, observe that on the event  $\mathcal{E}_S$ , we have

$$\|\varepsilon S_0\| \leq \varepsilon K_0 \sqrt{n};$$

$$\begin{aligned} \|\varepsilon^2 Z^\top MZ\| &\leq \varepsilon^2 \|Z\|^2 \|M\| \leq \varepsilon^2 \|S\|^2 \|M\| \\ &\leq \varepsilon^2 \cdot K_0^2 n \cdot \frac{K_1}{\varepsilon} \quad (\text{by definition of } \mathcal{E}_S \text{ and (4.22)}) \\ &= \varepsilon K_0^2 K_1 n. \end{aligned}$$

Therefore we can approximate  $a$  by the following simpler quantity:

$$(4.24) \quad \begin{aligned} a_0 &:= I_0 + \varepsilon Z^\top M u \tilde{D}_0, \\ \|a - a_0\|_2 &\leq \varepsilon K_0 \sqrt{n} + \varepsilon K_0^2 K_1 n \leq 2\varepsilon K_0^2 K_1 n. \end{aligned}$$

Hence we can replace  $a$  by  $a_0$  in our problem of estimating

$$(4.25) \quad s_{\min}(H_0 - YMX) = s_{\min}(a + R_0b) \geq s_{\min}(a_0 + R_0b) - 2\varepsilon K_0^2 K_1 n.$$

We have reduced the problem to the invertibility of the  $3 \times 3$  random matrix  $a + R_0b$ .

4.3.5. *Removing the local perturbation  $Z$ .* As we mentioned in the introduction, the argument in this case (when the minor is well invertible) relies on global perturbations only. This is the time when we remove the local perturbation  $Z$  from our problem. To this end, we express  $a_0 + R_0b$  as a function of  $Z$  using (4.24) and (4.23):

$$a_0 + R_0b = L + Z^\top \varepsilon M u \tilde{D}_0 - R_0 v \check{D} M \varepsilon Z,$$

where

$$(4.26) \quad L := I_0 + R_0(V_0 - v \check{D} M u) \tilde{D}_0.$$

If we condition on everything but  $Z$ , we can view  $a_0 + R_0b$  as a Gaussian perturbation of the fixed matrix  $L$ . It will then be easy to show that  $a_0 + R_0b$  is well invertible whenever  $L$  is. This will reduce the problem to the invertibility of  $L$ ; the local perturbation  $Z$  will thus be removed from the problem.

Formally, let  $\lambda_1 \in (0, 1)$  be a parameter to be chosen later; we define the event

$$\mathcal{E}_L := \{s_{\min}(L) \geq \lambda_1\}.$$

Note that  $\mathcal{E}_L$  is determined by  $R_0, Q_0$  and is independent of  $Z$ .

Let us condition on  $R_0, Q_0$  satisfying  $\mathcal{E}_L$ . Then

$$(4.27) \quad s_{\min}(a_0 + R_0b) \geq s_{\min}(L) \cdot s_{\min}(L^{-1}(a_0 + R_0b)) \geq \lambda_1 \cdot s_{\min}(I_0 + f(Z))$$

where

$$(4.28) \quad f(Z) := L^{-1} Z^\top \varepsilon M u \tilde{D}_0 - L^{-1} R_0 v \check{D} M \varepsilon Z$$

is a linear function of (the entries of)  $Z$ . A good invertibility of  $I_0 + f(Z)$  is guaranteed by the following lemma.

**Lemma 4.5** (Invertibility of Gaussian perturbations). *Let  $m \geq 1$ , and let  $f : \mathbb{R}^m \rightarrow \mathbb{C}^{3 \times 3}$  be a linear (matrix-valued) transformation. Assume that  $\|f\| \leq K$  for some  $K \geq 1$ , i.e.  $\|f(z)\|_{\text{HS}} \leq K\|z\|_2$  for all  $z \in \mathbb{R}^m$ . Let  $Z \sim N_{\mathbb{R}}(0, I_m)$ . Then*

$$\mathbb{P} \{s_{\min}(I + f(Z)) \leq t\} \leq CKt^{1/4}, \quad t > 0.$$

We defer the proof of this lemma to Section B.2.

We will use Lemma 4.5 with  $m = 3(n - 3)$ , rewriting the entries of the  $(n - 3) \times 3$  matrix  $Z$  as coordinates of a vector in  $\mathbb{R}^m$ . In order to apply this lemma, let us bound  $\|f(Z)\|_{\text{HS}}$  in (4.28). To this end, note that  $\|L^{-1}\| \leq \lambda_1^{-1}$  if the event  $\mathcal{E}_L$  occurs;  $\|M\| \leq K_1/\varepsilon$  by (4.22);  $\|u\| \leq \|U\| = 1$ ;  $\|v\| \leq \|V\| \leq 1$ ;  $\|\tilde{D}_0\| = \|D_0\| \leq \|D\| \leq K$ ;  $\|\tilde{D}\| \leq \|D\| \leq K$ ;  $\|R_0\| = 1$ . It follows that

$$\|f(Z)\|_{\text{HS}} \leq 2\lambda_1^{-1}KK_1\|Z\|_{\text{HS}}.$$

An application of Lemma 4.5 then yields

$$\mathbb{P}_Z \{s_{\min}(I_0 + f(Z)) \leq t\} \leq C\lambda_1^{-1}KK_1t^{1/4}, \quad t > 0.$$

Putting this together with (4.27), we have shown the following. Conditionally on  $R_0, Q_0$  satisfying  $\mathcal{E}_L$ , the matrix  $a_0 + R_0b$  is well invertible:

$$(4.29) \quad \mathbb{P}_Z \{s_{\min}(a_0 + R_0b) \leq \lambda_1 t\} \leq C\lambda_1^{-1}KK_1t^{1/4}, \quad t > 0.$$

This reduces the problem to showing that event  $\mathcal{E}_L$  is likely, namely that the random matrix  $L$  in (4.26) is well invertible. The local perturbation  $Z$  has been removed from the problem.

4.3.6. *Invertibility in dimension 3.* In showing that  $L$  is well invertible, the global perturbation  $R_0$  will be crucial. Recall that

$$L = I_0 + R_0B, \quad \text{where } B = (V_0 - v\check{D}Mu)\tilde{D}_0.$$

Then  $s_{\min}(L) = s_{\min}(B + R_0^{-1})$ . If we condition on everything but  $R_0$ , we arrive at the invertibility problem for the perturbation of the fixed matrix  $B$  by a random matrix  $R_0$  uniformly distributed in  $O(3)$ . This is the same kind of problem that our main theorems are about, however for  $3 \times 3$  matrices. But recall that in dimension 3 the main result has already been established in Theorem 4.1. It guarantees that  $B + R_0^{-1}$  is well invertible whenever  $B$  is not approximately complex orthogonal, i.e. whenever  $\|BB^T - I\| \gtrsim \|B\|^2$ . This argument reduces our problem to breaking complex orthogonality for  $B$ .

Formally, let  $\lambda_2 \in (0, 1)$  be a parameter to be chosen later; we define the event

$$\mathcal{E}_B := \{\|BB^T - I\| \geq \lambda_2\|B\|^2\}.$$

Note that  $\mathcal{E}_B$  is determined by  $Q_0$  and is independent of  $R_0$ .

Let us condition on  $Q_0$  satisfying  $\mathcal{E}_B$ . Theorem 4.1 then implies that

$$(4.30) \quad \mathbb{P}_{R_0}(\mathcal{E}_L^c) = \mathbb{P}_{R_0} \{s_{\min}(L) < \lambda_1\} = \mathbb{P}_{R_0} \{s_{\min}(B + R_0^{-1}) < \lambda_1\} \leq C(\lambda_1/\lambda_2)^c.$$

This reduces the problem to showing that  $\mathcal{E}_B$  is likely, i.e. that  $B$  is not approximately complex orthogonal.

4.3.7. *Breaking complex orthogonality.* Recall that

$$B = (V_0 - v\check{D}Mu)\check{D}_0 =: T\check{D}_0, \quad \check{D}_0 := Q_0D_0Q_0^T,$$

where  $Q_0$  is a random matrix uniformly distributed in  $SO(3)$ . Thus  $\check{D}_0$  is a randomized version of  $D_0$  obtained by a random change of basis.

Let us condition on everything but  $Q_0$ , leaving  $B$  fixed. The following general result states that if  $D_0$  is not near a multiple of identity, then  $T$  is not approximately complex orthogonal with high probability.

**Lemma 4.6** (Breaking complex orthogonality). *Let  $n \in \{2, 3\}$ . Let  $D = \text{diag}(d_i) \in \mathbb{C}^{n \times n}$ . Assume that*

$$\max_i |d_i| \leq K, \quad |d_1^2 - d_2^2| \geq \delta$$

for some  $K, \delta > 0$ . Let  $T \in \mathbb{C}^{n \times n}$ . Let  $Q$  be uniformly distributed in  $SO(n)$  and consider the random matrix  $B := A Q D Q^T$ . Then

$$(4.31) \quad \mathbb{P} \{ \|BB^T - I\| \leq t \|B\|^2 \} \leq C(tK^2/\delta)^c, \quad t > 0.$$

We defer the proof of this lemma to Section B.3.

Let us apply Lemma 4.6 for  $D_0 = \text{diag}(d_1, d_2, d_3)$ . Recall that the assumptions of the lemma are satisfied by (1.2) and (4.9). Then an application of the lemma with  $t = \lambda_2$  yields that

$$(4.32) \quad \mathbb{P}_{Q_0}(\mathcal{E}_B^c) = \mathbb{P}_{Q_0} \{ \|BB^T - I\| < \lambda_2 \|B\|^2 \} \leq C(\lambda_2 K^2/\delta)^c.$$

This was the remaining piece to be estimated, and now we can collect all the pieces together.

4.3.8. *Putting all the pieces together.* By (4.30) and (4.32), we have

$$\begin{aligned} \mathbb{P}_{Q_0, R_0}(\mathcal{E}_L^c) &= \mathbb{E}_{Q_0} \mathbb{P}_{R_0}(\mathcal{E}_L^c | Q_0) \leq \mathbb{E}_{Q_0} \mathbb{P}_{R_0}(\mathcal{E}_L^c | Q_0) \mathbf{1}_{\{Q_0 \text{ satisfies } \mathcal{E}_B\}} + \mathbb{P}_{Q_0}(\mathcal{E}_B^c) \\ &\leq C(\lambda_1/\lambda_2)^c + C(\lambda_2 K^2/\delta)^c. \end{aligned}$$

By a similar conditional argument, this estimate and (4.29) yield

$$\mathbb{P}_{Q_0, R_0, Z} \{ s_{\min}(a_0 + R_0 b) \leq \lambda_1 t \wedge \mathcal{E}_S \} \leq q,$$

where

$$(4.33) \quad q := C\lambda_1^{-1} K K_1 t^{1/4} + C(\lambda_1/\lambda_2)^c + C(\lambda_2 K^2/\delta)^c.$$

Obviously, we can choose  $C > 1$  and  $c < 1$ . By (4.25),

$$\mathbb{P}_{Z, Q_0, R_0} \{ s_{\min}(H_0 - YMX) < \lambda_1 t - 2\varepsilon K_0^2 K_1 n \wedge \mathcal{E}_S \} \leq q$$

and further by (4.21), we obtain

$$\mathbb{P}_{Z, Q_0, R_0} \left\{ \inf_{x \in S_{1,2,3}} \|\tilde{A}x\|_2 < \frac{\varepsilon(\lambda_1 t - 2\varepsilon K_0^2 K_1 n)}{3K K_1 \sqrt{n}} \wedge \mathcal{E}_S \right\} \leq q.$$

Thus we have successfully estimated  $p_{1,3}$  in (4.16) and in (4.11):

$$(4.34) \quad p_{1,3} \leq q \quad \text{for } \mu = \frac{\lambda_1 t - 2\varepsilon K_0^2 K_1 n}{3K K_1 \sqrt{n}}$$

and where  $q$  is defined in (4.33). By an identical argument, the same estimate holds for all  $p_{1,i}$  in the sum (4.10):

$$(4.35) \quad p_{1,i} \leq q, \quad i = 3, \dots, n.$$

Summarizing, we achieved the goal of this section, which was to show that  $A$  is well invertible on the set  $S_{1,2}$  in the case when there is a well-invertible minor  $A_{(1,2,i)}$ .

*Remark 4.7* (Doing the same for  $(n - 2) \times (n - 2)$  minors). One can carry on the argument of this section in a similar way for  $(n - 2) \times (n - 2)$  minors and thus obtain the same estimate for the probability

$$\mathbb{P} \left\{ \inf_{x \in S_{1,2}} \|Ax\|_2 \leq 2\mu\varepsilon \wedge \|(A_{(1,2)})^{-1}\| \leq \frac{K_1}{\varepsilon} \wedge \mathcal{E}_S \right\}$$

as we obtained in (4.34) for the probability  $p_{1,3}$  in (4.11).

**4.4. When all minors are poorly invertible: Going 1 + 2 dimensions up.** In this section we estimate the probability  $p_{1,0}$  in the decomposition (4.10), i.e.

$$(4.36) \quad p_{1,0} = \mathbb{P} \left\{ \inf_{x \in S_{1,2}} \|Ax\|_2 \leq 2\mu\varepsilon \wedge \|(A_{(1,2,i)})^{-1}\| > \frac{K_1}{\varepsilon} \forall i \in [3 : n] \wedge \mathcal{E}_S \right\}.$$

4.4.1. *Invertibility of a matrix with a poorly invertible minor.* The following analog of Lemma 4.4 for a poorly invertible minor will be helpful in estimating  $p_{1,0}$ . Unfortunately, it only works for  $(n - 1) \times (n - 1)$  minors rather than  $(n - 3) \times (n - 3)$  or  $(n - 2) \times (n - 2)$  minors.

**Lemma 4.8** (Invertibility of a matrix with a poorly invertible minor). *Consider an  $n \times n$  matrix*

$$H = \begin{bmatrix} H_0 & Y \\ X & \check{H} \end{bmatrix} \quad \text{where } H_0 \in \mathbb{C}, \check{H} \in \mathbb{C}^{(n-1) \times (n-1)}.$$

*Assume that  $X \sim N_{\mathbb{R}}(\nu, \varepsilon^2 I_{n-1})$  for some fixed  $\nu \in \mathbb{C}^{n-1}$  and  $\varepsilon > 0$ .<sup>2</sup> We assume also that  $\check{H}$  is a fixed matrix satisfying*

$$\|\check{H}^{-1}\| \geq L$$

*for some  $L > 0$ , while  $H_0$  and  $Y$  may be arbitrary, possibly random, and correlated with  $X$ . Then*

$$\mathbb{P} \left\{ \inf_{x \in S_1} \|Hx\|_2 \leq \frac{t\varepsilon}{\sqrt{n}} - \frac{1}{L} \right\} \leq Ct\sqrt{n}, \quad t > 0.$$

*Proof.* Choose  $x \in S_1$  which attains  $\inf_{x \in S_1} \|Hx\|_2 =: \delta$  and decompose it as

$$x =: \begin{bmatrix} x_0 \\ \check{x} \end{bmatrix} \quad \text{where } x_0 \in \mathbb{C}, \check{x} \in \mathbb{C}^{n-1}.$$

As in the proof of Lemma 4.4, we deduce that

$$\|\check{H}^{-1}Xx_0 + \check{x}\|_2 \leq \delta \|\check{H}^{-1}\|.$$

This yields

$$\|\check{x}\|_2 \geq \|\check{H}^{-1}Xx_0\|_2 - \delta \|\check{H}^{-1}\|.$$

Note that

$$\|\check{H}^{-1}Xx_0\|_2 = |x_0| \|\check{H}^{-1}X\| \geq \frac{1}{\sqrt{n}} \|\check{H}^{-1}X\|,$$

where the last inequality is due to  $x \in S_1$ .

---

<sup>2</sup>Although  $X$  is complex-valued,  $X - \nu$  is a real-valued variable distributed according to  $N(0, \varepsilon^2 I_{n-1})$ .

Further, we have  $\|\check{H}^{-1}X\| \sim \|\check{H}^{-1}\|_{\text{HS}}$  by standard concentration techniques. We state and prove such a result in Lemma B.2 in Appendix B. It yields that

$$\mathbb{P} \{ \|\check{H}^{-1}X\|_2 \leq t\varepsilon \|\check{H}^{-1}\|_{\text{HS}} \} \leq Ct\sqrt{n}, \quad t > 0.$$

Next, when this unlikely event does not occur, i.e. when  $\|\check{H}^{-1}X\|_2 > t\varepsilon \|\check{H}^{-1}\|_{\text{HS}}$ , we have

$$\|\check{x}\|_2 \geq \frac{t\varepsilon}{\sqrt{n}} \|\check{H}^{-1}\|_{\text{HS}} - \delta \|\check{H}^{-1}\| \geq \left( \frac{t\varepsilon}{\sqrt{n}} - \delta \right) \|\check{H}^{-1}\| \geq \left( \frac{t\varepsilon}{\sqrt{n}} - \delta \right) L.$$

On the other hand,  $\|\check{x}\|_2 \leq \|x\|_2 = 1$ . Substituting and rearranging the terms yields

$$\delta \geq \frac{t\varepsilon}{\sqrt{n}} - \frac{1}{L}.$$

This completes the proof of Lemma 4.8. □

4.4.2. *Going one dimension up.* As we outlined in Section 2.2.3, the probability  $p_{1,0}$  in (4.36) will be estimated in two steps. At the first step, which we carry on in this section, we explore the condition that all  $(n - 3) \times (n - 3)$  minors of  $A_{1,2}$  are poorly invertible:

$$\|(A_{(1,2,i)})^{-1}\| > \frac{K_1}{\varepsilon} \quad \forall i \in [3 : n].$$

Using Lemma 4.8 in dimension  $n - 2$ , we will conclude that the matrix  $A_{(1,2)}$  is well invertible on the set of vectors with a large  $i$ -th coordinate. Since this happens for all  $i$ , the matrix  $A_{(1,2)}$  is well invertible on all vectors; i.e.  $\|A_{(1,2)}^{-1}\|$  is not too large. This step will thus carry us one dimension up, from poor invertibility of all minors in dimension  $n - 3$  to a good invertibility of the minor in dimension  $n - 2$ .

Since we will be working in dimensions  $[3 : n]$  during this step, we introduce the appropriate notation analogous to (4.7) restricted to these dimensions. Thus  $S^{[3:n]}$  will denote the unit Euclidean sphere in  $\mathbb{C}^{[3:n]}$ , so

$$(4.37) \quad S^{[3:n]} = \bigcup_{i \in [3:n]} S_i^{[3:n]} \quad \text{where } S_i^{[3:n]} := \{x \in S^{[3:n]} : |x_i| \geq n^{-1/2}\}.$$

We apply Lemma 4.8 for

$$H = A_{(1,2)}, \quad \check{H} = A_{(1,2,3)}, \quad L = \frac{K_1}{\varepsilon}, \quad t = \frac{2\sqrt{n}}{K_1}.$$

Recall from Section 4.2.1 that

$$A = \begin{bmatrix} * & * & * & \dots \\ * & * & * & \dots \\ * & * & H_0 & Y \\ \vdots & \vdots & X & \check{H} \end{bmatrix} = VD + I + \varepsilon S,$$

where  $S$  is a skew-symmetric Gaussian random matrix (with i.i.d.  $N_{\mathbb{R}}(0, 1)$  above-diagonal entries). Let us condition on everything except the entries  $S_{ij}$  with  $i \in [4 : n]$ ,  $j = 3$  and with  $i = 3$ ,  $j \in [4 : n]$ , since these entries define the parts  $X, Y$  of  $H$ . Note that  $X = \nu + \varepsilon S^{(3)}$ , where the vector  $\nu \in \mathbb{C}^{n-3}$  is independent of  $S$  and  $S^{(3)}$  is a standard real Gaussian vector with coordinates  $S_{4,3}, \dots, S_{n,3}$ .

Lemma 4.8 used with  $t = \frac{2\sqrt{n}}{K_1}$  then implies that if  $\|(A_{(1,2,3)})^{-1}\| > K_1/\varepsilon$ , then

$$\mathbb{P}_{X,Y} \left\{ \inf_{x \in S_3^{[3:n]}} \|A_{(1,2)}x\|_2 \leq \frac{\varepsilon}{K_1} \right\} \leq \frac{Cn}{K_1}.$$

Therefore, unconditionally,

$$\mathbb{P} \left\{ \inf_{x \in S_3^{[3:n]}} \|A_{(1,2)}x\|_2 \leq \frac{\varepsilon}{K_1} \wedge \|(A_{(1,2,3)})^{-1}\| > \frac{K_1}{\varepsilon} \right\} \leq \frac{Cn}{K_1}.$$

By an identical argument, the dimension 3 here can be replaced by any other dimension  $i \in [3 : n]$ . Using a union bound over these  $i$  and (4.37), we conclude that

$$\mathbb{P} \left\{ \inf_{x \in S_3^{[3:n]}} \|A_{(1,2)}x\|_2 \leq \frac{\varepsilon}{K_1} \wedge \|(A_{(1,2,i)})^{-1}\| > \frac{K_1}{\varepsilon} \forall i \in [3 : n] \right\} \leq \frac{Cn^2}{K_1}.$$

This is of course the same as

$$(4.38) \quad \mathbb{P} \left\{ \|(A_{(1,2)})^{-1}\| > \frac{K_1}{\varepsilon} \wedge \|(A_{(1,2,i)})^{-1}\| > \frac{K_1}{\varepsilon} \forall i \in [3 : n] \right\} \leq \frac{Cn^2}{K_1}.$$

This concludes the first step: we have shown that in the situation of  $p_{1,0}$  when all minors  $A_{(1,2,i)}$  are poorly invertible, the minor  $A_{(1,2)}$  is well invertible.

4.4.3. *Going two more dimensions up.* At the second step, we move from the good invertibility of the minor  $A_{(1,2)}$  that we have just established to a good invertibility of the full matrix  $A$ . But we have already addressed exactly this problem in Section 4.3, except for the minor  $A_{(1,2,3)}$ . So no new argument will be needed in this case.

Formally, combining (4.36), (4.38), and the estimate (4.3) on  $\mathcal{E}_S$ , we obtain

$$p_{1,0} \leq \mathbb{P} \left\{ \inf_{x \in S_{1,2}} \|Ax\|_2 \leq 2\mu\varepsilon \wedge \|(A_{(1,2)})^{-1}\| \leq \frac{K_1}{\varepsilon} \wedge \mathcal{E}_S \right\} + \frac{Cn^2}{K_1} + 2 \exp(-cK_0^2n).$$

The probability here is very similar to the probability  $p_{1,3}$  in (4.11) and is bounded in the same way as in (4.34); see Remark 4.7. We conclude that

$$(4.39) \quad p_{1,0} \leq q + \frac{Cn^2}{K_1} + 2 \exp(-cK_0^2n)$$

where  $\mu$  and  $q$  are defined in (4.34) and (4.33), respectively.

We have successfully estimated  $p_{1,0}$  in the sum (4.10). This achieves the goal of this section, which was to show that  $A$  is well invertible on the set  $S_{1,2}$  in the case when there all minors  $A_{(1,2,i)}$  are poorly invertible.

**4.5. Combining the results for well-invertible and poorly invertible minors.** At this final stage of the proof, we combine the conclusions of Sections 4.3 and 4.4.

Recall from (4.10) that

$$p_1 \leq \sum_{i=3}^n p_{1,i} + p_{1,0}.$$

The terms in this sum were estimated in (4.35) and in (4.39). Combining these, we obtain

$$p_1 \leq nq + \frac{Cn^2}{K_1} + 2 \exp(-cK_0^2n).$$

An identical argument produces the same estimate for all  $p_i, i = 2, \dots, n$ , in (4.8). Thus

$$\begin{aligned}
 p &= \mathbb{P} \left\{ s_{\min}(D + U) \leq \mu\varepsilon \right\} \leq \sum_{i=1}^n p_i + 2 \exp(-cK_0^2 n) \\
 (4.40) \quad &\leq n^2 q + \frac{Cn^3}{K_1} + 2(n + 1) \exp(-cK_0^2 n).
 \end{aligned}$$

Recall that  $\mu$  and  $q$  are defined in (4.34) and (4.33), respectively, and  $C \geq 1, c \leq 1$  in these inequalities.

Finally, for  $t \in (0, 1)$ , we choose the parameters  $K_0 > 1, K_1 > 1, \varepsilon, \lambda_1, \lambda_2 \in (0, 1)$  to make the expression in (4.40) reasonably small. For example, one can choose

$$\begin{aligned}
 K_0 &= \log(1/t), & K_1 &= t^{-1/16}, \\
 \lambda_1 &= t^{1/16}, & \lambda_2 &= t^{1/32}, & \varepsilon &= \frac{t^{9/8}}{24K \log^2(1/t)n^{3/2}}.
 \end{aligned}$$

With this choice, we have

$$\begin{aligned}
 \mu &\geq \frac{t^{9/8}}{6K\sqrt{n}} \geq 2K_0^2 n\varepsilon, & q &\leq 3t^{c/32}(K^2/\delta)^c, \\
 p &\leq C_1 n^3 t^{c/32}(K^2/\delta)^c,
 \end{aligned}$$

and so (4.2) and (4.5) are satisfied, and (4.4) is satisfied whenever  $t < e^{-C_0}$ . Summarizing, we have shown that

$$\mathbb{P} \left\{ s_{\min}(D + U) \leq \frac{t^{9/4}}{144K^2 \log^4(1/t)n^2} \right\} \leq C_1 n^3 t^{c/32}(K^2/\delta)^c.$$

This quickly leads to the conclusion of Theorem 1.3. □

5. APPLICATION TO THE SINGLE RING THEOREM: PROOF OF COROLLARY 1.4

In this section we prove Corollary 1.4, which states that condition (SR3) can be completely eliminated from the Single Ring Theorem. Let  $D_n$  be a sequence of deterministic  $n \times n$  diagonal matrices. (The case of random  $D_n$  can be reduced to this by conditioning on  $D_n$ .) If  $z \neq 0$ , then

$$(5.1) \quad s_{\min}(U_n D_n V_n - zI_n) = |z| \cdot s_{\min}((1/z)D_n - U_n^{-1}V_n^{-1}),$$

where the matrix  $U_n^{-1}V_n^{-1}$  is uniformly distributed in  $U(n)$  or  $O(n)$ . Let us first consider the case where the matrices  $D_n$  are well invertible, thus we assume that

$$r := \inf_{n \in \mathbb{N}} s_{\min}(D_n) > 0.$$

In the complex case, an application of Theorem 1.1 yields the inequality

$$(5.2) \quad \mathbb{P} \left\{ s_{\min}(U_n D_n V_n - zI_n) \leq tr \right\} \leq t^c n^C, \quad 0 \leq t < 1/2,$$

which holds (uniformly) for all  $z \in \mathbb{C}$  and which implies condition (SR3). Indeed, Theorem 1.1 combined with (5.1) implies the inequality (5.2) for  $|z| \geq r/2$ . In the disc  $|z| < r/2$  we use the trivial estimate

$$s_{\min}(U_n D_n V_n - zI_n) \geq s_{\min}(U_n D_n V_n) - |z| > r/2,$$

which again implies (5.2).

Now consider the real case, still under the assumption that  $r > 0$ . Condition (SR1) allows us to assume that  $\|D_n\| \leq K$  for some  $K$  and for all  $n$ . Condition (SR2) and [10, Lemma 15] imply that  $|s_k(D_n) - 1| \geq 1/(4\kappa_1)$  for some  $1 \leq k \leq n$ . Hence

$$\inf_{V \in O(n)} \|D_n - V\| \geq \frac{1}{4\kappa_1}.$$

An application of Theorem 1.3 together with (5.1) shows that inequality (5.2) holds, which in turn implies condition (SR3). In this argument, we considered the matrix  $(1/z)D_n$ , which has *complex* entries. This was the reason for proving the more general Theorem 1.3 instead of the simpler Theorem 1.2.

It remains to analyze the case where the matrices  $D_n$  are poorly invertible, i.e. when  $\inf_{n \in \mathbb{N}} s_{\min}(D_n) = 0$ . In this case the condition (SR3) can be removed from the Single Ring Theorem using our results via the following argument, which was communicated to the authors by Ofer Zeitouni [24]. The proof of the Single Ring Theorem in [10] uses condition (SR3) only once, specifically in the proof of [10, Proposition 14] which is one of the main steps in the argument. Let us quote this proposition.

**Proposition 14** ([10]). *Let  $\nu_z^{(n)}$  be the symmetrized<sup>3</sup> empirical measure of the singular values of  $U_n D_n V_n - zI_n$ . Assume that the conditions (SR1), (SR2), and (SR3) of the Single Ring Theorem hold.*

- (i) *There exists a sequence of events  $\Omega_n$  with  $\mathbb{P}(\Omega_n) \rightarrow 1$  such that for Lebesgue almost every  $z \in \mathbb{C}$ , one has*

$$(5.3) \quad \lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{E} \int_0^\varepsilon \mathbf{1}_{\Omega_n} \log |x| d\nu_z^{(n)}(x) = 0.$$

*Consequently, for almost every  $z \in \mathbb{C}$  one has*

$$(5.4) \quad \int_{\mathbb{R}} \log |x| d\nu_z^{(n)}(x) \rightarrow \int_{\mathbb{R}} \log |x| d\nu_z(x)$$

*for some limit measure  $\nu_z$  in probability.*

- (ii) *For any  $R > 0$  and for any smooth deterministic function  $\varphi$  compactly supported in  $B_R = \{z \in \mathbb{C} : |z| \leq R\}$ , one has*

$$(5.5) \quad \int_{\mathbb{C}} \varphi(z) \int_{\mathbb{R}} \log |x| d\nu_z^{(n)}(x) dm(z) \rightarrow \int_{\mathbb{C}} \varphi(z) \int_{\mathbb{R}} \log |x| d\nu_z(x) dm(z).$$

Our task is to remove condition (SR3) from this proposition. Since the argument below is the same for unitary and orthogonal matrices, we will not distinguish between the real and the complex case.

Even without assuming (SR3), part (i) can be deduced from Theorems 1.1 and 1.2 by the argument of [10], since condition (5.3) pertains to a fixed  $z$ .

It remains to prove (ii) without condition (SR3). To this end, consider the probability measure  $\tilde{\mu}$  with the density

$$(5.6) \quad \frac{d\tilde{\mu}}{dm}(z) = \frac{1}{2\pi} \Delta \left( \int_{\mathbb{R}} \log |x| d\nu_z(x) \right).$$

This measure was introduced and studied in [10]. After the Single Ring Theorem is proved, it turns out that  $\tilde{\mu} = \mu_e$ , where  $\mu_e$  is the limit of the empirical measures

---

<sup>3</sup>Symmetrization here means that we consider the set of the singular values  $s_k$  together with their opposites  $-s_k$ .



of eigenvalues. However, at this point of the proof this identity is not established, so we have to distinguish between these two measures.

It was shown in [10] that for any smooth compactly supported function  $f : \mathbb{C} \rightarrow \mathbb{C}$  such that condition (SR3) holds with some  $\delta, \delta' > 0$  for almost all  $z \in \text{supp}(f)$ , one has

$$(5.7) \quad \int_{\mathbb{C}} f(z) d\mu_e^{(n)}(z) \rightarrow \int_{\mathbb{C}} f(z) d\tilde{\mu}(z).$$

The argument in the beginning of this section shows that if  $Q := \text{supp}(f) \subset B_R \setminus B_r$  for some  $r > 0$ , then (5.2) holds uniformly on  $Q$ , and therefore (5.7) holds for such  $f$ .

The proof of [10, Theorem 1] shows that it is enough to establish (ii) for all smooth compactly supported functions  $\varphi$  that can be represented as  $\varphi = \Delta\psi$ , where  $\psi$  is another smooth compactly supported function. Assume that (ii) fails; thus there exist  $\varepsilon > 0$ , a subsequence  $\{n_k\}_{k=1}^\infty$ , and a function  $\psi : \mathbb{C} \rightarrow \mathbb{C}$  as above, such that

$$(5.8) \quad \left| \int_{\mathbb{C}} \Delta\psi(z) \int_{\mathbb{R}} \log|x| d\nu_z^{(n_k)}(x) dm(z) - \int_{\mathbb{C}} \Delta\psi(z) \int_{\mathbb{R}} \log|x| d\nu_z(x) dm(z) \right| > \varepsilon.$$

Recall the following identity [10, formula (5)]:

$$(5.9) \quad \int_{\mathbb{C}} \psi(z) d\mu_e^{(n)}(z) = \frac{1}{2\pi} \int_{\mathbb{C}} \Delta\psi(z) \int_{\mathbb{R}} \log|z| d\nu_z^{(n)}(x) dm(z).$$

Condition (SR1) implies that the sequence of measures  $\mu_e^{(n_k)}$  is tight, so we can extract a further subsequence  $\{\mu_e^{(n_{k_l})}\}_{l=1}^\infty$  which converges weakly to a probability measure  $\mu$ .

We claim that  $\mu = \tilde{\mu}$ . Indeed, let  $f : \mathbb{C} \rightarrow [0, 1]$  be a smooth function supported in  $B_R \setminus B_r$  for some  $r > 0$ . Then the weak convergence implies

$$\int_{\mathbb{C}} f(z) d\mu_e^{(n_{k_l})}(z) \rightarrow \int_{\mathbb{C}} f(z) d\mu(z).$$

Since  $f$  satisfies (5.7), we obtain

$$\int_{\mathbb{C}} f(z) d\mu(z) = \int_{\mathbb{C}} f(z) d\tilde{\mu}(z).$$

This means that the measure  $\mu$  coincides with  $\tilde{\mu}$  on  $\mathbb{C} \setminus \{0\}$ . Since both  $\mu$  and  $\tilde{\mu}$  are probability measures,  $\mu = \tilde{\mu}$ .

Since  $\tilde{\mu}$  is absolutely continuous, we can choose  $\tau > 0$  so that  $\tilde{\mu}(B_\tau) < \frac{\varepsilon}{8\pi\|\psi\|_\infty}$ . Let  $\eta : \mathbb{C} \rightarrow [0, 1]$  be a smooth function such that  $\text{supp}(\eta) \subset B_\tau$  and  $\eta(z) = 1$  for any  $z \in B_{\tau/2}$ . Then

$$(5.10) \quad \int_{\mathbb{C}} \eta(z) d\tilde{\mu}(z) < \frac{\varepsilon}{8\pi\|\psi\|_\infty},$$

and therefore

$$(5.11) \quad \int_{\mathbb{C}} \eta(z) d\mu_e^{(n_{k_l})}(z) < \frac{\varepsilon}{8\pi\|\psi\|_\infty}$$

for all sufficiently large  $l$ . Let us estimate the quantity in (5.8):

$$\begin{aligned} & \left| \int_{\mathbb{C}} \Delta\psi(z) \int_{\mathbb{R}} \log|x| d\nu_z^{(n_{k_l})}(x) dm(z) - \int_{\mathbb{C}} \Delta\psi(z) \int_{\mathbb{R}} \log|x| d\nu_z(x) dm(z) \right| \\ & \leq \left| \int_{\mathbb{C}} \Delta((1-\eta)\psi)(z) \int_{\mathbb{R}} \log|x| d\nu_z^{(n_{k_l})}(x) dm(z) \right. \\ & \qquad \qquad \qquad \left. - \int_{\mathbb{C}} \Delta((1-\eta)\psi)(z) \int_{\mathbb{R}} \log|x| d\nu_z(x) dm(z) \right| \\ & + \left| \int_{\mathbb{C}} \Delta(\eta\psi)(z) \int_{\mathbb{R}} \log|x| d\nu_z^{(n_{k_l})}(x) dm(z) \right| \\ & + \left| \int_{\mathbb{C}} \Delta(\eta\psi)(z) \int_{\mathbb{R}} \log|x| d\nu_z(x) dm(z) \right|. \end{aligned}$$

Consider the terms on the right-hand side separately. By (5.6) and (5.10), we have

$$\begin{aligned} \left| \int_{\mathbb{C}} \Delta(\eta\psi)(z) \int_{\mathbb{R}} \log|x| d\nu_z(x) dm(z) \right| &= 2\pi \left| \int_{\mathbb{C}} (\eta\psi)(z) d\tilde{\mu}(z) \right| \\ &\leq 2\pi \|\psi\|_{\infty} \cdot \left| \int_{\mathbb{C}} \eta(z) d\tilde{\mu}(z) \right| < \frac{\varepsilon}{4}. \end{aligned}$$

Similarly, (5.9) and (5.11) imply that for large  $l$

$$\left| \int_{\mathbb{C}} \Delta(\eta\psi)(z) \int_{\mathbb{R}} \log|x| d\nu_z^{(n_{k_l})}(x) dm(z) \right| < \frac{\varepsilon}{4}.$$

The function  $\tilde{\varphi} = \Delta((1-\eta)\psi)$  is supported in the annulus  $B_R \setminus B_{\tau/2}$ . This function satisfies (5.7), so using (5.6) and (5.9), we obtain

$$\begin{aligned} & \int_{\mathbb{C}} \tilde{\varphi}(z) \int_{\mathbb{R}} \log|x| d\nu_z^{(n)}(x) dm(z) - \int_{\mathbb{C}} \tilde{\varphi}(z) \int_{\mathbb{R}} \log|x| d\nu_z(x) dm(z) \\ & = 2\pi \int_{\mathbb{C}} ((1-\eta)\psi)(z) d\mu_e^{(n_{k_l})}(z) - 2\pi \int_{\mathbb{C}} ((1-\eta)\psi)(z) d\mu(z) \rightarrow 0. \end{aligned}$$

The combination of these inequalities yields

$$\begin{aligned} & \limsup_{l \rightarrow \infty} \left| \int_{\mathbb{C}} \Delta\psi(z) \int_{\mathbb{R}} \log|x| d\nu_z^{(n_{k_l})}(x) dm(z) - \int_{\mathbb{C}} \Delta\psi(z) \int_{\mathbb{R}} \log|x| d\nu_z(x) dm(z) \right| \\ & < \frac{\varepsilon}{2}, \end{aligned}$$

which contradicts (5.8). □

*Remark 5.1.* Convergence of the empirical measures of eigenvalues  $\mu_e^{(n)}$  to the limit measure  $\mu_e$  does not imply the convergence of the eigenvalues to an annulus. Indeed, there may be outliers which do not affect the limit measure. For example, assume that  $\{D_n\}_{n=1}^{\infty}$  is a sequence of diagonal matrices

$$D_n = \text{diag}(d_1, \dots, d_{n-1}, 0),$$

where  $d_1, \dots, d_{n-1}$  are independent random variables uniformly distributed in  $[1, 2]$ . Then the Single Ring Theorem asserts that the support of the measure  $\mu_e$  is the annulus  $\sqrt{2} \leq |z| \leq \sqrt{7/3}$ ; see (1.4). At the same time, all matrices  $A_n = U_n D_n V_n$  have eigenvalue 0.

Guionnet and Zeitouni [11] established sufficient conditions for the convergence of the spectrum of  $A_n$  to an annulus. Assume that the matrices  $A_n$  satisfy (SR1), (SR2), and (SR3), and in addition:

(SR4) Assume that

$$\begin{aligned} \left(\int_0^\infty x^{-2} d\mu_s^{(n)}(x)\right)^{-1/2} &\rightarrow a = \left(\int_0^\infty x^{-2} d\mu_s(x)\right)^{-1/2}, \\ \left(\int_0^\infty x^2 d\mu_s^{(n)}(x)\right)^{1/2} &\rightarrow b = \left(\int_0^\infty x^2 d\mu_s(x)\right)^{1/2}, \end{aligned}$$

and  $\inf_n s_{\min}(D_n) > 0$  whenever  $a > 0$ .

Then [11, Theorem 2] claims that the spectrum of  $A_n$  converges to the annulus  $a \leq |z| \leq b$  in probability.

Arguing as before, one can eliminate the condition (SR3) from this list. The other conditions are formulated in terms of the matrices  $D_n$  only.

### APPENDIX A. ORTHOGONAL PERTURBATIONS IN LOW DIMENSIONS

In this section we prove Theorem 4.1, which is a slightly stronger version of the main Theorem 1.3 in dimensions  $n = 2$  and  $n = 3$ . The argument will be based on Remez-type inequalities.

**A.1. Remez-type inequalities.** Remez inequality and its variants capture the following phenomenon: if a polynomial of a fixed degree is small on a set of given measure, then it remains to be small on a larger set (usually an interval). We refer to [7], [8] for an extensive discussion of these inequalities.

We will use two versions of Remez-type inequalities, for multivariate polynomials on a convex body and on the sphere. The first result is due to Ganzburg and Brudnyi [1], [2]; see [7, Section 4.1].

**Theorem A.1** (Remez-type inequality on a convex body). *Let  $V \subset \mathbb{R}^m$  be a convex body, let  $E \subseteq V$  be a measurable set, and let  $f$  be a real polynomial on  $\mathbb{R}^m$  of degree  $n$ . Then*

$$\sup_{x \in V} |f(x)| \leq \left(\frac{4m|V|}{|E|}\right)^n \sup_{x \in E} |f(x)|.$$

Here  $|E|$  and  $|V|$  denote the  $m$ -dimensional Lebesgue measures of these sets.

The second result can be found in [7]; see (3.3) and Theorem 4.2 there.

**Theorem A.2** (Remez-type inequality on the sphere). *Let  $m \in \{1, 2\}$ , let  $E \subseteq S^m$  be a measurable set, and let  $f$  be a real polynomial on  $\mathbb{R}^{m+1}$  of degree  $n$ . Then*

$$\sup_{x \in S^m} |f(x)| \leq \left(\frac{C_1}{|E|}\right)^{2n} \sup_{x \in E} |f(x)|.$$

Here  $|E|$  denotes the  $m$ -dimensional Lebesgue measure of  $E$ .

*Remark A.3.* By a simple argument based on Fubini’s theorem, a similar Remez-type inequality can be proved for the real three-dimensional torus  $T_3 := S^1 \times S^2 \subset \mathbb{R}^5$  equipped with the product measure:

$$(A.1) \quad \sup_{x \in T_3} |f(x)| \leq \left(\frac{C_1}{|E|}\right)^{4n} \sup_{x \in E} |f(x)|.$$

**A.2. Vanishing determinant.** Before we can prove Theorem 4.1, we establish a simpler result, which is deterministic and which concerns the determinant instead of the smallest singular value. The determinant is simpler to handle because it can be easily expressed in terms of the matrix entries.

**Lemma A.4** (Vanishing determinant). *Let  $B$  be a fixed  $n \times n$  complex matrix, where  $n \in \{2, 3\}$ . Assume that  $\|B\| \geq 1/2$ . Let  $\varepsilon > 0$  and assume that*

$$|\det(B + U)| \leq \varepsilon \quad \text{for all } U \in SO(n).$$

Then

$$\|BB^T - I\| \leq C\varepsilon\|B\|.$$

*Proof.* To make this proof more readable, we will write  $a \lesssim b$  if  $a \leq Cb$  for a suitable absolute constant  $C$ , and we will write  $a \approx_\varepsilon b$  if  $|a - b| \lesssim \varepsilon$ .

**Dimension  $n = 2$ .** Let us represent

$$U = U(\phi) = \begin{bmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{bmatrix}.$$

Then  $\det(B + U)$  is a trigonometric polynomial

$$\det(B + U) = k_0 + k_1 \cos \phi + k_2 \sin \phi$$

whose coefficients can be expressed in terms of the coefficients of  $B$ :

$$k_0 = \det(B) + 1; \quad k_1 = B_{11} + B_{22}; \quad k_2 = B_{12} - B_{21}.$$

By assumption, the modulus of this trigonometric polynomial is bounded by  $\varepsilon$ . Therefore all of its coefficients are also bounded; i.e.

$$|k_i| \lesssim \varepsilon, \quad i = 1, 2, 3.$$

It is enough to check that all entries of  $BB^T$  are close to the corresponding entries of  $I$ . We will check this for entries  $(1, 1)$  and  $(1, 2)$ ; others are similar. Then

$$(BB^T)_{11} = B_{11}^2 + B_{12}^2 \approx_{\varepsilon'} -B_{11}B_{22} + B_{12}B_{21}$$

where we used that  $|k_1| \lesssim \varepsilon$ ,  $|k_2| \lesssim \varepsilon$ , and thus the resulting error can be estimated as

$$\varepsilon' \lesssim \varepsilon(|B_{11}| + |B_{12}|) \lesssim \varepsilon\|B\|.$$

But

$$-B_{11}B_{22} + B_{12}B_{21} = -\det(B) \approx_\varepsilon 1,$$

where we used that  $|k_0| \lesssim \varepsilon$ . We have shown that

$$|(BB^T)_{11} - 1| \lesssim \varepsilon\|B\| + \varepsilon \lesssim \varepsilon\|B\|,$$

as required.

Similarly we can estimate

$$(BB^T)_{12} = B_{11}B_{21} + B_{12}B_{22} \approx_{\varepsilon'} B_{11}B_{12} - B_{12}B_{11} = 0.$$

Repeating this procedure for all entries, we have shown that

$$|(BB^T)_{ij} - I_{ij}| \lesssim \varepsilon\|B\|$$

for all  $i, j$ . This immediately implies the conclusion of the lemma in dimension  $n = 2$ .

**Dimension  $n = 3$ .** We claim that

$$(A.2) \quad \det(B) \approx_\varepsilon -1; \quad B_{ij} \approx_\varepsilon (-1)^{i+j+1} \det(B^{ij}), \quad i, j \in \{1, 2, 3\},$$

where  $B^{ij}$  denotes the minor obtained by removing the  $i$ -th row and  $j$ -th column from  $B$ .

Let us prove (A.2) for  $i = j = 1$ ; for other entries the argument is similar. Let

$$U = U(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & \sin \phi \\ 0 & -\sin \phi & \cos \phi \end{bmatrix}.$$

Then as before,  $\det(B + U)$  is a trigonometric polynomial

$$\det(B + U) = k_0 + k_1 \cos \phi + k_2 \sin \phi$$

whose coefficients can be expressed in terms of the coefficients of  $B$ . Our argument will only be based on the free coefficient  $k_0$ , which one can quickly show to equal

$$k_0 = \det \begin{bmatrix} B_{11} + 1 & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} + B_{11} + 1 = \det(B) + \det(B^{11}) + B_{11} + 1.$$

As before, the assumption yields that  $|k_0| \lesssim \varepsilon$ , so

$$(A.3) \quad \det(B) + \det(B^{11}) + B_{11} + 1 \approx_\varepsilon 0.$$

Repeating the same argument for

$$U = U(\phi) = \begin{bmatrix} -1 & 0 & 0 \\ 0 & \cos \phi & \sin \phi \\ 0 & \sin \phi & -\cos \phi \end{bmatrix}$$

yields

$$(A.4) \quad \det(B) - \det(B^{11}) - B_{11} + 1 \approx_\varepsilon 0.$$

Estimates (A.3) and (A.4) together imply that

$$\det(B) \approx_\varepsilon -1; \quad B_{11} \approx_\varepsilon -\det(B^{11}).$$

This implies claim (A.2) for  $i = j = 1$ ; for other entries the argument is similar.

Now we can estimate the entries of  $B^\top B$ . Indeed, by (A.2) we have

$$(A.5) \quad \begin{aligned} (BB^\top)_{11} &= B_{11}^2 + B_{12}^2 + B_{13}^2 \\ &\approx_{\varepsilon'} -B_{11} \det(B^{11}) + B_{12} \det(B^{12}) - B_{13} \det(B^{13}), \end{aligned}$$

where the error  $\varepsilon'$  can be estimated as

$$\varepsilon' \lesssim \varepsilon(|B_{11}| + |B_{12}| + |B_{13}|) \lesssim \varepsilon \|B\|.$$

Further, the expression in (A.5) equals  $-\det(B)$ , which can be seen by expanding the determinant along the first row. Finally,  $-\det(B) \approx_\varepsilon 1$  by (A.2). We have shown that

$$|(B^\top B)_{11} - 1| \lesssim \varepsilon \|B\| + \varepsilon \lesssim \varepsilon \|B\|,$$

as required.

Similarly we can estimate

$$\begin{aligned} (BB^\top)_{12} &= B_{11}B_{21} + B_{12}B_{22} + B_{13}B_{23} \\ &\approx_{\varepsilon'} B_{11} \det(B^{21}) - B_{12} \det(B^{22}) + B_{13} \det(B^{23}) \\ &= B_{11}(B_{12}B_{33} - B_{32}B_{13}) - B_{12}(B_{11}B_{33} - B_{31}B_{13}) + B_{13}(B_{11}B_{32} - B_{31}B_{12}) \\ &= 0 \end{aligned}$$

(all terms cancel).

Repeating this procedure for all entries, we have shown that

$$|(BB^T)_{ij} - I_{ij}| \lesssim \varepsilon \|B\|$$

for all  $i, j$ . This immediately implies the conclusion of the lemma in dimension  $n = 3$ . □

**A.3. Proof of Theorem 4.1.** Let us fix  $t$ ; without loss of generality, we can assume that  $t < \delta/100$ . Let us assume that  $B + U$  is poorly invertible with significant probability:

$$(A.6) \quad \mathbb{P} \{s_{\min}(B + U) \leq t\} > p(\delta, t)$$

where  $p(\delta, t) \in (0, 1)$  is to be chosen later. Without loss of generality we may assume that  $U$  is distributed uniformly in  $SO(n)$  rather than  $O(n)$ . Indeed, since  $O(n)$  can be decomposed into two conjugacy classes  $SO(n)$  and  $O(n) \setminus SO(n)$ , the inequality (A.6) must hold over at least one of these classes. Multiplying one of the rows of  $B + U$  by  $-1$  if necessary, one can assume that it holds for  $SO(n)$ .

Note that  $\|B\| \geq 1/2$ ; otherwise  $s_{\min}(B + U) \geq 1 - \|B\| \geq 1/2 > t$  for all  $U \in O(n)$ , which violates (A.6).

**A.3.1. Dimension  $n = 2$ .** In this case the result follows easily from Lemma A.4 and Remez inequality. Indeed, the event  $s_{\min}(B + U) \leq t$  implies

$$|\det(B + U)| = s_{\min}(B + U)\|B + U\| \leq t(\|B\| + 1) \leq 3t\|B\|.$$

Therefore, by (A.6) we have

$$(A.7) \quad \mathbb{P} \{\det(B + U) \leq 3t\|B\|\} > p(\delta, t).$$

A random uniform rotation  $U = \begin{bmatrix} x & y \\ -y & x \end{bmatrix} \in SO(2)$  is determined by a random uniform point  $(x, y)$  on the real sphere  $S^1$ . Now,  $\det(D + U)$  is a complex-valued quadratic polynomial in variables  $x, y$  that is restricted to the real sphere  $S^1$ . Hence  $|\det(D + U)|^2$  is a real-valued polynomial of degree 4 restricted to the real sphere  $S^1$ . Therefore, we can apply the Remez-type inequality, Theorem A.2, for the subset  $E := \{U : |\det(D + U)|^2/\|B\|^2 \leq 3t\}$  of  $S^1$  which satisfies  $|E| \geq 2\pi p(\delta, t)$  according to (A.7). It follows that

$$|\det(B + U)| \leq \left(\frac{C_1}{p(\delta, t)}\right)^{C_0} t\|B\| \quad \text{for all } U \in SO(2).$$

An application of Lemma A.4 then gives

$$\|BB^T - I\| \leq C_2 \left(\frac{C_1}{p(\delta, t)}\right)^{C_0} t\|B\|^2.$$

On the other hand, assumption (4.1) states that the left-hand side is bounded below by  $\delta\|B\|^2$ . It follows that

$$(A.8) \quad \delta \leq C_2 \left(\frac{C_1}{p(\delta, t)}\right)^{C_0} t.$$

Now we can choose  $p(\delta, t) = C(t/\delta)^c$  with sufficiently large absolute constant  $C$  and sufficiently small absolute constant  $c > 0$  so that inequality (A.8) is violated. Therefore (A.6) fails with this choice of  $p(\delta, t)$ , and consequently we have

$$\mathbb{P} \{s_{\min}(B + U) \leq t\} \leq C(t/\delta)^c,$$

as claimed.

A.3.2. *Dimension  $n = 3$ : middle singular value.* This time, the determinant is the product of three singular values. So repeating the previous argument would produce an extra factor of  $\|B\|$ , which would force us to require that

$$\|BB^T - I\| \geq \delta \|B\|^3$$

instead of (4.1).

The weak point of this argument is that it ignores the *middle* singular value of  $B$ , replacing it by the largest one. We will now be more careful. Let  $s_1 \geq s_2 \geq s_3 \geq 0$  denote the singular values of  $B$ .

Assume the event  $s_{\min}(B+U) \leq t$  holds. Since  $\|U\| = 1$ , the triangle inequality, Weyl's inequality, and the assumption imply that the three singular values of  $B+U$  are bounded, one by  $s_1 + 1 \leq \|B\| + 1 \leq 3\|B\|$ , another by  $s_2 + 1$ , and the remaining one by  $t$ . Thus

$$|\det(B+U)| \leq 3t(s_2 + 1)\|B\|.$$

Let  $K \geq 2$  be a parameter to be chosen later. Suppose first that  $s_2 \leq K$  holds. Then  $|\det(B+U)| \leq 6tK\|B\|$ , and we shall apply Remez inequality. In order to do this, we can realize  $U \in SO(3)$  as a random uniform rotation of the  $(x, y)$ -plane followed by an independent rotation that maps the  $z$ -axis to a uniform random direction. Thus  $U$  is determined by a random point  $(x, y, z_1, z_2, z_3)$  in the real three-dimensional torus  $T_3 = S^1 \times S^2$ , chosen according to the uniform (product) distribution. Here  $(x, y) \in S^1$  and  $(z_1, z_2, z_3) \in S^2$  determine the two rotations we described above.<sup>4</sup>

We regard  $|\det(B+U)|^2$  as a real polynomial in five variables  $x, y, z_1, z_2, z_3$  and constant degree, which is restricted to  $T_3$ . Thus we can apply the Remez-type inequality for the torus, (A.1), and an argument similar to the case  $n = 2$  yields

$$(A.9) \quad \mathbb{P}\{s_{\min}(B+U) \leq t\} \leq C(tK/\delta)^c.$$

Now we assume that  $s_2 \geq K$ . We will show that, for an appropriately chosen  $K$ , this case is impossible; i.e.  $B+U$  cannot be poorly invertible with considerable probability.

A.3.3. *Reducing to one dimension.* Since  $s_1 \geq s_2 \geq K \geq 2$ , it must be that  $s_3 \leq 2$ ; otherwise all singular values of  $B$  are bounded below by 2, which clearly implies that  $s_{\min}(B+U) \geq 1$  for all  $U \in O(3)$ . This will allow us to reduce our problem to one dimension. To this end, we consider the singular value decomposition of  $B$ ,

$$B = s_1 q_1 p_1^* + s_2 q_2 p_2^* + s_3 q_3 p_3^*,$$

where  $\{p_1, p_2, p_3\}$  and  $\{q_1, q_2, q_3\}$  are orthonormal bases in  $\mathbb{C}^3$ .

Assume the event  $s_{\min}(B+U) \leq t$  holds. Then there exists  $x \in \mathbb{C}^3$ ,  $\|x\|_2 = 1$ , such that

$$\|(B+U)x\|_2 \leq t.$$

We are going to show that  $x$  is close to  $p_3$ , up to a unit scalar factor. To see this, note that  $\|Bx\|_2 \leq 1 + t \leq 2$ , so

$$(A.10) \quad \begin{aligned} 4 &\geq \|Bx\|_2^2 = s_1^2 |p_1^* x|^2 + s_2^2 |p_2^* x|^2 + s_3^2 |p_3^* x|^2 \geq K^2 (|p_1^* x|^2 + |p_2^* x|^2) \\ &= K^2 (1 - |p_3^* x|^2). \end{aligned}$$

<sup>4</sup>This construction and its higher-dimensional generalization follow the 1897 description of the Haar measure on  $SO(n)$  by Hurwitz; see [3].

It follows that

$$(A.11) \quad 1 - \frac{4}{K^2} \leq |p_3^*x| \leq 1$$

(the right-hand side holds since  $\|p_3^*\|_2 = \|x\|_2 = 1$ ). Let  $\eta := p_3^*x/|p_3^*x|$ ; then

$$\begin{aligned} \|x - \eta p_3\|_2^2 &= \|x/\eta - p_3\|_2^2 = |p_1^*(x/\eta - p_3)|^2 + |p_2^*(x/\eta - p_3)|^2 + |p_3^*(x/\eta - p_3)|^2 \\ &= |p_1^*x|^2 + |p_2^*x|^2 + \left| |p_3^*x| - 1 \right|^2 \quad (\text{by orthogonality and the definition of } \eta) \\ &\leq \frac{4}{K^2} + \frac{16}{K^4} \quad (\text{by (A.10) and (A.11)}) \\ &\leq \frac{8}{K^2}. \end{aligned}$$

Now, by the triangle inequality,

$$(A.12) \quad |q_3^*(B + U)p_3| = |q_3^*(B + U)\eta p_3| \leq |q_3^*(B + U)x| + |q_3^*(B + U)(x - \eta p_3)|.$$

The first term is bounded by  $\|q_3^*\|_2\|(B + U)x\|_2 \leq t$ . The second term is bounded by

$$\|q_3^*(B + U)\|_2\|x - \eta p_3\|_2 \leq (\|q_3^*B\|_2 + 1)\frac{\sqrt{8}}{K} = (s_3 + 1)\frac{\sqrt{8}}{K} \leq \frac{3\sqrt{8}}{K} \leq \frac{9}{K}.$$

Therefore the expression in (A.12) is bounded by  $t + 9/K$ .

Summarizing, we have found vectors  $u, v \in \mathbb{C}^3$ ,  $\|u\|_2 = \|v\|_2 = 1$ , such that the event  $s_{\min}(B + U) \leq t$  implies

$$|u^\top(B + U)v| \leq t + 9/K.$$

Note that the vectors  $u = (q_3^*)^\top$ ,  $v = p_3$  are fixed; they depend on  $B$  only. By (A.6), we have shown that

$$\mathbb{P} \{ |u^\top(B + U)v| \leq t + 9/K \} \geq p(\delta, t).$$

We can apply Remez inequality for  $|u^\top(B + U)v|^2$ , which is a quadratic polynomial in the entries of  $U$ . It yields

$$(A.13) \quad |u^\top(B + U)v| \leq \left( \frac{C_1}{p(\delta, t)} \right)^{C_0} (t + 9/K) \quad \text{for all } U \in SO(3).$$

Let  $c_0 \in (0, 1)$  be a small absolute constant. Now we can choose

$$(A.14) \quad p(\delta, t) = C(t/\delta)^c, \quad K = 4(\delta/t)^{1/2}$$

with sufficiently large absolute constant  $C$  and sufficiently small absolute constant  $c > 0$  so that the right-hand side in (A.13) is bounded by  $c_0$ . Summarizing, we have shown that

$$(A.15) \quad |u^\top(B + U)v| \leq c_0 \quad \text{for all } U \in SO(3).$$

We are going to show that this is impossible. In the remainder of the proof, we shall write  $a \ll 1$  to mean that  $a$  can be made arbitrarily small by a suitable choice of  $c_0$ , i.e. that  $a \leq f(c_0)$  for some fixed real-valued positive function (which does not depend on anything) and such that  $f(x) \rightarrow 0$  as  $x \rightarrow 0_+$ .



A.3.4. *Testing on various  $U$ .* Let us test (A.15) on  $U = U(\phi) = \begin{bmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{bmatrix}$ . Writing the bilinear form as a function of  $\phi$ , we obtain

$$u^\top(B + U)v = k + (u_1v_1 + u_2v_2) \cos \phi + (u_1v_2 - u_2v_1) \sin \phi$$

where  $k = k(B, u, v)$  does not depend on  $\phi$ . Since this trigonometric polynomial is small for all  $\phi$ , its coefficients must also be small; thus

$$|u_1v_1 + u_2v_2| \ll 1, \quad |u_1v_2 - u_2v_1| \ll 1.$$

We can write this in terms of a matrix-vector product as

$$\left\| \begin{bmatrix} u_1 & u_2 \\ -u_2 & u_1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right\|_2 \ll 1.$$

Since  $c_0$  is small, it follows that either the matrix  $\begin{bmatrix} u_1 & u_2 \\ -u_2 & u_1 \end{bmatrix}$  is poorly invertible (its smallest singular value is small) or the vector  $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$  has small norm. Since  $\|u\|_2 = 1$ , the norm of the matrix is bounded by  $\sqrt{2}$ . Hence the poor invertibility of the matrix is equivalent to the smallness of its determinant, which is  $u_1^2 + u_2^2$ . Formally, we conclude that

$$(A.16) \quad \text{either } |v_1|^2 + |v_2|^2 \ll 1 \quad \text{or } |u_1^2 + u_2^2| \ll 1.$$

Assume that  $|v_1|^2 + |v_2|^2 \ll 1$ ; since  $\|v\|_2 = 1$ , this implies  $|v_3| \geq 1/2$ . Now testing (A.15) on  $U = U(\phi) = \begin{bmatrix} \cos \phi & 0 & \sin \phi \\ 0 & 1 & 0 \\ -\sin \phi & 0 & \cos \phi \end{bmatrix}$ , a similar argument yields

$$|u_1v_1 + u_3v_3| \ll 1, \quad |u_1v_3 - u_3v_1| \ll 1.$$

Since  $|u_1| \leq 1$ ,  $|u_3| \leq 1$ ,  $|v_1| \ll 1$ , and  $|v_3| \geq 1/2$ , this system implies

$$|u_1| \ll 1, \quad |u_3| \ll 1.$$

Similarly, testing on  $U = U(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & \sin \phi \\ 0 & -\sin \phi & \cos \phi \end{bmatrix}$ , the same argument yields

$$|u_2| \ll 1, \quad |u_3| \ll 1.$$

So we proved that  $|u_1| \ll 1$ ,  $|u_2| \ll 1$ ,  $|u_3| \ll 1$ . But this is impossible since  $\|u\|_2 = 1$ .

We have thus shown that in (A.16) the first option never holds, so the second must hold. In other words, we have deduced from (A.15) that

$$(A.17) \quad |u_1^2 + u_2^2| \ll 1.$$

Using a similar argument (for rotations  $U$  in coordinates 1, 3 and 2, 3), we can also deduce that

$$(A.18) \quad |u_1^2 + u_3^2| \ll 1, \quad |u_2^2 + u_3^2| \ll 1.$$

Inequalities (A.17) and (A.18) imply that

$$|u_1^2| \ll 1, \quad |u_2^2| \ll 1, \quad |u_3^2| \ll 1.$$

But this contradicts the identity  $\|u\|_2 = 1$ .

This shows that (A.15) is impossible for a suitable choice of absolute constant  $c_0$ .

**A.3.5. Conclusion of the proof.** Let us recall the logic of the argument above. We assumed in (A.6) that  $B + U$  is poorly invertible with significant probability,  $\mathbb{P}\{s_{\min}(B + U) \leq t\} > p(\delta, t)$ . With the choice  $p(\delta, t) = C(t/\delta)^c$ ,  $K = 4(\delta/t)^{1/2}$  made in (A.14), we showed that either (A.9) holds (in the case  $s_2 \leq K$ ), i.e.  $\mathbb{P}\{s_{\min}(B + U) \leq t\} \leq C(tK/\delta)^c$ , or a contradiction appears (in the case  $s_2 \geq K$ ). Therefore, one always has

$$\mathbb{P}\{s_{\min}(B + U) \leq t\} \leq \max(p(\delta, t), C(tK/\delta)^c).$$

Due to our choice of  $p(\delta, t)$  and  $K$ , the right-hand side is bounded by  $C(tK/\delta)^{c/2}$ . This completes the proof of Theorem 4.1.  $\square$

APPENDIX B. SOME TOOLS USED IN THE PROOF OF THEOREM 1.3

In this appendix we shall prove auxiliary results used in the proof of Theorem 1.3. These include: Lemma B.2 on small ball probabilities for Gaussian random vectors (which we used in the proof of Lemma 4.8), Lemma 4.5 on invertibility of Gaussian perturbations, and Lemma 4.6 on breaking complex orthogonality by a random change of basis. Some of the proofs of these results follow standard arguments, but the statements are difficult to locate in the literature.

**B.1. Small ball probabilities.**

**Lemma B.1.** *Let  $X \sim N_{\mathbb{R}}(\mu, \sigma^2)$  for some  $\mu \in \mathbb{R}$ ,  $\sigma > 0$ . Then*

$$\mathbb{P}\{|X| \leq t\sigma\} \leq t, \quad t > 0.$$

*Proof.* The result follows since the density of  $X$  is bounded by  $1/\sigma\sqrt{2\pi}$ .  $\square$

**Lemma B.2.** *Let  $Z \sim N_{\mathbb{R}}(\mu, \sigma^2 I_n)$  for some  $\mu \in \mathbb{C}^n$  and  $\sigma > 0$ .<sup>5</sup> Then*

$$\mathbb{P}\{\|MZ\|_2 \leq t\sigma\|M\|_{\text{HS}}\} \leq Ct\sqrt{n}, \quad t > 0.$$

*Proof.* By rescaling we can assume that  $\sigma = 1$ .

First we give the argument in the real case, for  $\mu \in \mathbb{R}^n$ ,  $M \in \mathbb{R}^{n \times n}$ . Let  $M_i^T$  denote the  $i$ -th row of  $M$ , and let  $\mu = (\mu_1, \dots, \mu_n)$ . Choose  $i \in [n]$  such that  $\|M_i\|_2 \geq \|M\|_{\text{HS}}/\sqrt{n}$ . Note that  $M_i^T Z \sim N_{\mathbb{R}}(\nu_i, \|M_i\|_2^2)$  for some  $\nu_i \in \mathbb{R}^n$ . Lemma B.1 yields that

$$\mathbb{P}\{|M_i^T Z| \leq \tau\|M_i\|_2\} \leq Ct, \quad t > 0.$$

Since  $\|MZ\|_2 \geq |M_i^T Z|$  and  $\|M_i\|_2 \geq \|M\|_{\text{HS}}/\sqrt{n}$ , this quickly leads to the completion of the proof.

The complex case can be proved by decomposing  $\mu$  and  $M$  into real and imaginary parts and applying the real version of the lemma to each part separately.  $\square$

**B.2. Invertibility of random Gaussian perturbations.** In this appendix we prove Lemma 4.5.

First we note that without loss of generality, we can assume that  $m = 18$ . Indeed, since  $f$  is linear, it can be represented as

$$f(z) = [f(z)_{ij}]_{i,j=1}^3 = [a_{ij}^T z + \sqrt{-1} b_{ij}^T z]_{i,j=1}^3,$$

where  $a_{ij}$  and  $b_{ij}$  are some fixed vectors in  $\mathbb{R}^m$ . By the rotation invariance of  $Z$ , the joint distribution of the Gaussian random variables  $a_{ij}^T Z$  and  $b_{ij}^T Z$  is determined by the inner products of the vectors  $a_{ij}$  and  $b_{ij}$ . There are 18 of these vectors; so

---

<sup>5</sup>This means that  $X - \mu$  is a real-valued variable distributed according to  $N(0, \sigma^2 I_{n-1})$ .

we can isometrically realize them in  $\mathbb{R}^{18}$ . It follows that the distribution of  $f(Z)$  is preserved, and thus we can assume that  $m = 18$ .

Let  $R \geq 1$  be a parameter to be chosen later. By a standard Gaussian concentration inequality,  $\|Z\|_2 \leq R$  with probability at least  $1 - 2\exp(-cR^2)$ . On this event, the matrix in question is well bounded:  $\|I + f(Z)\| \leq 1 + \|f(Z)\|_{\text{HS}} \leq 2KR$ , and consequently we have

$$|\det(I + f(Z))| \leq s_{\min}(I + f(Z)) \cdot (2KR)^2.$$

Therefore we can estimate the probability in question as follows:

$$(B.1) \quad \mathbb{P} \{s_{\min}(I + f(Z)) \leq t\} \\ \leq \mathbb{P} \{|\det(I + f(Z))| \leq (2KR)^2 t, \|Z\| \leq R\} + 2\exp(-cR^2).$$

Since  $f$  is linear,  $|\det(I + f(Z))|^2$  is a real polynomial in  $Z \in \mathbb{R}^{18}$  of degree 6, and thus we can apply Remez inequality, Theorem A.1. We are interested in the Gaussian measure of the set

$$E := \{Z \in \mathbb{R}^{18} : |\det(I + f(Z))| \leq (2KR)^2 t, \|Z\| \leq R\}$$

which is a subset of

$$V := \{Z \in \mathbb{R}^{18} : \|Z\| \leq R\}.$$

The conclusion of Theorem A.1 is in terms of the Lebesgue rather than Gaussian measures of these sets:

$$|\det(I + f(Z))|^2 \leq \left(\frac{C_1|V|}{|E|}\right)^6 \cdot ((2KR)^2 t)^2 \quad \text{for all } Z \in V.$$

Taking square roots and substituting  $Z = 0$  in this inequality, we obtain

$$1 \leq \left(\frac{C_1|V|}{|E|}\right)^3 \cdot (2KR)^2 t;$$

thus

$$|E| \leq C_1|V| \cdot ((2KR)^2 t)^{1/3} \leq C_2 R^{18} \cdot ((2KR)^2 t)^{1/3},$$

where the last inequality follows from the definition of  $V$ . Further, note that the (standard) Gaussian measure of  $E$  is bounded by the Lebesgue measure  $|E|$ , because the density is bounded by  $(2\pi)^{-9} \leq 1$ . Recalling the definition of  $E$ , we have shown that

$$\mathbb{P} \{|\det(I + f(Z))| \leq (2KR)^2 t, \|Z\| \leq R\} \leq C_2 R^{18} \cdot ((2KR)^2 t)^{1/3}.$$

Substituting this back into (B.1), we obtain

$$\mathbb{P} \{|\det(I + f(Z))| \leq t\} \leq C_2 R^{18} \cdot ((2KR)^2 t)^{1/3} + 2\exp(-cR^2).$$

Finally, we can optimize the parameter  $R \geq 1$ , choosing for example  $R = t^{-1/1000}$  to conclude that

$$\mathbb{P} \{|\det(I + f(Z))| \leq t\} \leq C_3 K^{2/3} t^{1/4}.$$

This completes the proof of Lemma 4.5. □

**B.3. Breaking complex orthogonality.** In this section we prove Lemma 4.6 about breaking complex orthogonality by a random change of basis.

We will present the argument in dimension  $n = 3$ ; the dimension  $n = 2$  is very similar. Without loss of generality, we can assume that  $t < 1/2$ . Note that by assumption,

$$\|B\| \leq \|T\| \|D\| \leq K \|T\|.$$

Then the probability on the left side of (4.31) is bounded by

$$\mathbb{P} \{ \|BB^T - I\| \leq K^2 \|T\|^2 t \} = \mathbb{P} \left\{ \|\widehat{T}\widehat{T}^T - I\| \leq K^2 \|T\|^2 t \right\}, \quad \text{where } \widehat{T} = TQD.$$

We can pass to Hilbert-Schmidt norms (recall that all matrices are  $3 \times 3$  here) and further bound this probability by

$$\mathbb{P} \{ \|\widehat{T}\widehat{T}^T - I\|_{\text{HS}} \leq 3K^2 \|T\|_{\text{HS}}^2 t \}.$$

Assume that the conclusion of the lemma fails, so this probability is larger than  $C(tK^2/\delta)^c$ . We are going to apply Remez inequality and conclude that  $\|BB^T - I\|$  is small with probability one. Recalling the Hurwitz description of a uniform random rotation  $Q \in SO(3)$  which we used in Section A.3.2, we can parameterize  $Q$  by a uniform random point on the real torus  $T_3 = S^1 \times S^2 \subset \mathbb{R}^5$ . Under this parametrization,  $(\|\widehat{T}\widehat{T}^T - I\|_{\text{HS}}/3K^2 \|T\|_{\text{HS}}^2)^2$  becomes a polynomial in five variables and with constant degree restricted to  $T_3$ .

Our assumption above is that this polynomial is bounded by  $t^2$  on a subset of  $T_3$  of measure larger than  $C(tK^2/\delta)^c$ . Then the Remez-type inequality for the torus (A.1) implies that the polynomial is bounded on the entire  $T_3$  by

$$\left( \frac{C_1}{C(tK^2/\delta)^c} \right)^{C_0} t^2 \leq \left( \frac{\delta}{10^4 K^2} \right)^2$$

where the last inequality follows by a suitable choice of a large absolute constant  $C$  and a small absolute constant  $c$  in the statement of the lemma. This means that

$$\|\widehat{T}\widehat{T}^T - I\|_{\text{HS}} \leq 3K^2 \|T\|_{\text{HS}}^2 \cdot \frac{\delta}{10^4 K^2} \leq \frac{\delta}{500} \|T\|_{\text{HS}}^2 \quad \text{for all } Q \in SO(3).$$

There is an entry of  $T$  such that  $|T_{ij}| \geq \frac{1}{3} \|T\|_{\text{HS}}$ . Since the conclusion of the lemma is invariant under permutations of the rows of  $T$ , we can permute the rows in such a way that  $T_{ij}$  is on the diagonal,  $i = j$ . Furthermore, for simplicity we can assume that  $i = j = 1$ ; the general case is similar. We have

$$(B.2) \quad |(\widehat{T}\widehat{T}^T)_{11} - 1| \leq \frac{\delta}{500} \|T\|_{\text{HS}}^2 \quad \text{for all } Q \in SO(3).$$

We shall work with  $Q$  of the form  $Q = Q_1 Q_2$  where  $Q_1, Q_2 \in SO(3)$ . We shall use  $Q_1$  to mix the entries of  $T$  and  $Q_2$  to test the inequality (B.2). Let

$$Q_1 = Q_1(\phi) = \begin{bmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \phi \in [0, 2\pi],$$

and consider the matrix

$$G := TQ_1.$$

Since  $G_{11} = T_{11} \cos \phi - T_{12} \sin \phi$  and  $G_{12} = T_{11} \sin \phi + T_{12} \cos \phi$ , one can find  $\phi$  (and thus  $Q_1$ ) so that

$$(B.3) \quad |G_{11}^2 - G_{12}^2| \geq \frac{1}{9} |T_{11}|^2 \geq \frac{1}{81} \|T\|_{\text{HS}}^2.$$

Recall that  $\widehat{T} = TQ_1Q_2D = GQ_2D$ . Substituting into inequality (B.2)  $Q_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$  and  $Q_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ , we obtain

$$|d_1^2 G_{11}^2 + d_2^2 G_{12}^2 + d_3^2 G_{13}^2 - 1| \leq \frac{\delta}{500} \|T\|_{\text{HS}}^2;$$

$$|d_1^2 G_{12}^2 + d_2^2 G_{11}^2 + d_3^2 G_{13}^2 - 1| \leq \frac{\delta}{500} \|T\|_{\text{HS}}^2.$$

We subtract the second inequality from the first and conclude that

$$(B.4) \quad |(d_1^2 - d_2^2)(G_{11}^2 - G_{12}^2)| \leq \frac{\delta}{250} \|T\|_{\text{HS}}^2.$$

On the other hand, recall that  $|d_1^2 - d_2^2| \geq \delta$  by assumption and  $|G_{11}^2 - G_{12}^2| \geq \frac{1}{81} \|T\|_{\text{HS}}^2$  by (B.3). Hence  $|(d_1^2 - d_2^2)(G_{11}^2 - G_{12}^2)| \geq \frac{\delta}{81} \|T\|_{\text{HS}}^2$ . This contradicts (B.4). The proof of Lemma 4.6 is complete.  $\square$

#### ACKNOWLEDGEMENTS

The authors are grateful to Ofer Zeitouni for drawing their attention to this problem and for many useful discussions and comments. The second author learned about the problem at the IMA Workshop on High Dimensional Phenomena in September 2011; he is grateful to the IMA for its hospitality. The authors are grateful to Anirban Basak, who found an inaccuracy in the earlier version of this paper, specifically in the application of Theorem 1.2 to the Single Ring Theorem over reals. Amir Dembo communicated this to the authors, for which they are thankful.

#### REFERENCES

- [1] Ju. A. Brudnyi and M. I. Ganzburg, *A certain extremal problem for polynomials in  $n$  variables*, Izv. Akad. Nauk SSSR Ser. Mat. **37** (1973), 344–355 (Russian). MR0352825 (50 #5311)
- [2] Yu. A. Brudnyi and M. I. Ganzburg, *On the exact inequality for polynomials of many variables*. In: Proceedings of 7th Winter Meeting on Function Theory and Functional Analysis, Drogobych, 1974. Moscow, 1976, pp. 118–123 (Russian).
- [3] Persi Diaconis and Laurent Saloff-Coste, *Bounds for Kac’s master equation*, Comm. Math. Phys. **209** (2000), no. 3, 729–755, DOI 10.1007/s002200050036. MR1743614 (2002e:60107)
- [4] László Erdős, Benjamin Schlein, and Horng-Tzer Yau, *Local semicircle law and complete delocalization for Wigner random matrices*, Comm. Math. Phys. **287** (2009), no. 2, 641–655, DOI 10.1007/s00220-008-0636-9. MR2481753 (2010f:60018)
- [5] László Erdős, Benjamin Schlein, and Horng-Tzer Yau, *Wegner estimate and level repulsion for Wigner random matrices*, Int. Math. Res. Not. IMRN **3** (2010), 436–479, DOI 10.1093/imrn/rnp136. MR2587574 (2011h:60016)
- [6] Joshua Feinberg and A. Zee, *Non-Gaussian non-Hermitian random matrix theory: phase transition and addition formalism*, Nuclear Phys. B **501** (1997), no. 3, 643–669, DOI 10.1016/S0550-3213(97)00419-7. MR1477381 (98k:82091)
- [7] M. I. Ganzburg, *Polynomial inequalities on measurable sets and their applications*, Constr. Approx. **17** (2001), no. 2, 275–306, DOI 10.1007/s003650010020. MR1814358 (2002d:26017)
- [8] Michael I. Ganzburg, *Polynomial inequalities on measurable sets and their applications. II. Weighted measures*, J. Approx. Theory **106** (2000), no. 1, 77–109, DOI 10.1006/jath.2000.3484. MR1778070 (2002d:26018)
- [9] Friedrich Götze and Alexander Tikhomirov, *The circular law for random matrices*, Ann. Probab. **38** (2010), no. 4, 1444–1491, DOI 10.1214/09-AOP522. MR2663633 (2012a:60011)
- [10] Alice Guionnet, Manjunath Krishnapur, and Ofer Zeitouni, *The single ring theorem*, Ann. of Math. (2) **174** (2011), no. 2, 1189–1217, DOI 10.4007/annals.2011.174.2.10. MR2831116

- [11] Alice Guionnet and Ofer Zeitouni, *Support convergence in the single ring theorem*, Probab. Theory Related Fields **154** (2012), no. 3-4, 661–675, DOI 10.1007/s00440-011-0380-5. MR3000558
- [12] Uffe Haagerup and Flemming Larsen, *Brown's spectral distribution measure for  $R$ -diagonal elements in finite von Neumann algebras*, J. Funct. Anal. **176** (2000), no. 2, 331–367, DOI 10.1006/jfan.2000.3610. MR1784419 (2001i:46106)
- [13] Michel Ledoux and Michel Talagrand, *Probability in Banach spaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 23, Springer-Verlag, Berlin, 1991. Isoperimetry and processes. MR1102015 (93c:60001)
- [14] Hoi H. Nguyen, *On the least singular value of random symmetric matrices*, Electron. J. Probab. **17** (2012), no. 53, 19, DOI 10.1214/EJP.v17-2165. MR2955045
- [15] Mark Rudelson, *Invertibility of random matrices: norm of the inverse*, Ann. of Math. (2) **168** (2008), no. 2, 575–600, DOI 10.4007/annals.2008.168.575. MR2434885 (2010f:46021)
- [16] Mark Rudelson and Roman Vershynin, *The Littlewood-Offord problem and invertibility of random matrices*, Adv. Math. **218** (2008), no. 2, 600–633, DOI 10.1016/j.aim.2008.01.010. MR2407948 (2010g:60048)
- [17] Mark Rudelson and Roman Vershynin, *Non-asymptotic theory of random matrices: extreme singular values*, Proceedings of the International Congress of Mathematicians. Volume III, Hindustan Book Agency, New Delhi, 2010, pp. 1576–1602. MR2827856 (2012g:60016)
- [18] Terence Tao and Van H. Vu, *Inverse Littlewood-Offord theorems and the condition number of random discrete matrices*, Ann. of Math. (2) **169** (2009), no. 2, 595–632, DOI 10.4007/annals.2009.169.595. MR2480613 (2010j:60110)
- [19] Terence Tao and Van Vu, *Random matrices: the distribution of the smallest singular values*, Geom. Funct. Anal. **20** (2010), no. 1, 260–297, DOI 10.1007/s00039-010-0057-8. MR2647142 (2011m:60020)
- [20] Terence Tao and Van Vu, *Random matrices: universality of ESDs and the circular law*, Ann. Probab. **38** (2010), no. 5, 2023–2065, DOI 10.1214/10-AOP534. With an appendix by Manjunath Krishnapur. MR2722794 (2011e:60017)
- [21] Terence Tao and Van Vu, *Random matrices: universality of local eigenvalue statistics*, Acta Math. **206** (2011), no. 1, 127–204, DOI 10.1007/s11511-011-0061-3. MR2784665 (2012d:60016)
- [22] Roman Vershynin, *Introduction to the non-asymptotic analysis of random matrices*, Compressed sensing, Cambridge Univ. Press, Cambridge, 2012, pp. 210–268. MR2963170
- [23] R. Vershynin, *Invertibility of symmetric random matrices*, arXiv:1102.0300v4, Random Structures and Algorithms, to appear.
- [24] O. Zeitouni, personal communication.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MICHIGAN, 530 CHURCH STREET, ANN ARBOR, MICHIGAN 48109

*E-mail address:* rudelson@umich.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MICHIGAN, 530 CHURCH STREET, ANN ARBOR, MICHIGAN 48109

*E-mail address:* romanv@umich.edu